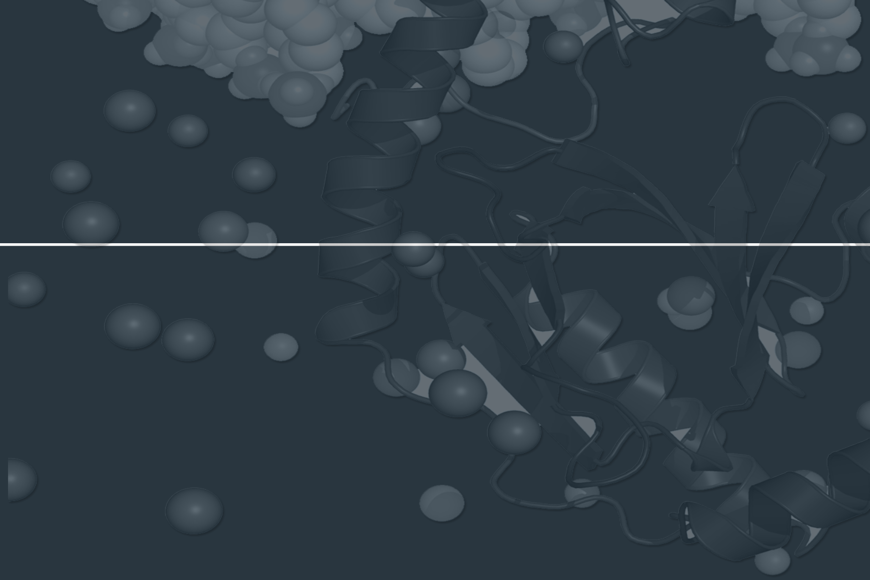


PITTSBURGH SUPERCOMPUTING CENTER
300 S. CRAIG STREET
PITTSBURGH, PENNSYLVANIA 15213

PITTSBURGH SUPERCOMPUTING CENTER

PEOPLE. SCIENCE. COLLABORATION.

SPRING 2016

INTEL AD

PITTSBURGH SUPERCOMPUTING CENTER provides university, government and industrial researchers with access to several of the most powerful systems for high performance computing, communications and data storage and handling available to scientists and engineers nationwide for unclassified research. PSC advances the state of the art in high performance computing, communications and data analytics and offers a flexible environment for solving the largest and most challenging problems in computational science. As a leading partner in XSEDE, the Extreme Science and Engineering Discovery Environment, the National Science Foundation's cyberinfrastructure program, PSC works with other XSEDE participants to harness the full range of information technologies to enable discovery in U.S. science and engineering.

PSC.EDU

CONTENTS

SPRING
2016

FROM THE DIRECTORS

TICK, TOCK GOES THE PSC CLOCK …

Some new systems augment previous research capabilities through improved performance, although they are very similar to previous systems. They are the ticks. But it is the TOCKS, the systems that break new ground in technology and architecture, that can, in addition to improving performance for existing application areas, have a disruptive effect. These are the systems that open advanced IT capabilities for a wide range of new fields of scientific research.

The Bridges system, now entering production and whose initial scientific output is described in this 30th anniversary issue of People. Science. Collaboration. (p. X), is just such a system. With its new hardware, architecture, software and management capabilities it is already supporting multiple new, exciting areas of scientific research.

Using heterogeneous hardware components flexibly configured with powerful system management software, Bridges can fill a far wider range of research needs than would be possible with a statically configured system of identical components. Using software to reconfigure a tightly integrated (thanks to the first application of Intel's new OmniPath architecture) system of very different components (GPU nodes, extreme memory nodes offering 12 TB of coherent shared memory, database nodes, etc.), Bridges can become a number of different supercomputers tailored for different users and computational tasks. Included is a growing number of "new community" users that have never before needed HPC—and, therefore, have never before had a system quite tailored to their needs. PSC's previous systems have been friendly to new communities; Bridges is, in a virtual sense, created for them, giving them unparalleled ability to port existing successful computational methodologies into HPC rather than having to rewrite their codes and procedures.

PSC also continues to help pave the way to super-specialized HPC systems that do a single task far more quickly than possible with an all-purpose supercomputer. Thanks to an agreement with D. E. Shaw Research (DESRES) and a new National Institutes of Health grant, we will continue hosting the Anton molecular dynamics simulation system until October (p. X). More importantly, at that point we will begin hosting a far-more powerful DESRES Anton 2, capable of modeling much larger subcellular structures than just proteins or small groups of proteins.

While we look forward to the future typified by these two very different but very powerful machines, our 30th anniversary also provides an opportunity to look back at the progress we've made. In this volume we have paired most of the new stories with PSC "greatest hits" about earlier work in similar fields. The story about Anton 2 includes a "PSC Lookback" sidebar that harkens back to University of Pittsburgh researchers' use of one of our earliest machines, the Cray Y-MP, in the 1990s in an early molecular dynamics study of the interaction of the EcoR1 protein with DNA. A new study of driver-injury modeling appears alongside Alcoa's work with our Cray C90 to model aluminum auto body designs in the late 1990s (p. X). A report of how key stocks can be used to better predict movement in the market pairs with 2009 Carnegie Mellon University work using PSC's Pople system to show how

Michael Levine (left) and Ralph Roskies, PSC co-scientific directors

THE STAFF

JACOB Czech
GREGORY Hood
ROZITA Laghaei
ALEXANDER Ropelewski
ARTHUR Wetzel

MARCELA Madrid
DEB Nigra
RICHARD Costa
ROBERTO Gomez
ANIRBAN Jana
THOMAS Maiden
DAVE O'neal

SERGIU Saneilevici
YANG Wang

ROBIN Scibek
PHILLIP Blood
STEPHEN Deems
KENNETH Hackworth
NICHOLAS Nystrom
JOHN Urbanic
JOEL Welling

ANJANA Kar
SUSAN Litzinger

easily social security numbers could be guessed just from publically available information on the Web (p. X). PSC's collaboration with Texas A&M researchers on coolant flow through a promising new, inherently safer, nuclear reactor design recalls a 1993 model of solar turbulence using the C90 (p. X). An exciting new 3D reconstruction of subcellular connections between excitatory neurons in the visual cortex, made possible with PSC hardware, software and expertise is reminiscent of early C90 work in reconstructing the 3D anatomy of the beating heart (p. X). And a University of Delaware study with Anton of how cholesterol affects biological membranes appears with a PSC Lookback on University of Illinois work using our LeMieux system (at the time, the world's second-fastest computer) that showed how the aquaporin membrane channel excludes hydrogen ions—worked cited in aquaporin-discoverer Peter Agre's 2003 Nobel Prize acceptance speech (p. X). See the table of contents for more of these stories and their associated Lookback pieces (p. X).

Not all of the stories in this celebratory issue can easily be related to early PSC work, though, and that's part of the point of Bridges. A University of Illinois group used our now-retired Blacklight system, the transitional Greenfield system and soon Bridges to glean clues about the life experiences of black women in historical periods when they could not share their lives in print (p. X).

This work is exemplary of the new users whom, along with those in "traditional" supercomputing fields, we expect Bridges to serve by design. We look forward to helping and working with researchers in fields never before employing high-performance computing to use Bridges to make transformational contributions. And with the help of the XSEDE network's programs, sites and personnel, Bridges will reach an ever-expanding community of new users.

Education and workforce development continue to accompany research and technological development as major foci for PSC's efforts. This year we are proud to note that the leaders of PSC's bioinformatics education group have received a prestigious Carnegie Science Center

Award for STEM Education (p. X). Accompanying this article is a pointer to a few of the many successful bioinformatics researchers and educators who have benefitted from our educational programs. In addition, our Director of Education, Outreach and Training, Cheryl Begandy, has been recognized with the 2016 YWCA Greater Pittsburgh Tribute to Women Leadership Award for Science & Technology.

We at PSC enter our fourth decade (!) excited to see where new systems, new science and above all the continued ingenuity of our people will take us next.

As always we would like to acknowledge all our funders, especially the NSF, the NIH, the Commonwealth of Pennsylvania. We would also like to hear any feedback you have, on our work or this publication. You can send any comments or suggestions via our feedback page at psc.edu/index.php/contact/607. You can also contribute to PSC's nonprofit, academic mission at psc.edu/donate.

ANDREW Adams
EDWARD Berger
ROBERT Budden
BURT Cubbison
SHANE Filus
BRYON Gill
EDWARD Hanna
BRIAN Johanson
ROB Light
JAMES Marsteller
CLINTON Perrone
J RAY Scott
DEREK Simmel

JASON Sommerfield
RICHARD Underwood
BRYAN Webb
EDWARD Wozniak
ZHIHUI Zhang

KATHY Benninger
RICHARD Angeletti
STEPHEN Cunningham
KENNETH Goodwin
MICHAEL Lambert
BRYAN Learn
CHRIS Rapier

NATHAN Robinson
JENNIFER Bakal
LEILA Haidari
SHAWN Brown
JAY DePasse
JIM Leonard
ELIJAH Zenkov

CHERYL Begandy
VIVIAN Benton
PALLAVI Ishwad
SHANDRA Williams
KENNETH Chiacchia

ERICA Anderson
MICHELLE Devlin
DAVID Kapcin
RAY Nardozzi
LUCILLE Jarzynka
PAT Sudac

JAMES Kasdorf
DAVID Moses
JANET Brown
ROBERT Stock

OPENING BRIDGES

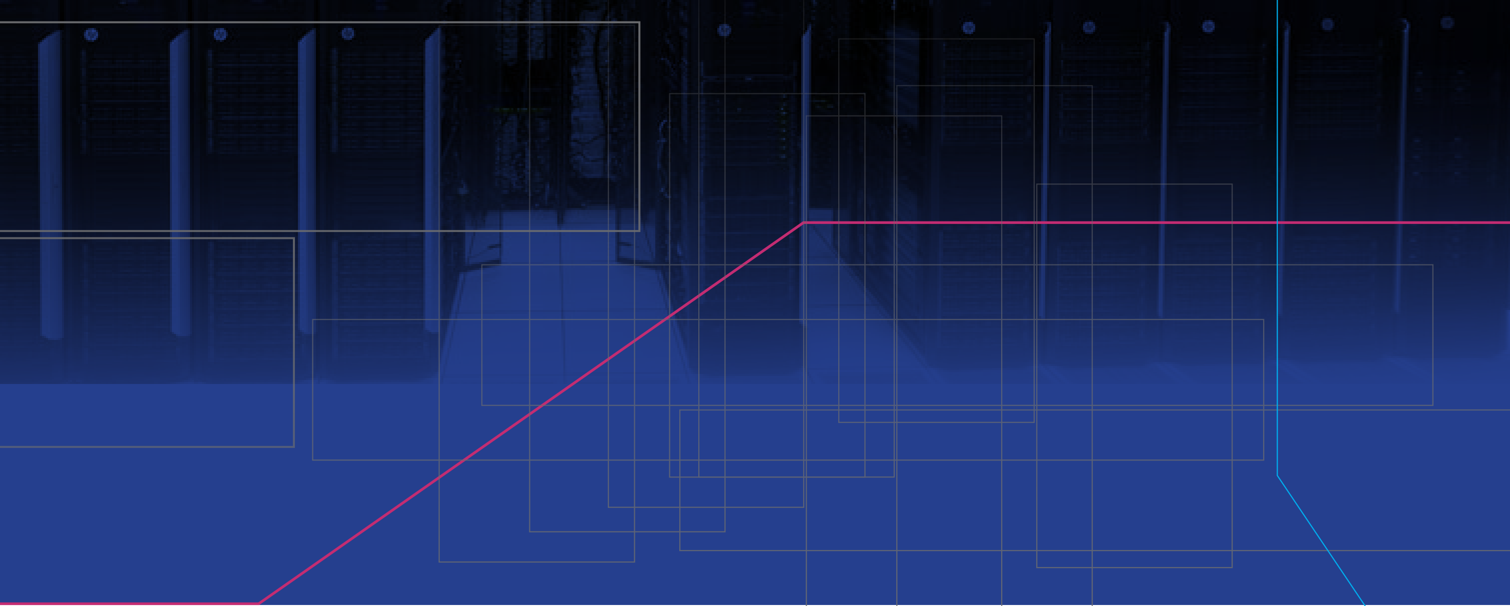
First Users Score Early Successes on PSC's Game-Changing Supercomputer

Bridges, PSC's new National Science Foundation-funded computing system, has seen its first month of use by the national scientific community. In that short time, users have reported progress in fields such as genomics, public health, chemistry, machine learning and more. As of April 22, 196 projects have been allocated time on Bridges, and we expect many more to come.

Rather than just an incremental improvement in performance, Bridges represents a new way of doing business in high-performance computing (HPC). Researchers can adapt its flexible architecture to their specific needs, in effect creating a "custom supercomputer." Users can select from the system's features including massive computing power, components optimized for different types of computation and common computational tools not normally found on supercomputers, such as databases and popular Big Data software packages. And all of this comes with an unsurpassed ease of entry for researchers who never before needed HPC. Following is a sampling of Bridges' early successes.

Bridges' first users were the infectious disease experts of the National Institutes of Health-funded MIDAS network. In their first Public Health Hackathon at PSC in February, they gave twelve teams from across the U.S. and India the task of **using Bridges to visualize data in a way that transformed understanding of an issue in public health.** A team from Carnegie Mellon University's Department of Statistics took first place with their SPEW VIEW tool, which maps the historical spread of diseases in the U.S. Teams from the University of Pittsburgh Department of Biomedical Informatics and PSC's Public Health Applications Group took second and third place, respectively. Omni-Path Architecture and Dell sponsored the hackathon.

Timothy Hele and Eric Fuemmeler of Cornell University used Bridges to calculate **the electronic structure of TIPS pentacene, a large organic semiconductor molecule with applications in solar power cells.** On Bridges, they were able to run two major computations in one job with the same memory requirements, saving time and improving the productivity of their research. Previously, they hadn't been able to run those calculations on any other supercomputer available to them.



The PSC Public Health Application Group's own Jay DePasse has used Bridges to **model the possible benefits of flu vaccine choice** in Washington D.C., Allegheny County, Pa., and Salt Lake City. DePasse used "agent-based modeling," in which every person in an area is represented by a realistic virtual human in the simulation. He tested the results of offering both adults and children a choice between the new quadrivalent vaccine—which protects against more strains but is also more expensive—and the earlier trivalent vaccine. His initial results suggest that this choice policy would be more cost-effective than alternatives such as no choice of vaccine, choice offered to children only and choice offered to adults only.

The new quadrivalent flu vaccine protects against more virus subtypes, but is more expensive. Whoisjohngalt [Public domain] via Wikimedia Commons.

James Denvir and Swanthana Rekulapa of Marshall University in West Virginia assembled **the genetic sequences of two species, the Narcissus flycatcher and the critically endangered Sumatran rhinoceros.** They used a "*de novo* assembly" method, in which scientists first sequence millions of small fragments of DNA. Then the method uses brute computational force to piece together the fragments' order via the sequences where they overlap. Using Bridges' 3-terabyte large memory nodes, the researchers pieced together the 1 billion DNA bases of the bird genome in 6.6 hours—almost five times faster than possible with the systems previously available to them. The rhino assembly was also faster than possible elsewhere, with 3 billion bases assembled in 11 hours.

Wenxuan Zhong and Xin Xing of the University of Georgia used Bridges to assemble **378 billion base pairs of bacterial DNA from the intestines of healthy patients and those with diabetes.** Such "metagenome assembly" doesn't even try to chemically separate the DNA from many microbial species in a sample. Instead, the scientists sequence short DNA fragments of all the species at once, using computation to sort out the different microbes' sequences as they assemble them. This massive task leveraged Bridges' Intel Omni-Path internal connections—the first such installation in the world—between 20 computational nodes to finish the calculation in a blistering 16 hours. The team is now using Bridges to test a new statistical method on the sequence data to identify critical differences in gut microbes associated with diabetes.

Escherichia coli, the common gut bacterium. Rocky Mountain Laboratories, NIAID, NIH [Public domain], via Wikimedia Commons

RESCUED HISTORY

Massive Data Analysis Uncovering Black Women's Experiences

We say, "History is written by the victors." But it's probably more true to say it's written by the people who have the opportunity to write.

One example of this is the study of Black women, their lives and their experiences. Many earlier documents don't mention them directly, though these works may offer clues. Those that do mention Black women often are historically obscure, hidden away in vast library collections and unintentionally misleadingly titled or cataloged. Until recently researchers had no good way of recovering this "lost history."

Ruby Mendenhall of the University of Illinois at Urbana-Champaign is leading a collaboration of social scientists, humanities scholars and digital researchers that hopes to harness the power of high-performance computing to find and understand the historical experiences of Black women, searching two massive databases of written works from the 18th through 20th centuries. The work also offers a common toolbox that can help other digital humanities projects.

PSC staff helped this research with the support of several programs within the NSF's XSEDE network of supercomputing centers: Extended Collaborative Support Service; Novel and Innovative Projects Program.

"With a Big Data approach we get a chance to make use of hundreds of thousands of texts—journals, books, periodicals," says Mendenhall. "The number is greater than what you would normally be able to look at during an entire career."

POWERING UP

Mendenhall's team realized that to search tens or even hundreds of thousands of books, articles and letters, they'd need considerably more computing power than available on a typical campus "cluster" of commodity computers. They consulted with colleagues on campus who were members of the NSF's XSEDE network of supercomputing centers. With these new collaborators, they identified PSC's now-retired Blacklight supercomputer as a good fit for their project. With help from PSC's Sergiu Sanielevici, they adapted their earlier work to Blacklight and then moved the project to PSC's interim Greenfield system, a precursor to the new Bridges system (see p. Z).

"We chose Blacklight specifically because the tools we're using need huge amounts of shared memory," a strength of that system, says Mark Van Moer, an XSEDE staff member at the National Center for Supercomputing Applications at the University of Illinois who worked as the team's visualization specialist. PSC's continued focus on memory-intensive computing with Greenfield, and soon the new Bridges system, support such work well.

"History is written by the victors."

Sculptor Edmondia Lewis (1844-1907), bottom right, was the first woman of African and Native American descent to achieve notoriety in the fine arts world. She spent most of her career in Rome. Henry Rocher – National Portrait Gallery, Smithsonian Institution, Public Domain (PD-1870)

Using Blacklight, the researchers analyzed 20,000 documents known to contain information about Black women in the HathiTrust and JSTOR databases to create a computational model. They're now using this model on Greenfield to study the entire 800,000 documents in both databases.

MAKING SENSE OF WORDS VIA NUMBERS, GRAPHICS

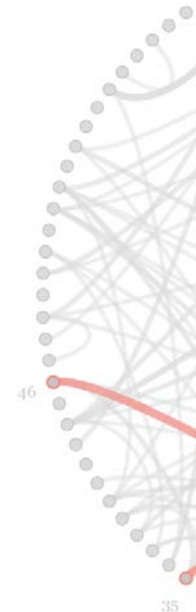
To make sense of the numbers, the investigators turned to two sets of computational techniques: topic modeling and data visualization.

Topic modeling looks at how often certain key words associate with each other. For example, a book that contains the word "negro"—at the time considered the most respectful term to describe Black men and women—the word "vote" and the word "women" might offer clues about Black women's participation in the women's suffrage movement. Mike Black, then at the University of Illinois and currently at the University of Massachusetts, headed this topic-modeling project for the team.

"We're hoping, in the next stage, to ramp up and check these topics against the larger corpus of works," Mendenhall adds. "Do the 'recovery' part."

Van Moer's data visualization work is building ways of displaying results in a way that helps humans make more intuitive sense of them. A "tree map" displays key words in boxes that correspond to each word's frequency (right). A "network graph" charts how often key words appear close to each other, also offering insight into how those words are being used and what they mean in context (p. 13).

Yet another visualization technique plots key terms in histographs that allow users to track the emergence and prominence of a given topic over time.



MAKING SENSE OF THE NUMBERS

An initial finding on Blacklight confirmed the prediction that the same documents referenced the post-World-War-I Black Women's Club and New Negro movements. This raises interesting questions about how the two movements, which historians knew were contemporaneous, may have interacted. The Illinois researchers hope to begin answering these questions in their current work on Greenfield, as well as proposed work on Bridges.

Tree maps such as this one, help researchers view the statistics relating to word frequency in a way that spurs insight.

Network graphs such as this one are another visualization method that shows how strongly topics correlate across documents with a topic of interest—in this case, the team's "Topic 21," a cluster of concepts spanning labor, employment and industry.

topic21	topic35	topic46	topic 90	topic15
worker	county	farm	house	work
labor	city	land	family	day
employment	state	farmer	area	labor
job	town	acre	home	time
percent	york	crop	city	pay
industry	public	agricultural	build	hour
defense	district	county	live	week
service	mayor	family	room	wage
work	relief	agriculture	unit	make
increase	local	cotton	community	month
employ	population	labor	project	condition
train	person	state	income	year
unemployment	citizen	area	rend	employ
wage	community	migrant	move	find
employee	place	rural	neighborhood	case
department	residence	year	facility	money
employer	resident	tenant	urban	service
occupation	board	migration	low	receive
rate	settlement	make	occupy	leave
number	large	large	resident	care
industrial	welfare	grower	lodge	employer
production	number	small	condition	order
earnings	part	california	apartment	good
woman	aid	camp	neighborhood	require

"The beauty of computation and Big Data lies in how it complements the traditional close reading," says Nicole Brown, a postdoctoral fellow in Mendenhall's group who is interpreting the computational results in light of Black feminist theory. "The two methods complement each other to give you a full picture of what's going on."

"Working with the social science and humanities people has been a real eye opener in a lot of ways," adds Van Moer. "In the previous seven years I pretty much worked with physical scientists. Humanities and social science researchers have to be worried about not just what the numbers mean at a surface level. They have a whole theory behind how you go about interpreting things as it relates to the larger society—that's really an interesting aspect of the project for me."

Another of the group's goals will be to create a set of computational tools that researchers in many fields will be able to use to search various texts for topics of interest—and to understand how those topics interrelate. Topic modeling and visualization methods can be modules in a larger toolbox for digital humanities research.

"We're generally interested in Black women and their life experience," Mendenhall says. "But we also see this as a tool that social scientists and people in the humanities can use to study many topics."

Stitching Thought

PSC Powers Harvard's, Allen Institute's 3D-Reconstruction of Excitatory Visual Neuron Wiring

WHY IT'S IMPORTANT:

One of the mysteries of brain function is how we make sense of the jumble of images that confront our eyes. Neuroscientists have discovered that most individual nerve cells in the brain respond to specific elements in the visual environment. For example, a nerve cell may fire in response to vertical lines—another, in response to horizontal or slanted lines. Researchers suspected that the mammalian cortex amplifies this signal by having nerve cells that respond to similar elements excite each other. This mutual excitation may help those elements stand out and prime the network for their further processing. But scientists had no anatomical evidence that this actually happens.

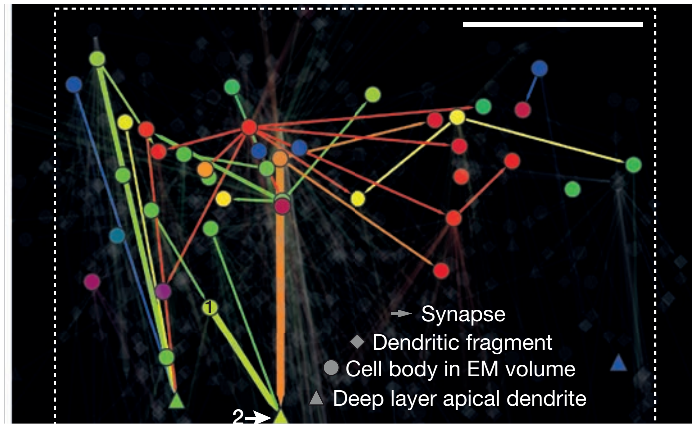
> The challenge here was to go from a block of brain tissue to a coherent, three-dimensional digital volume that someone could then visualize on a computer and trace individual nerve-cell axons and dendrites. This involves careful cutting of the tissue block into extremely thin slices, staining them and then transferring these slices into an electron microscope where they are imaged one small area at a time.

TOGETHER

> In this process there are inevitable physical distortions, and, if one simply stacked the images together, nothing would line up properly. Supercomputing allows us to register these images to one another and reconstruct a well-aligned volume.
>
> —Greg Hood, PSC

HOW PSC HELPED:

A team led by Wei-Chung Allen Lee of Harvard University and R. Clay Reid of the Allen Institute for Brain Science in Seattle identified brain nerve cells that respond to visual elements in living mice. Then they took a series of millions of microscope images of ultra-thin (about 40-nanometer-thick) tissue slices around these nerve cells.



Stylized drawing showing how the excitatory visual neurons connect with each other. Arrows represent excitatory connections, their thickness showing how many other neurons each connects with. Colors code for what orientation of visual lines each neuron responds to; circles are target neurons, triangles large-caliber neuron extensions that exit the image, and diamonds are other neurons. Adapted by permission from Macmillan Publishers Ltd: Lee et al. (2016) Anatomy and function of an excitatory network in the visual cortex. *Nature* 532(7599):370-374.

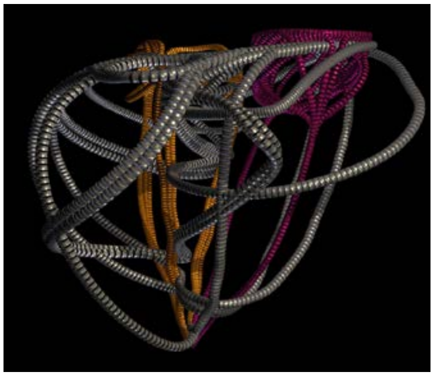
In a collaboration facilitated by PSC's Art Wetzel, PSC computational scientist Greg Hood helped them to reconnect these images into a three-dimensional volume using PSC's AlignTK software. But because these slices are fragile, microscopic tears and other artifacts happened, requiring manual intervention to correct. So researchers had to move back and forth between computation and manual "repair" of the images until the quality of the aligned volume was good enough to trace the connections between the nerve cells. The team reported in the journal *Nature* in March 2016 that mammalian excitatory nerve cells that respond to a given visual feature do indeed make more and larger connections to other excitatory nerve cells that are tuned to respond to similar visual elements.

DEEPER DIVE: A COMPUTATIONAL TOUR DE FORCE

In their *Nature* paper, the Harvard and Allen Institute group and coauthor Hood have produced a reconstruction of an excitatory nerve-cell network in the brain's cortex at a subcellular level. The work was a tour de force of two-photon fluorescence microscopy (to identify the nerve cells responding to a given orientation of visual lines), electron microscopy and computation. To connect the electron microscope images into a 3D reconstruction, the scientists first had to merge thousands of individual images for each slice into a single large image of the entire slice. They next aligned the adjacent slices. They could then combine these pairwise alignments over the entire stack to calculate exactly how to correct the distortion present in each slice and place its corrected image back into an aligned stack.

The group only reconstructed 0.03 cubic millimeters of the mouse brain, a volume that would go into a teaspoon about 167,000 times. But this still resulted in about 10 million camera images, amounting to roughly 100 Terabytes of raw data—about the memory required for nearly 30 million high-resolution, large-format photographs.

Greg Hood's work was supported by a National Institutes of Health grant to the National Center for Multiscale Modeling of Biological Systems (MMBIOS).



PSC LOOKBACK:
RECONSTRUCTING THE BEATING HEART

Sometimes, a breakthrough in computational research pays an extra dividend. That's what Charles Peskin and David McQueen of New York University got when they employed PSC's Cray C90 supercomputer in a 1993 effort to model the human heart in three dimensions. Their main goal was to improve the simulated flow of blood through the aortic valve.

The C90 allowed them to increase the number of points in the mesh-like grid used in the calculations, and the extra resolution did indeed render the blood motion through the valve more realistically. But it also fixed problems they'd been having with the right side of their simulated heart—a breakthrough that made the whole heartbeat work correctly for the first time in 15 years of research. Their accomplishment gained them the 1994 Computerworld Smithsonian Award for Breakthrough Computational Science; Peskin also received the 1994 Sidney Fernbach award.

> There's really limited information you can get from a crash-test dummy—you get only about 20 data points. Our human body model gives us much more, predicting injuries in organs that aren't in that dummy, such as lung contusions.
>
> —Ashley A. Weaver, Wake Forest University

CIREN to the Rescue

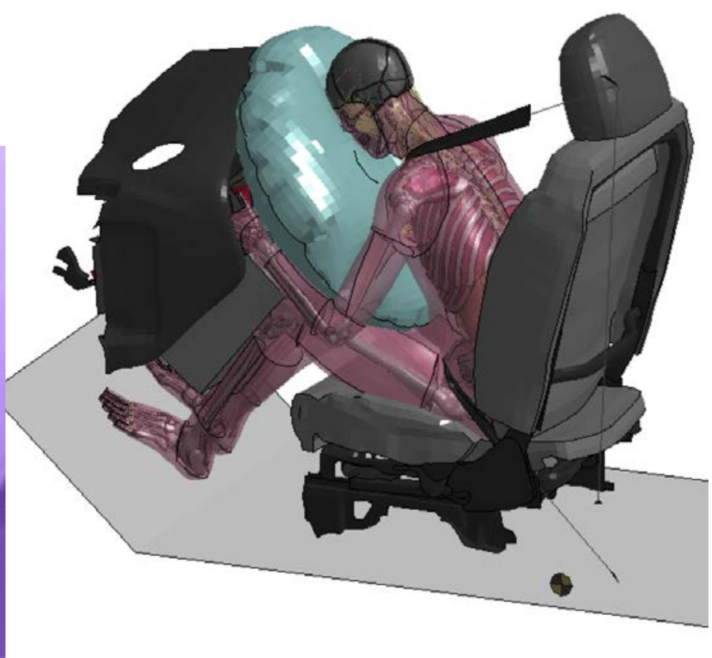
Blacklight simulations give new insight to crash injuries, better prevention

WHY IT'S IMPORTANT:

Over 33,000 Americans die in motor vehicle crashes annually, according to the U.S. Centers for Disease Control. Modern restraint systems have decreased deaths, but some deaths and injuries remain—and restraints themselves can cause some injuries. "Crash-test dummies" can help engineers design safer cars but provide only limited information about the forces experienced by the body in an impact. Improved computer models of vehicle crashes can predict how restraints and other safety systems could work better. They could also help researchers to duplicate the effects of thousands of changes that would be far too slow to test in physical crash tests.

HOW PSC HELPED:

The Crash Injury Research and Engineering Network (CIREN) has created a database of real-world vehicle crashes for researchers to test using computer models. Ashley A. Weaver of Wake Forest University, a CIREN center, used XSEDE's Extended Collaborative Support Service to select PSC's Blacklight supercomputer to run thousands of simulations drawn from hundreds of cases in virtual autos simulating the Toyota Camry and Chevrolet Cobalt. Using the Total Human Model for Safety, they showed that the model can reproduce real-world injury patterns and predict details that crash-test dummies can't provide. Along the way they showed how injury-causing stress moves from the foot to the lower leg as a driver's head comes forward into a frontal airbag, and that more reclined seating positions can lead to higher risk of head and chest injuries. Weaver and colleagues published their initial findings in *Traffic Injury Prevention* in October 2015.



Immediate results of the Blacklight simulations suggested that more reclined positions produce higher risk of head and chest injuries in higher-speed collisions.

PSC staff helped this research with the support of several programs within the NSF's XSEDE network of supercomputing centers: Extended Collaborative Support Service; Novel and Innovative Projects Program.

PSC LOOKBACK: MATERIAL DIFFERENCES

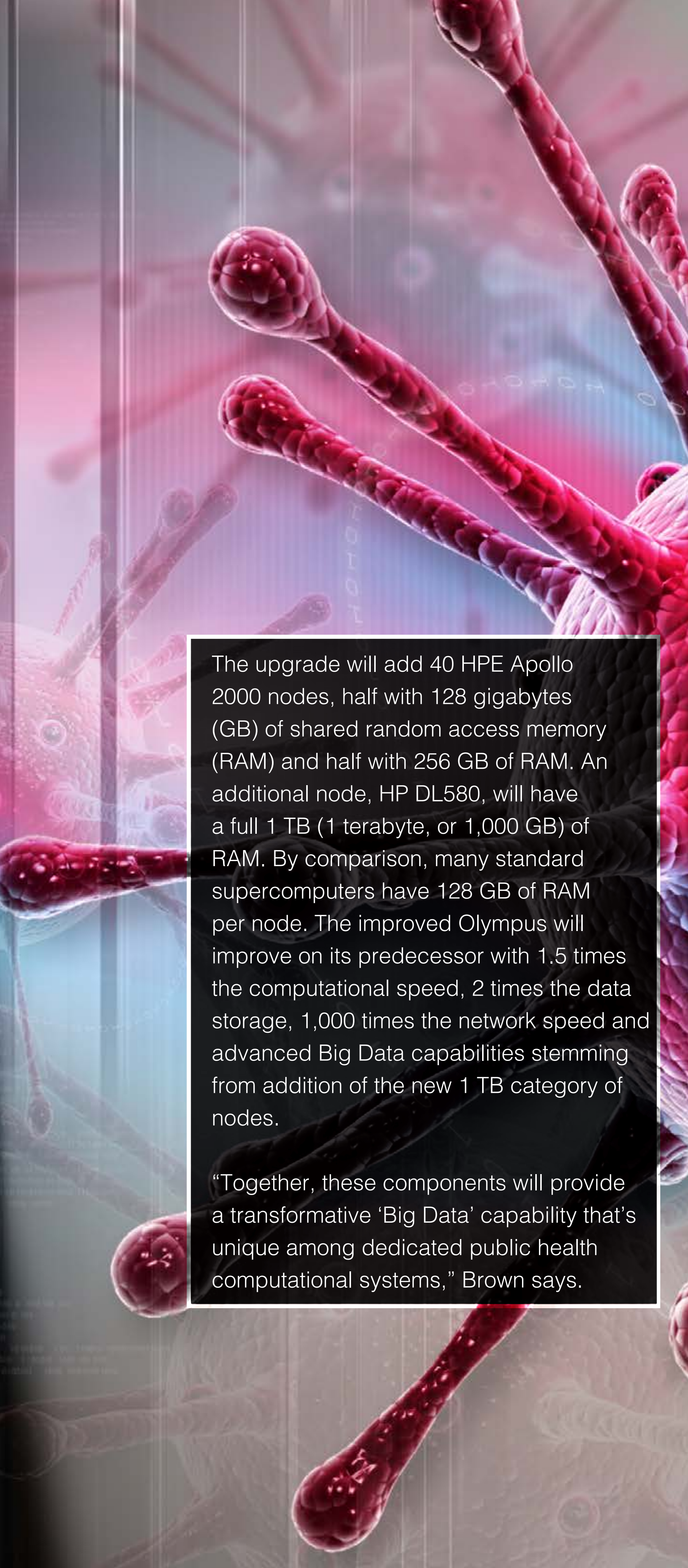
PSC has been involved in industrial-related projects like CIREN almost since its inception. A partnership with aluminum manufacturer Alcoa began in 1987, with PSC computing resources helping optimize the design of the familiar modern, lightweight aluminum beverage can. Another fruit of this partnership was the design of the aluminum "space-frame" construction of the Plymouth Prowler and the Audi A8 autos in 1998. By performing initial design steps on PSC's Cray C90, Alcoa developed these components at a savings of hundreds of thousands of dollars and weeks of trial-and-error design. These projects were even more important for Alcoa's bigger picture: with plastics challenging aluminum's domination of lightweight, high-strength structural materials at the time, supercomputing helped keep the metal competitive and then some.

Scaling (UP) Olympus

PSC Public Health Applications Group Leverages Bridges Construction to Upgrade Modeling Cluster

PSC's work with Hewlett Packard Enterprise (HPE) and Intel on its new Bridges system (see p. X) has paid dividends for the public health community in the U.S. and worldwide. Leveraging the ongoing relationship with these vendors, PSC's Public Health Applications Group is upgrading Olympus, the dedicated computer cluster it administers for the Models of Infectious Disease Agent Study (MIDAS) Research Network. The MIDAS network is a National Institutes of Health-funded international consortium of infectious disease transmission modelers.

"The upgrade to Olympus will provide computing capability to MIDAS and greatly increase the performance and capacity of the cluster," says PSC's Shawn Brown, leader of the MIDAS MISSION Group, which organizes MIDAS software developers around the world to share software, provide best-practices in development and validation and gather requirements for user-focused tools. "The system is particularly designed to provide state-of-the-art capability for performing public health simulations of infectious disease spread and how to stop them."

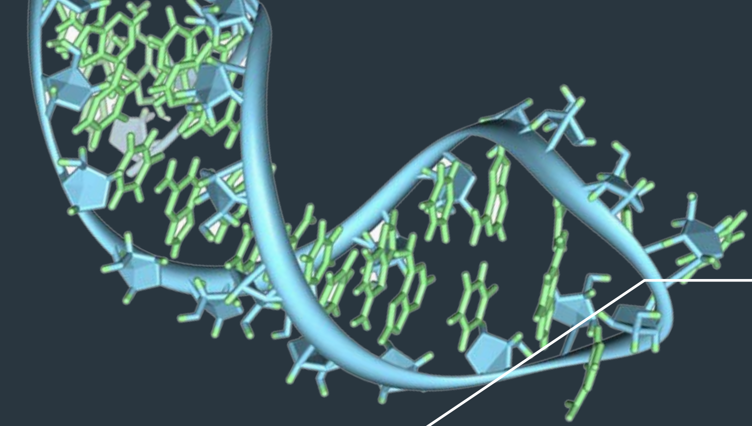


The upgrade will add 40 HPE Apollo 2000 nodes, half with 128 gigabytes (GB) of shared random access memory (RAM) and half with 256 GB of RAM. An additional node, HP DL580, will have a full 1 TB (1 terabyte, or 1,000 GB) of RAM. By comparison, many standard supercomputers have 128 GB of RAM per node. The improved Olympus will improve on its predecessor with 1.5 times the computational speed, 2 times the data storage, 1,000 times the network speed and advanced Big Data capabilities stemming from addition of the new 1 TB category of nodes.

“Together, these components will provide a transformative ‘Big Data’ capability that’s unique among dedicated public health computational systems,” Brown says.

PSC LOOKBACK: CAPTURING THE TWISTER

While PSC’s Public Health Applications Group is a relatively new effort, it isn’t the center’s first foray into improving public safety. In 2004, Ming Xue of the University of Oklahoma used PSC’s Hewlett Packard-delivered LeMieux supercomputer to create the first realistic simulation of tornado formation. In a period when tornado warnings were three times as likely to be false alarms as lifesaving alerts, PSC’s terascale system allowed Xue to simulate a “supercell” thunderstorm at a scale of 20 to 25 meters. That was far better than the 100-meter resolution of previous simulations and the first that simulated at a finer level than the typical tornado path width of 46 meters. Just like the real 1977 storm it was modeling, the LeMieux simulation formed a funnel cloud, paving the way to today’s lifesaving tornado warnings.



CELEBRATING

YEARS

INNOVATION THROU

2001

1992 1993 1995 1995

1992

1990

1986 1989

Cray X-MP/48
(1986 - 1989)

Cray Y-MP/832
(1989 - 1993)

CM-2
(1990 - 1992)

CM-5
(1992 - 1993)

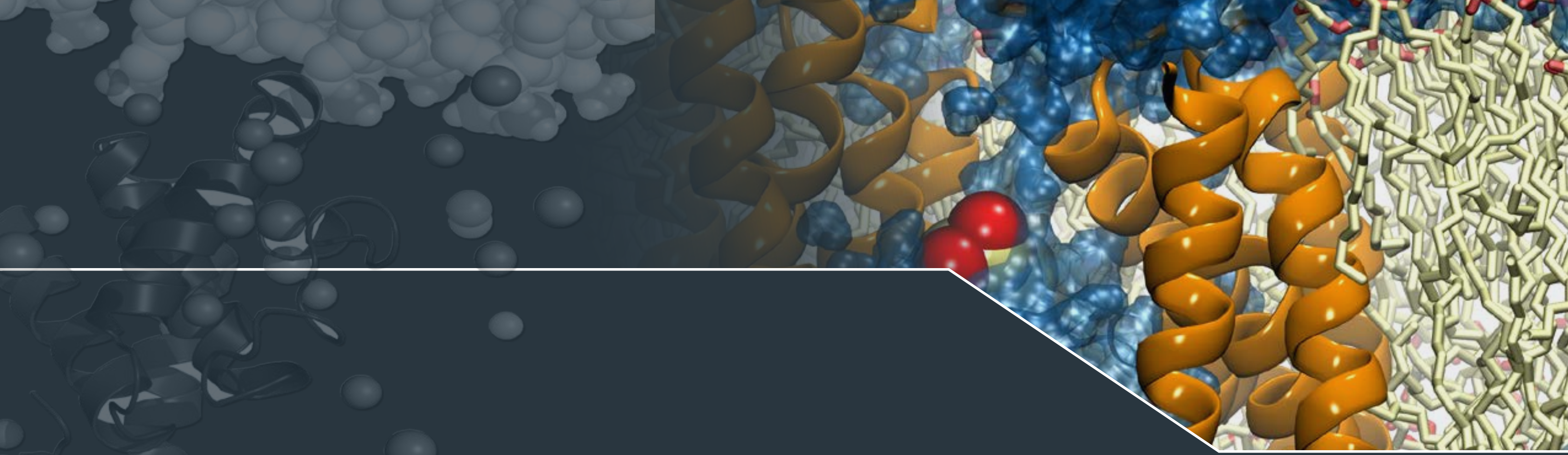
Cray C90
(1992 - 1999)

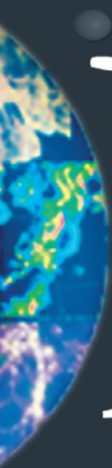
Cray T3D
(1993 - 1999)

Cray J90s
(1995 - 2002)

Cray T3E / Jaromir
(1995 - 2004)

Lemieux / TCS
(2001 - 2006)





PSC

…GH COLLABORATION

2016

BRIDGES
(2016 -)

2010

Blacklight
(2010 - 2015)

2009

Warhol
(2009 - 2013)

2008

Pople
(2008 - 2011)

Salk
(2008 - 2015)

2005

Cray XT3 / BigBen
(2005 - 2010)

2003

Marvels / Jonas & Rachel
(2003 - 2008)

Not Your Grandfather's

PSC Studies of Coolant Flow Promise Improved Efficiency, Reliability in Advanced Nuclear Reactor Design

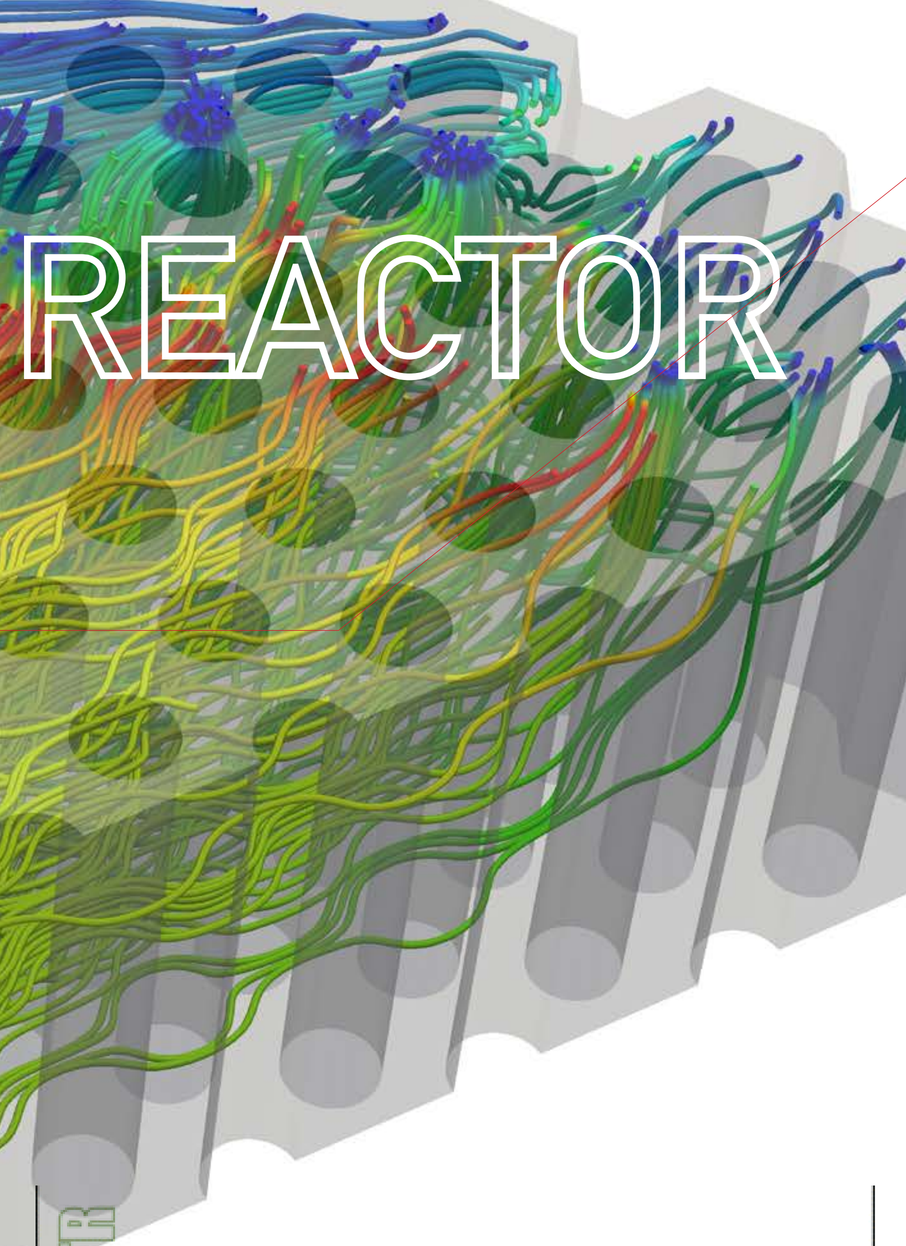
WHY IT'S IMPORTANT:

Nuclear power is seeing a Renaissance worldwide. About 20 percent of the U.S. electric supply comes from nuclear—the nation's largest carbon-free source of electrical energy. "Fourth generation" reactor designs such as the Very-High Temperature Reactor (VHTR) promise to be much more safe, sustainable and efficient over the technology of the 1980s and 1990s that underlies currently operating nuclear power plants in the U.S. In addition to electricity, VHTRs may also produce hydrogen and process heat for industrial use, really increasing their economic value. They will generate much less nuclear waste. And they will be proliferation-resistant since they will employ a new type of nuclear fuel that will be very hard to use to make a fission bomb. As with any new technology, though, advanced reactor designs pose many engineering challenges. In particular, the high pressure and temperature and highly turbulent helium coolant in the VHTR could put severe strains on the reactor's structure, limiting the operational lifetime of components and increasing costs.

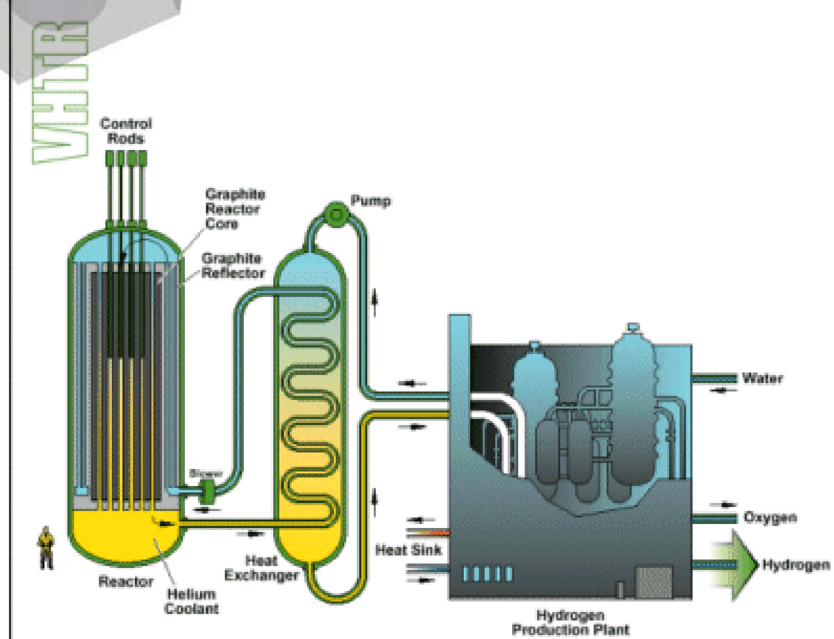
HOW PSC HELPED:

PSC's Anirban Jana is a co-principal investigator and computational lead of a study of coolant flow in the VHTR design. Working with principal investigator Mark Kimber, who recently moved from the University of Pittsburgh to Texas A&M University, Jana has used the OpenFOAM software on PSC's supercomputers Blacklight and Bridges, as well as the Texas Advanced Computing Center's Stampede, to simulate jets of hot coolant emerging from the reactor core. These simulations faithfully reproduce the turbulent mixing of the coolant jets, promising important computational tool for understanding flow of coolant and heat transfer in the VHTR. They will also help in modifying the reactor design to reduce turbulence and extend the lifespan of the reactor's components, as well as improve the efficiency and reduce the cost of maintaining nuclear reactors.

REACTOR



Coolant flow through the VHTR's lower plenum



The VHTR design offers more economy, efficiency, safety and security against nuclear proliferation than earlier-generation reactors. U.S. Department of Energy Nuclear Energy Research Advisory Committee [Public domain], via Wikimedia Commons.

PSC LOOKBACK: THE TURBULENCE UNDERNEATH

No one would mistake the Sun—a fiery furnace that bombards the Earth with about as much energy as the largest thermonuclear bomb ever detonated every second— for a peaceful place. But the Sun presents us with a surprisingly calm face all the same. Its smooth, "cellular" surface belies the churning, turbulent action that theorists in the early 1990s said *had* to lie underneath. Juri Toomre and colleagues at the University of Colorado used PSC's Cray C90 to simulate a three-dimensional section of the solar convection zone, finding that the smooth surface emerged as a result of turbulent flows from below compressing gases above. The work suggested that turbulence is the major mode of energy transport in the Sun's interior, explaining a number of otherwise puzzling observations and earning the researchers' paper a coveted cover image in the journal *Science* in 1995.

Spotting the Signal

Blacklight Helps Discover Sparse Clues for Stock Performance

WHY IT'S IMPORTANT:

From the smallest retirement investors to the highest-tech hedge fund manager on Wall Street, we're all trying to anticipate the market. Should I buy stock A? If I do, when should I sell it? Finance researchers have long suspected the most talented traders use their experience and intuition to pick out market details—performance of key stocks, for example—that help them guess whether a particular stock of interest is likely to go up or down. But how do they identify such "predictor" stocks, and how few predictors do you need? Or could it be the idea of predictor stocks is a myth?

HOW PSC HELPED:

University of Illinois finance professors Adam D. Clark-Joseph, Alex Chinco and Mao Ye teamed up to study whether "sparse signals"—small numbers of predictor stocks—actually exist in the markets. With help from PSC staff, Clark-Joseph and his co-authors leveraged the large shared memory and numbers of processors of PSC's Blacklight to test this idea using nine months of minute-by-minute New York Stock Exchange returns for thousands of stocks. Out of the roughly 2,200 stocks they analyzed each month, the researchers could statistically identify about 12 predictor stocks at a time that helped forecast the performance of a given "target" stock. But these predictors were only briefly relevant to the target, with 90 percent of predictor stocks

Our intuition is that a lot of these signals are appearing and [experienced traders] know what they mean, but they're too fleeting to get picked up by most traditional statistical methods…Where the supercomputer comes in is that we were able to analyze the…data among ome 7,600 variables relatively quickly.

—Adam Daniel Clark-Joseph, University of Illinois

remaining relevant for four minutes or less. Adding the fleeting predictor-stock information improved forecasts by a factor of nearly 1.5 compared with standard forecasting methods that target a stock's historical performance alone. Even more interesting, more or less the same group of predictor stocks were relevant to numerous other stocks at a given moment.

Future work, including using PSC's Bridges system, will explore the deeper structure that could underlie these fleeting relationships between seemingly unrelated stocks.

PSC staff helped this research with the support of several programs within the NSF's XSEDE network of supercomputing centers.

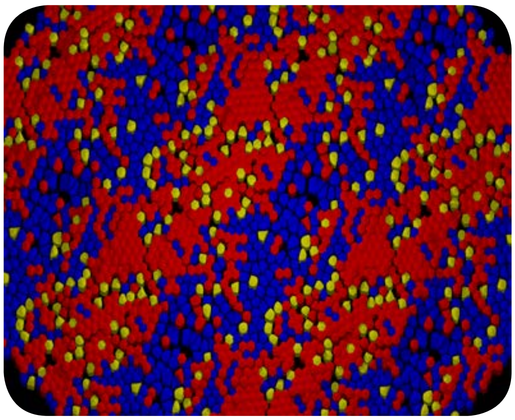
Extended Collaborative Support Service Novel and Innovative Projects Program Campus Champions Program.

PSC LOOKBACK: SECURING PRIVACY

Over roughly the past decade, PSC's systems have pioneered the use of massive, shared memory to allow researchers to make many calculations against vast datasets in parallel. In 2009, work by Alessandro Acquisti and collaborators at Carnegie Mellon University using PSC's Pople system—PSC's first large-memory supercomputer—sounded a clarion warning. Publically available information on the Internet, they found, can be used in some cases to predict individuals' social security numbers. Pople's combination of many processors and large shared memory allowed the team to run many variations of their alorithms against the same data in parallel. The fact that many of us may be vulnerable to identity theft without actually posting sensitive information to the Web gained national attention.

Island Time

Extended Simulation on Anton Shows How Cell-Surface Molecules Cluster



In this view of the simulated cell membrane from above, added cholesterol (yellow) causes other lipids to clump into "rafts."

WHY IT'S IMPORTANT:

Virtually every process in human health and disease relies on signals getting across cell membranes—the flexible "bags" that enclose the contents of all our cells. The immune system's response to infection, the spread of cancer cells and the communication between nerve cells that underlies brain function and mental illness are among the many life processes that involve such signals.

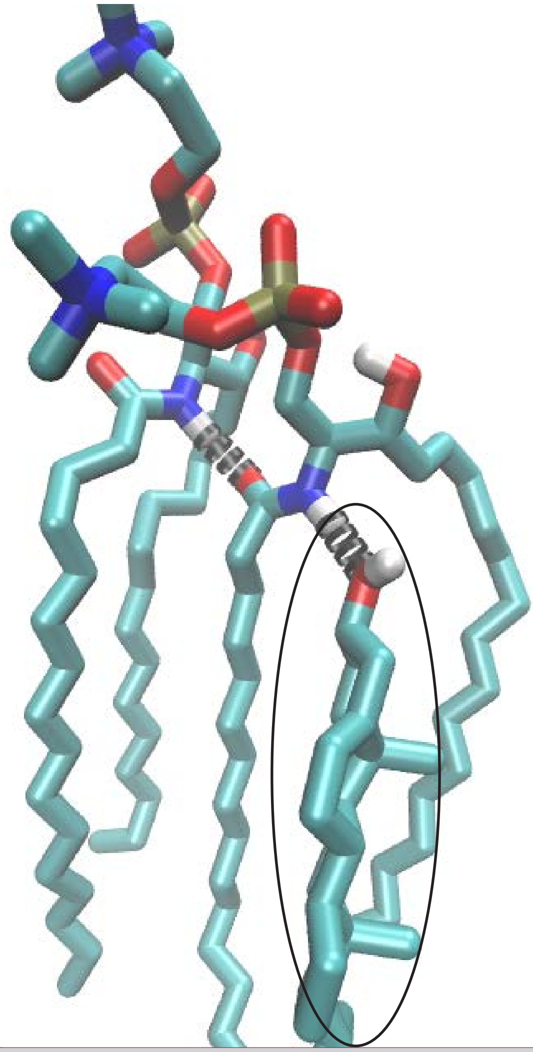
Many biochemists believe cholesterol plays a role in cross-membrane signals by limiting how proteins and other molecules in the membrane can move, causing them to clump into "islands." Other experts have pointed out that we have very little direct evidence for this membrane clustering. Understanding whether this phenomenon actually happens and how it changes signals can give us clues about how to intervene in a way that reverses diseases involving cell-surface signaling.

> When you think of cholesterol, most people think of heart attacks, statins, and so on. But it turns out your [healthy] cell membranes are composed mostly of cholesterol, by percentage of molecules. What's all this cholesterol doing there?
>
> —Edward Lyman, University of Delaware

HOW PSC HELPED:

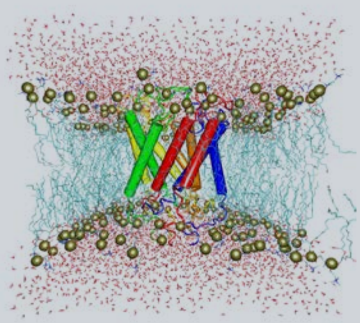
Edward Lyman and colleagues at the University of Delaware used the D.E. Shaw Research Anton supercomputer hosted at PSC to see how changing the composition of cholesterol and other fatty molecules, or lipids, in a simulated membrane affects the movement of these molecules. They simulated the membrane for close to 100 microseconds—a very long time in such molecular dynamics simulations, and only possible thanks to Anton's specialized architecture. In the simulations, adding cholesterol did in fact cause the other lipids in the membrane to sort and form islands that would likely limit protein movement. This result, featured on the cover of the *Biophysical Journal*, explained previous experimental findings and will help guide new experiments aimed at understanding how clustering affects signal proteins.

In this view of cholesterol (circled) interacting with other lipids, the cholesterol causes the other lipids to tilt, disrupting their interactions and causing rafts to form.



PSC LOOKBACK: HOW WATER DITCHES ITS DATE

One of the most notable accomplishments that scientists have achieved using PSC's resources was the early-2000s simulation of the aquaporin protein by Klaus Schulten and colleagues at the University of Illinois, using PSC's LeMieux supercomputer. The group solved a mystery about how aquaporin, a channel in the cell membrane, could speed water across the membrane at a blistering pace without allowing water's usual consort, the hydrogen ion, through as well. A pirouette at the center of the channel, they found, allowed water molecules to ditch their steady date, giving aquaporin its remarkable specificity. The work made it into the prestigious journal *Science* and was cited by Peter Agre, the discoverer of aquaporin, in his 2003 Nobel Prize acceptance speech.



The Aquaporin protein.

Kindling the Kindlers

Carnegie Science Award Recognizes PSC STEM Education Programs

PSC has received the 2016 Carnegie Science Award for Leadership in STEM Education. The award recognized three PSC staff members: Pallavi Ishwad, Hugh Nicholas and Alexander Ropelewski.

The Carnegie Science Center, a leader in science and technology education, attracts more than 700,000 visitors each year and is one of the top science museums in the country. The Center established its annual awards in 1997 to recognize individuals and organizations in western Pennsylvania that have made outstanding contributions in science and technology. The STEM (Science, Technology, Engineering and Math) Education Award "recognizes an individual, team or organization that demonstrates leadership in building literacy in science, technology, engineering and math."

"We're thrilled to see Hugh, Alex and Pallavi's work recognized by the STEM Leadership award," says Cheryl Begandy, PSC's director of communications and industrial relations. "Their work does more than educate students in bioinformatics; by increasing awareness of this discipline, it helps prepare young people—and our workforce—for the 21st Century."

PSC's Bioinformatics Education Team has developed and implemented bioinformatics curriculums at the graduate, undergraduate and high school levels.

BRIDGING THE BIOINFORMATICS GAP

The PSC STEM team began training researchers in bioinformatics—the use of computational methods to analyze and interpret biological data—as early as 1987. In 2001, they initiated a bioinformatics program at three minority-serving institutions (MSIs) through the NIH-funded MARC (Minority Access to Research Careers) program.

"A major obstacle to preparing minority students for the STEM workforce is lack of access to relevant courses, degree


PSC's 2015 MARC interns

programs and research opportunities at MSIs," says Ropelewski, a longstanding MARC instructor and director of PSC's Biomedical Applications Group. "These institutions often don't have the resources to offer students access to advanced instrumentation and computational resources needed for bioinformatics work."

The MARC program was designed to remedy this, focusing on creating both curriculum and research opportunities for faculty and students at MSIs—schools at which the undergraduate student body consists of over 50 percent minority students. To date, PSC's MARC program has provided bioinformatics training and mentoring to hundreds of students and researchers at MSIs and helped dozens of institutions across the country develop bioinformatics classes or full curriculums. These schools include: North Carolina Central University; Howard University; University of Puerto Rico Medical Sciences Campus; Morgan State University; University of Puerto Rico at Mayaguez; University of Texas at El Paso; University of Texas at San Antonio; Johnson C. Smith University; North Carolina A&T State University; Langston University; Tennessee State University; Jackson State University; and Universidad Metropolitana, Puerto Rico.

STARTING HIGH SCHOOL STUDENTS ON THE STEM PATH

In 2007, PSC staff adapted MARC for a high school audience through the BEST (Bioinformatics Education for STudents) Program. Today 10 area high schools are offering a full or integrated bioinformatics curriculum thanks to BEST. Training for teachers in the BEST program involves a summer workshop. The High School curriculum features an in-depth introduction to bioinformatics with real-life emphasis on 21st-century career awareness and readiness.

"Like MARC, BEST provides an innovative approach for acquiring 21st century STEM skills at the high school level," says Ishwad, who leads the BEST program. "We developed the BEST high school curriculum with a multidisciplinary focus, integrating concepts from biology, chemistry, physics, mathematics and computing. This is unique among high school STEM offerings, but more accurately reflects the integrative nature of the STEM disciplines."

The BEST curriculum has to date been adopted in 10 western Pennsylvania schools: Pittsburgh Public School District—SciTech Academy; Our Lady of the Sacred Heart High School, Coraopolis, Pa.; Oakland Catholic High School, Pittsburgh, Pa.; PA Cyber Charter School, Midland, Pa.; Ellwood City Area School District, Ellwood City, Pa.; Central Catholic High School, Pittsburgh, Pa.; Plum Borough School District, Plum, Pa.; Fort Cherry School District, McDonald, Pa.; Frazier High School, Perryopolis, Pa.; and Winchester-Thurston High School, Pittsburgh, Pa.

The PSC team received their award at a ceremony at Pittsburgh's Carnegie Music Hall in May 2016.

PSC LOOKBACK: ALUMNI CONTINUE TO KINDLE

PSC's STEM education efforts have enjoyed a number of organizational achievements over the year. But the true value of these programs can be seen in the accomplishments of their alumni, who include:

Cynthia Jeffries, who participated in the MARC program in 2007 as a graduate student at Jackson State University. She presented her PSC work at the International Society for Computational Biology meeting in Toronto in 2008. Today she's a scientist with Oak Ridge National Laboratory's bioinformatics group.

Danielle Auth, who was a high school student at BEST-participant Our Lady of the Sacred Heart High School. She presented her PSC internship work at the TeraGrid '11 conference in Salt Lake City. In 2015, she graduated from Gannon University with bioinformatics as her major.

Amanda Sciorillo, who participated in a BEST bioinformatics course through the Pennsylvania Cyber Charter School. Her experience helped her get an internship at the Children's Hospital of Philadelphia Research Center. This year she entered Marymount University as a Clare Boothe Luce Program scholar.

Luis Vazquez-Quiñones, who took the first MARC bioinformatics class taught at the University of Puerto Rico Medical Sciences Campus in 2001. He became a professor at the Universidad Metropolitana in Puerto Rico, starting a bioinformatics program there in 2009.

A Broader View of Life's Instants

Anton 2 Will Increase Speed and Size of Molecular Simulations

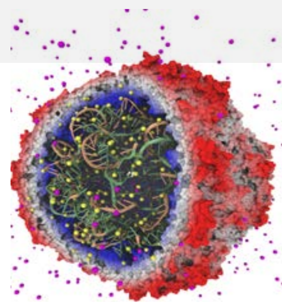
When it comes to simulating the motions of large biomolecules, the D. E. Shaw Research (DESRES) Anton supercomputer hosted at PSC has been in a class by itself. It's allowed researchers to identify unexpected but crucial molecular contortions that underlie previously puzzling properties of medically and biologically important molecules. Anton has powered 277 simulation projects by 127 different principal investigators across the U.S., resulting in more than 120 peer-reviewed research papers. Three of these studies appeared in the scientific journal *Nature*, one of the international scientific community's premier publications.

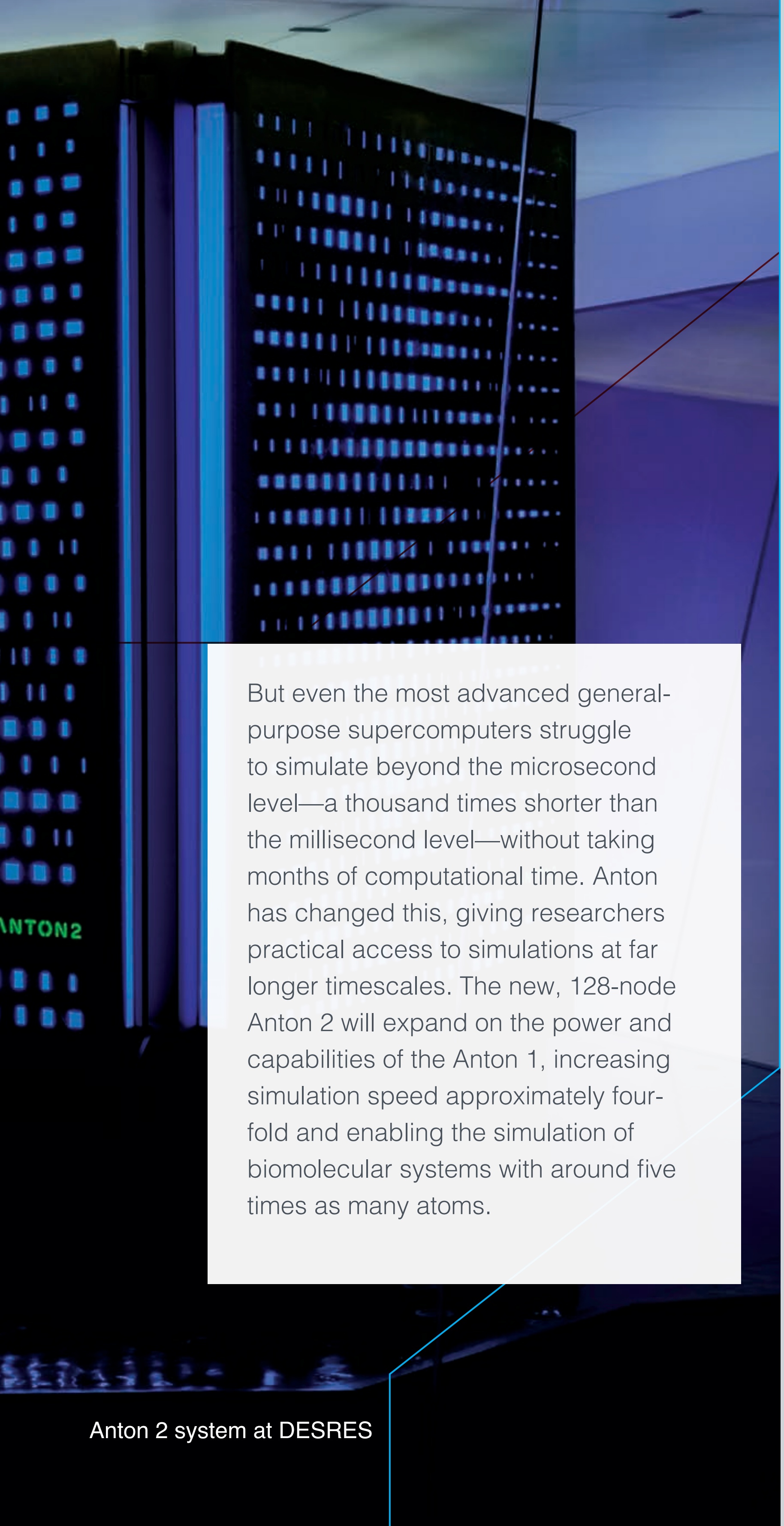
Now DESRES will provide a next-generation Anton 2 supercomputer to PSC without cost for use by the biomedical research community. In addition, a $1.8-million National Institutes of Health grant to PSC will provide operational funding for the new system, as well as extend the operation of the Anton 1 supercomputer currently at PSC until the new Anton 2 is deployed, expected in the Fall of 2016.

A specialized system for modeling the function and dynamics of biomolecules, the Anton 2 machine at PSC will be the only one of its kind publicly available to U.S. scientists.

"Many life processes important to understanding the molecular basis of cellular events occur over timescales exceeding a millisecond in duration," says PSC's Phil Blood, principal investigator of the new grant. "Anton 2's performance for molecular simulation will exceed that of current general-purpose supercomputing systems by orders of magnitude, enabling the study of biological processes not otherwise possible and offering new possibilities in drug discovery and development."

Molecular dynamics simulations can provide insights into the behavior of proteins, cell membranes, nucleic acids, and other molecules at the atomic scale.



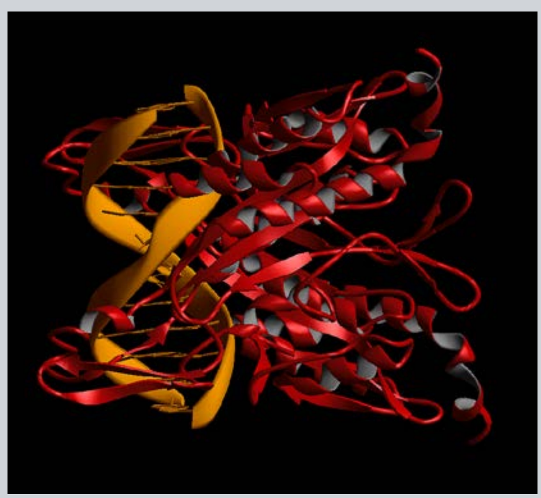


But even the most advanced general-purpose supercomputers struggle to simulate beyond the microsecond level—a thousand times shorter than the millisecond level—without taking months of computational time. Anton has changed this, giving researchers practical access to simulations at far longer timescales. The new, 128-node Anton 2 will expand on the power and capabilities of the Anton 1, increasing simulation speed approximately four-fold and enabling the simulation of biomolecular systems with around five times as many atoms.

Anton 2 system at DESRES

PSC LOOKBACK:
ECOR1 AND ITS KINKS

In molecular biology circles, the EcoR1 is a superstar. A humble DNA-cutting enzyme from the human gut bacterium E. coli, its unique ability to mix and match DNA sequences at will, spurred the genetic engineering industry. EcoR1 provided John Rosenberg and colleagues at the University of Pittsburgh with a window into how proteins interact with DNA, either to cut it as EcoR1 does or to affect the activity of genes as the critical DNA regulatory proteins do. EcoR1 proved a fruitful exemplar of protein-DNA interaction; in first-of-its-kind work with PSC's Cray Y-MP in the early 1990s, Rosenberg's team showed how merely binding to EcoR1 created a "kink" in the DNA helix, demonstrating that proteins could make surprisingly large changes in DNA's shape.



The EcoR1 protein

Guarding Openness

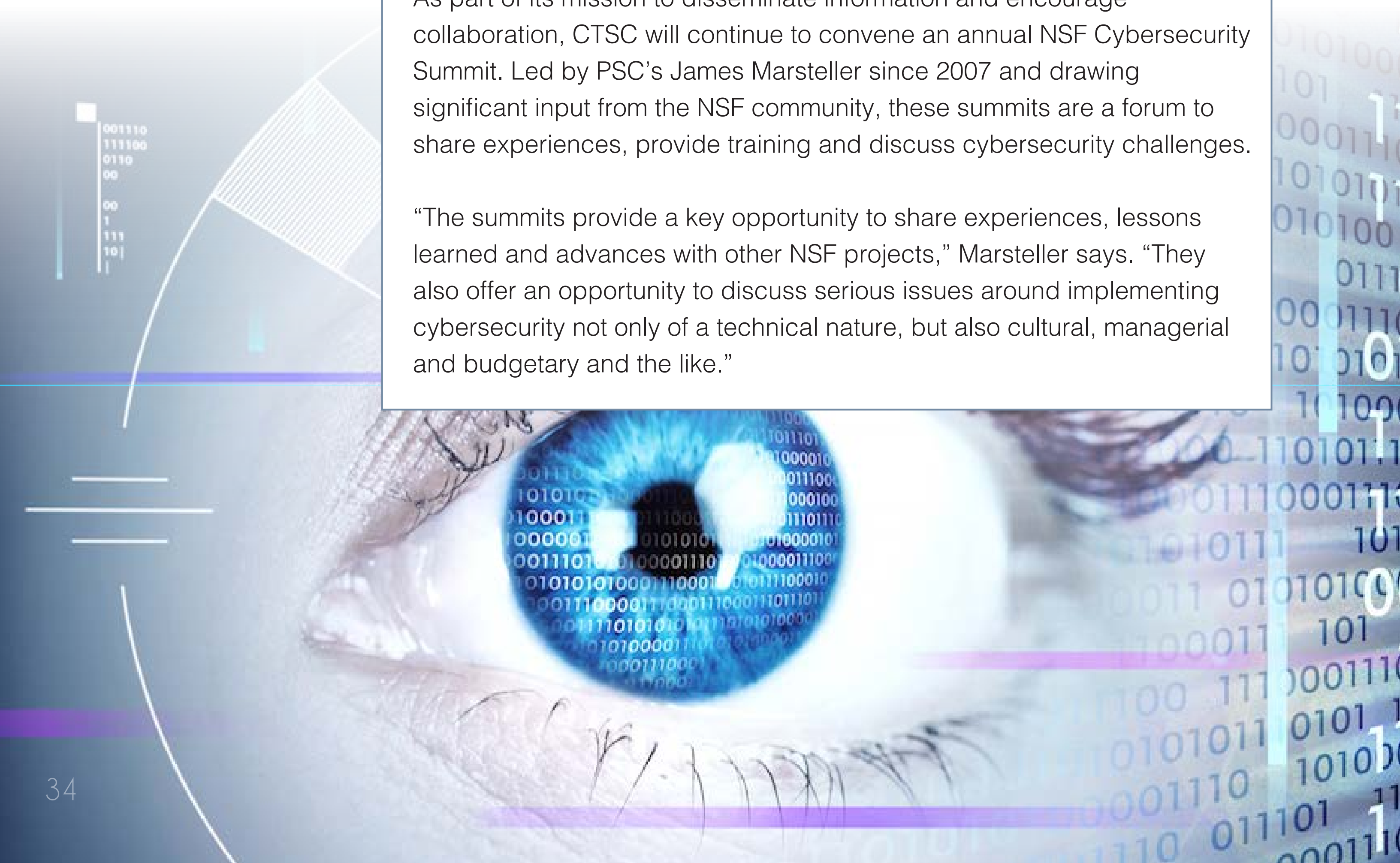
$5-million Grant Establishes Cybersecurity Center of Excellence

When it comes to cybersecurity, academic research organizations have a difficult mission. An open flow of ideas and data is critical to sparking scientific advances. Such an open culture though, by definition, poses greater security risks.

Thanks to a $5-million National Science Foundation (NSF) grant, PSC and three allied institutions will establish the NSF Cybersecurity Center of Excellence (CTSC) to provide expertise to research organizations to develop, share and enact cybersecurity measures to help safeguard the security of more than $7 billion in research funded by the NSF. The grant was awarded to Indiana University, the National Center for Supercomputing Applications, PSC and the University of Wisconsin-Madison.

As part of its mission to disseminate information and encourage collaboration, CTSC will continue to convene an annual NSF Cybersecurity Summit. Led by PSC's James Marsteller since 2007 and drawing significant input from the NSF community, these summits are a forum to share experiences, provide training and discuss cybersecurity challenges.

"The summits provide a key opportunity to share experiences, lessons learned and advances with other NSF projects," Marsteller says. "They also offer an opportunity to discuss serious issues around implementing cybersecurity not only of a technical nature, but also cultural, managerial and budgetary and the like."



Opening the Net

Grant Continues PSC Work to Improve Network Reliability

The TCP/IP code that underlies the Internet is near-miraculous, giving us a system that works with remarkable effectiveness, reliability and ease.

Except, of course, when it doesn't. One drawback is it works invisibly. Even network administrators can't open its hood and see what's going on when things go wrong.

A new, $300,000 National Science Foundation grant will take PSC's ongoing project to "open up" TCP/IP to the next phase. The grant will fund development of Test Rig 2.0, an automated system to provide network administrators with information about faulty connections, greatly speeding network diagnosis and repair.

"When researchers encounter network problems they naturally reach out to network engineers," says PSC's Chris Rapier, principal investigator in the grant. "However, the engineers have to rely on the user to provide them with enough information to properly diagnose the problem. This often means multiple rounds of email, phone calls and tests."

TestRig will allow users to test their connections and send the results automatically, while maintaining their data security. "The goal is to make this as easy as possible" for users and administrators, Rapier adds.

TestRig joins a growing stable of PSC-led innovations in network administration that includes DANCES, which will allow large data users to schedule data transfers to minimize network disruption, and Web10G, an intuitive dashboard to help users identify data slowdowns so they know when to ask for help.

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

We gratefully acknowledge significant support from:

The National Science Foundation
The National Institutes of Health
The National Oceanographic and Atmospheric Administration
The U. S. Department of Defense
The U. S. Department of Energy
The Commonwealth of Pennsylvania
The Bill and Melinda Gates Foundation
The Buhl Foundation
D. E. Shaw Research
Hewlett Packard Enterprise
Intel
Silicon Graphics, Inc.
Supermicro
Uber Advanced Technologies Center
The U.S. Agency for International Development

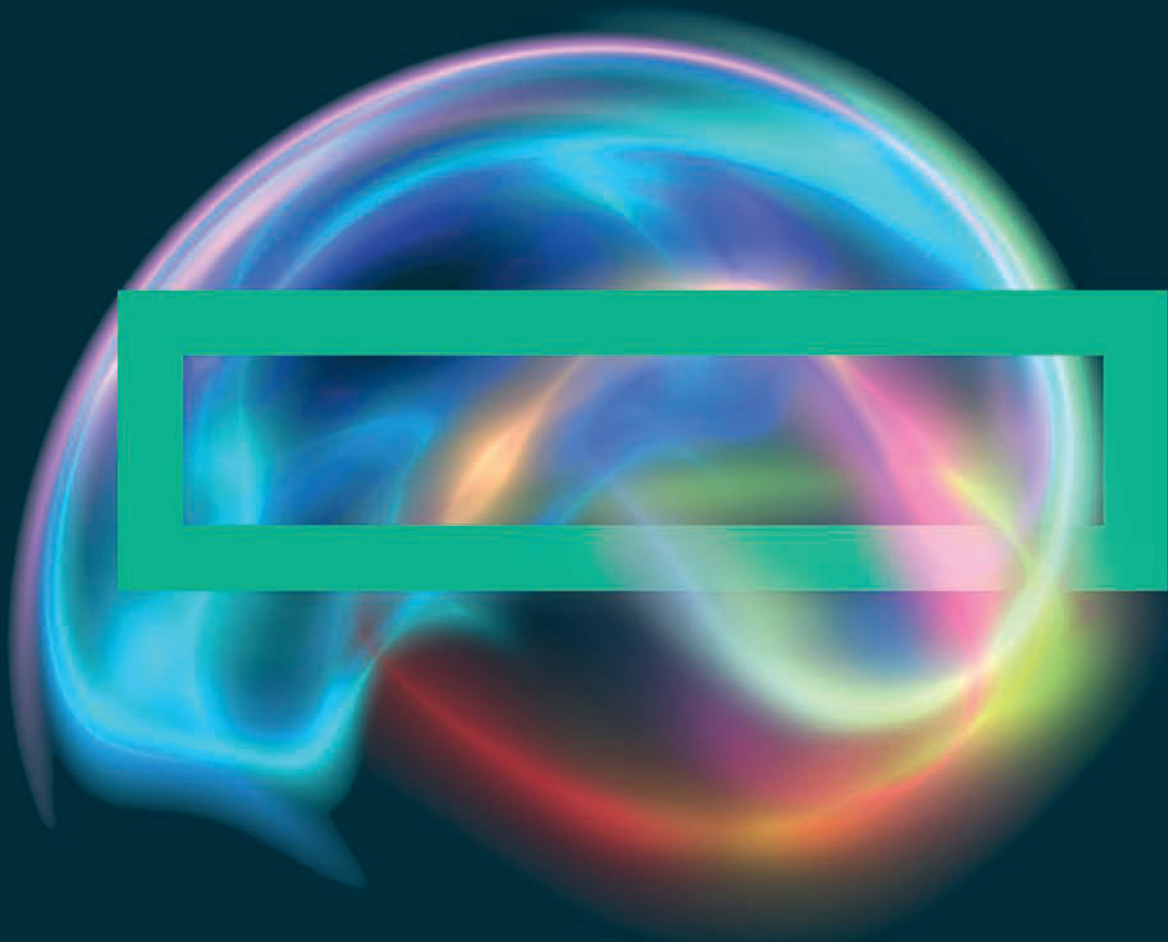
Project Coordinator: Cheryl Begandy
Managing Editor/Production Manager: Vivian Benton
Senior Science Writer: Ken Chiacchia
Design: Shandra Williams

Printing: heeter

Printed on Sappy McCoy paper, a premium sheet with 10 percent post-consumer waste fiber with vegetable-based inks

Accelerating supercomputing

Hewlett Packard Enterprise is solving the toughest compute challenges. Together with Intel,® we build more than twice as many of the world's most powerful supercomputers as the nearest competitor.[1]



Accelerating next

Hewlett Packard Enterprise

intel®

Learn more at hpe.com/info/hpc

[1]Based on the Top500 list (Top500.org), published November 17, 2015.