GATEWAY to DISCOVERY
Tackling Today's Grand Research Challenges

SDSC
ANNUAL REPORT FY2015/16

Where Stars Converge

Fiscal year 2015-16 was a watershed for the San Diego Supercomputer Center (SDSC). We marked our 30th anniversary with the successful launch of *Comet*, our latest supercomputer, funded by the National Science Foundation (NSF). Today, *Comet* is well on its way to meeting a key goal of serving the "long tail of science", providing high-performance research computing to at least 10,000 academic users, advancing discovery across research fields that range from astrophysics and alternative energy solutions to genomics, finance, and the social sciences.

As principal investigator of the *Comet* program, I could not be more proud of how this supercomputer already has opened new vistas of discovery to such a broad community of researchers. This Annual Report provides several examples of how *Comet* is spreading its computational prowess across the nation, not only to traditional big-time users, but to the vast majority of researchers who otherwise would not have access to advanced computing.

Comet is only one of many shining examples of excellence here at SDSC. As a long-time astrophysicist, I'm inclined to use the night sky as a metaphor for things of importance to me. For SDSC, I might compare *Comet* to looking at the sky on a clear night and first seeing only the brightest flashes of light... and then one's eyes adjust. Before long, the viewer sees other sparkling objects and may even identify some of the more familiar patterns as stars when they converge to form constellations and galaxies. It's the seeming relationships among all these brilliant gems which form something larger than just the sum of each twinkle in the night sky—it's a universe that's at once diverse and yet unified. To me, that's SDSC.

The convergence of stars is also a good metaphor for our new strategic planning initiative, which we undertook during the 2015-16 fiscal year. Our new three-to-five year strategic plan builds directly on SDSC's core strengths as a national resource for high-performance computing and interactions with the computational science community to develop new capabilities and increase value to UC San Diego, the University of California system, the State of California, and the national research community. Our strategic priorities below capture the essence of this plan:

Versatile Computing

This priority builds on SDSC's core strengths in advanced computing, which is now essential to nearly every field of research, with demand for computing far outstripping availability of resources. We are guided by some key trends as we pioneer the development of more versatile computing systems that are easily accessed by an ever-expanding user base:

- Research communities of users increasingly are turning to science gateways, or application-specific, web-based interfaces for accessing supercomputers that let users focus on their science, not the systems.

- **High-throughput computing (HTC)** as a critical tool in the analysis of data from large-scale experiments, notably ones being conducted in the area of astrophysics and genomics.

- **Cloud computing**, which has become an important element of an overall research computing ecosystem. SDSC has a leadership role in helping the research community understand and use this resource.

Data Science and Engineering

Data-driven science has emerged as the single most transformational force in advanced computing. Large scientific experiments, gene sequencers, machine and human sensors, and the Internet of Things (IoT) present those responsible for building cyberinfrastructure tools with major technological, application and workforce challenges in developing end-to-end solutions for a wide range of scientific research questions. This strategic priority, to include streaming data and other initiatives, aligns with national research priorities such as 'smart' cities and communities (metro science); developing new research tools; wearable biomedical devices; the IoT; and research innovations at the nexus of food, energy, and water systems.

Life Sciences Computing

SDSC is dedicated to fostering new collaborations with the life sciences research community to develop advanced research platforms and services. This priority aligns with national efforts such as the NSF's 'Understanding the Brain' initiative. It also supports more efficient inter-campus collaborations such as UC San Diego's strategic priority of "exploring the basis of human knowledge and creativity", while providing UC-wide economies of scale.

SDSC's Vision Statement
To deliver lasting impact across the greater scientific community by creating end-to-end computational and data solutions to meet the biggest research challenges of our time.

In closing, I am confident that this Annual Report conveys the sheer diversity of SDSC's interests and accomplishments. They reflect how SDSC continues to position itself as a leader in today's world of research, which now requires a seamless integration of advanced computation, data analysis, and of course, more collaboration than ever. Converging the stars, if you will.

Michael L. Norman
SDSC Director

New Appointments…

SDSC made two new appointments to its management staff during the 2015-16 fiscal year.

Shawn Strande, who worked for both SDSC and UC San Diego in various capacities since 1992, was named the Center's new deputy director, succeeding Richard L. Moore, who retired after 13 years with SDSC. Strande rejoined SDSC in August 2015, returning from the National Center for Atmospheric Research (NCAR), where since mid-2013 he was responsible for the center's computational infrastructure. Prior to joining NCAR, Strande was project manager for SDSC's data-intensive *Gordon* supercomputer, the result of a $20 million NSF grant awarded in late 2009.

"Shawn has a proven track record in coordinating large-scale procurement and proposal efforts, as well as fostering collaborations across campus, the entire UC system, and among our industry partners, said SDSC Director Michael Norman.

In October 2015, Karen R. Flammer joined SDSC as Director of Education at SDSC, succeeding Diane Baxter, who retired in November of that year. Flammer, a research physicist at UC San Diego who obtained her Ph.D. in Space Physics from the university in 1988, joins SDSC following a wide array of accomplishments throughout the education field during the past 20 years, including experience as the senior director of large-scale K-16 STEM (science, technology, engineering, and math) outreach and professional development programs. In 2001, she co-founded, along with Sally Ride and Tam O'Shaughnessy, Sally Ride Science, a STEM education technology company dedicated to inspiring young people by promoting STEM literacy.

"We are delighted that Karen has joined SDSC to continue to guide what is already an excellent and very diverse education program nurtured by Diane Baxter during the past 10 years," Norman said.

Yifeng Cui

SDSC

π Person of the Year

S DSC Researcher Yifeng Cui was recently named SDSC's third "Pi Person of the Year" for his research on computationally challenging problems in seismic research. Named after the π symbol, this award recognizes researchers who have one 'leg' in a science domain and the other in cyberinfrastructure technology.

Cui, who joined SDSC in 2001, is the founding director of the High Performance GeoComputing Laboratory at SDSC, and an adjunct professor at San Diego State University. He is an internationally recognized expert in the development of large-scale earthquake simulation software and helped to establish the Southern California Earthquake Center (SCEC) as a world leader in advancing high-performance computing activities in earthquake system science. Cui also directs the new Intel Parallel Computing Center at SDSC and has presented more than 40 invited talks at international conferences and prominent venues.

Cui is the recipient of NVIDIA Global Impact Award in 2015, the IDC HPC Innovation Excellence Award in 2013, the SciDAC OASCR Awards in 2009 and 2011, and an ACM Gordon Bell finalist in 2010 for parallel computing research. He received his B.S. in Meteorology from Nanjing Institute of Meteorology, and a Ph.D. in Hydrology from the University of Freiburg.

What is the thrust of activity at the High Performance GeoComputing Laboratory?

Cui: Modeling of earthquake dynamics is one of the most challenging computational problems in science today. State-of-the-art large-scale ground motion simulations require the most powerful supercomputers to solve equations that involve hundreds of billion elements, or unknowns. To be accurate, simulations must span an enormous range of geographic scale, from just a few meters near an earthquake's source to hundreds of miles across a region. We also need time scales from hundredths of a second to capture the higher frequencies.

Earlier this year Intel chose to open its second Parallel Computing Center at SDSC, this time to focus on seismic research. Can you describe some of the benefits of this new program?

Cui: Being part of this Intel program gives us the opportunity to prepare our earthquake computational models for next-generation architectures that can help researchers find ways to minimize potential loss of life, injuries, and property damage. Our strong partnership with Intel helps us to substantially broaden the frequency band covered by such detailed simulations, allowing us to predict ground motions at frequencies relevant for common dwellings in the future. The program continues a very productive and longstanding interdisciplinary collaboration between SDSC and the Southern California Earthquake Center, one of the largest open research collaborations in geoscience. In addition to UC San Diego, the new Intel center includes researchers from the University of Southern California, San Diego State University, and the University of California Riverside.

How has the accuracy and effectiveness of earthquake simulations improved over the years, and what are the main drivers for such improvements?

Cui: Producing realistic seismograms at high frequencies requires advanced physics. The simulations of the recent ShakeOut earthquake scenario, for example, have shown that nonlinearity could reduce the earlier predictions of long period (0 – 0.5 Hz) ground motions in the Los Angeles basin by 30 – 70%. We cannot predict earthquakes, but realistic, physics-based earthquake ground motion simulations can help answer questions such as what will be the shaking intensity in the Los Angeles basin from a magnitude 7.8 earthquake on the southern San Andreas fault. The full three-dimensional treatment of seismic wave propagation allows us to identify the ruptures that dominate the hazard at a particular site, and to retrieve the actual seismograms, which can then be used to drive full-physics engineering models. The 3D modeling brings the computational challenges of physics-based probabilistic seismic hazard analysis into sharp focus. Even in today's world, where supercomputers can perform one quadrillion calculations per second and sift through data at speeds never thought of, earthquake simulations still require orders of magnitude more compute power, especially for higher-frequency seismic events, along with steady advances in physics research to improve the accuracy of those simulations..

What are some of your key goals in seismic research?

Cui: One goal is to create more accurate simulations of earthquakes that occur at higher frequencies. We can run 2-Hz wall-to-wall scenarios on a southern California region with today's computing capabilities, but our goal is to include frequencies as high as 10-Hz. That is when our research becomes relevant to things such as designing and engineering more resilient buildings. Pipes, electrical systems, and critical infrastructures such as those in hospitals or nuclear power plants can be sensitive to higher frequencies. California comprises 75 percent of the nation's long-term earthquake risk. Our practical goal is to translate the advanced modeling capabilities to improve operational earthquake forecasting, and help identify effective ways to reduce seismic risk.

What do you for relaxation or fun when you need a seismic-free day?

Cui: I enjoy hiking and bicycling. I'm fortunate to be in San Diego, a perfect place for outdoor activities!
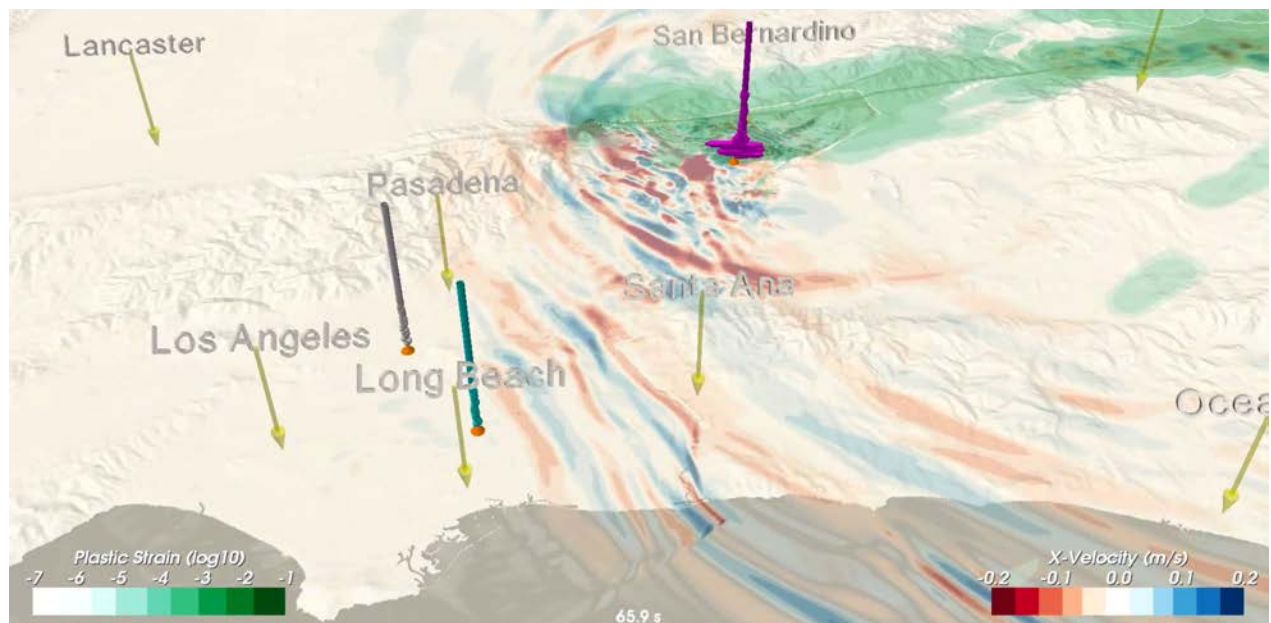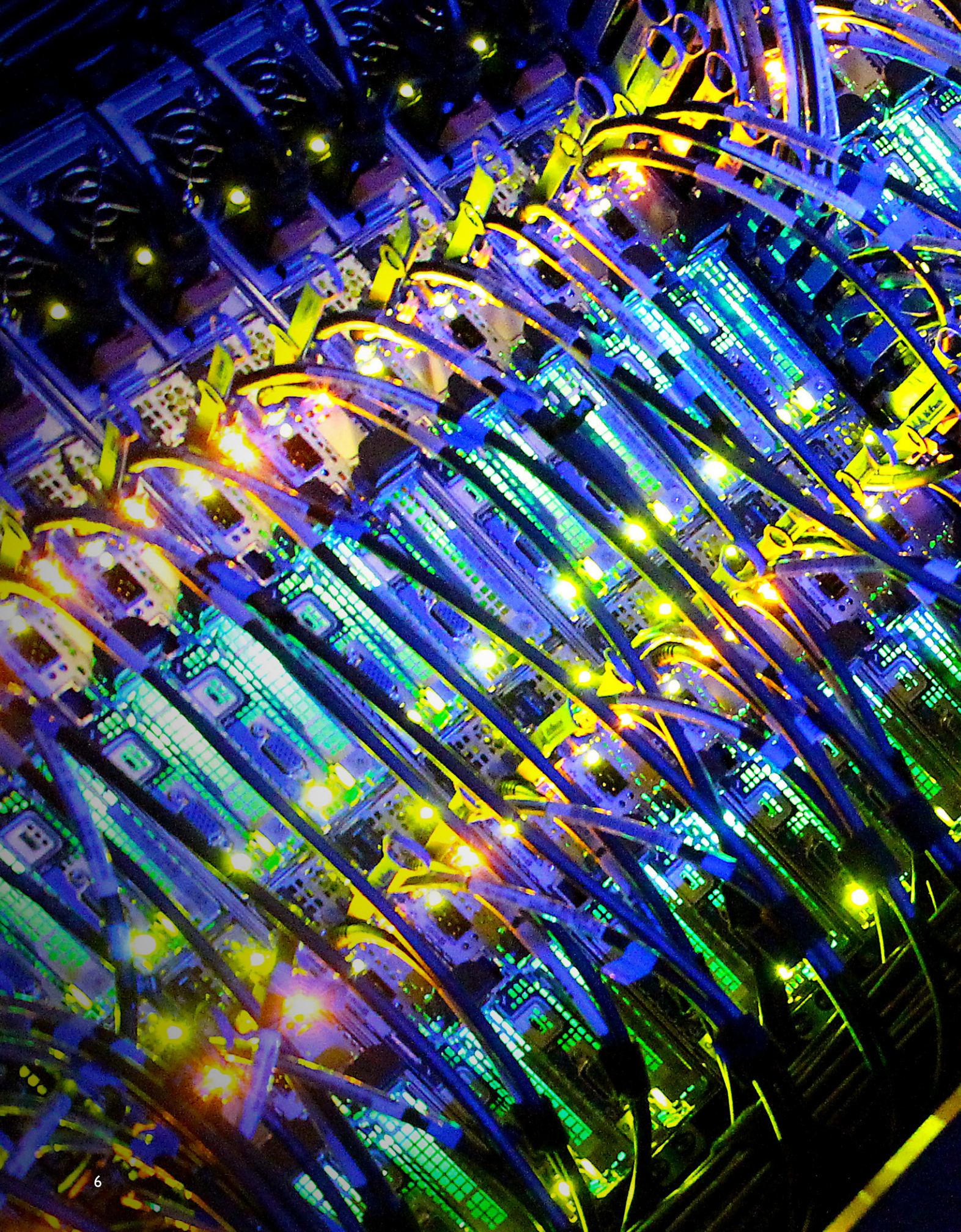
GATEWAYS
TO DISCOVERY:
SDSC's Computational Resources

SDSC is home to a wide array of advanced computational, storage, and networking resources, all aimed at accelerating scientific discovery at the local, state, national, and even global levels through collaborations across academia, industry, and government.

SDSC marked its 30th anniversary in late 2015, using the event to formally launch its *Comet* supercomputer, the result of a National Science Foundation (NSF) grant valued at about $24 million. *Comet* joined *Gordon*, a data-intensive supercomputer that came online in 2012, also as the result of an NSF grant.

While SDSC is committed to remaining at the forefront of high-performance computing, the Center has also focused on fostering new or expanded collaborations in the area of large data management and expertise. Moreover, our ever-increasing complex technological world has ushered in a new era of rapid scientific investigation, fueled by innovative ideas to tackle the grand research challenges of our time.

"Today, many fields of data-intensive research not only require high-performance computing, but the ability to sift, analyze, and share the massive tidal wave of data overwhelming researchers across multiple research domains and vast geographic regions," said SDSC Director Michael Norman. "To keep up with that accelerating pace of scientific investigation, we as a nation require a next-generation cyberinfrastructure that supports effective and efficient collaborations that will drive us toward grand challenge discoveries."

(left) SDSC's *Comet* supercomputer is capable of about two quadriliion calculations per second.

(right) Inside SDSC's 19,000 square foot datacenter.
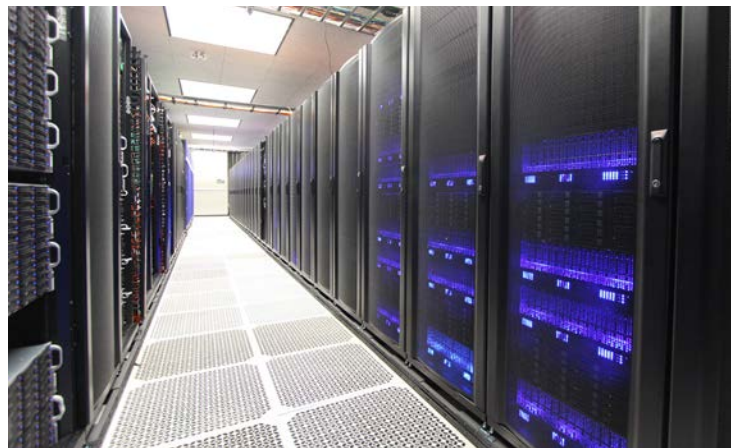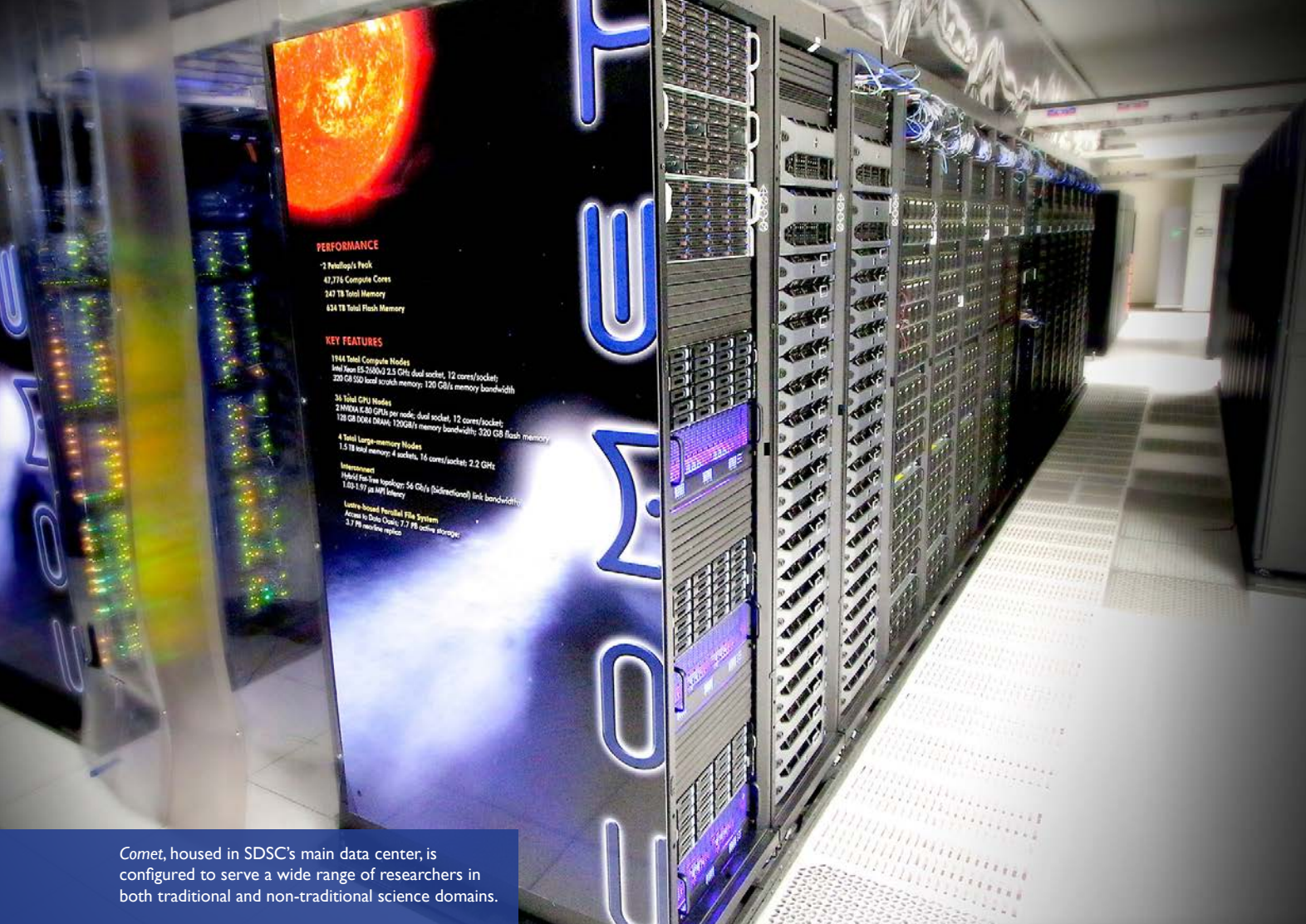
Comet, housed in SDSC's main data center, is configured to serve a wide range of researchers in both traditional and non-traditional science domains.

Comet

HPC for the 99 Percent

Comet is SDSC's latest high-performance computing (HPC) resource, a petascale supercomputer designed to transform advanced scientific computing by expanding access and capacity among traditional as well as non-traditional research domains. *Comet* is capable of an overall peak performance of two petaflops—that's two quadrillion calculations per second.

Comet joins SDSC's *Gordon* supercomputer as another key resource within the NSF's XSEDE (Extreme Science and Engineering Discovery Environment) program, which comprises the most advanced collection of integrated digital resources and services in the world.

"*Comet* represents what we like to call HPC for the 99 percent," said Norman, the program's principal investigator. "It's about providing high-performance computing to a much larger research community, while meeting the needs of underserved researchers in domains which have not traditionally relied on supercomputers to help solve problems."

Science Gateways Leader

Comet is already blazing new paths of discovery, thanks in part to its role as a providing an assortment of 'science gateways.' Simply described, science gateways provide researchers web browser access to applications and data used by specific research communities. Gateways make it possible to run the available applications on supercomputers such as *Comet* so results are achieved quickly, even with large data sets.

NSF Awards $15 Million to Create Science Gateways Community Institute Under a Multi-Partner Project led by SDSC. Read more on page 40.

Browser access offered by gateways also allows researchers to focus on their scientific problem without having to learn the details of how supercomputers work or how to access and organize the data needed.

"It's possible to support gateways across many disciplines because of the variety of hardware and support for complex,

customized software environments on *Comet*," according to Nancy Wilkins-Diehr, an SDSC associate director and co-director of XSEDE's Extended Collaborative Support Services. "This is a great benefit to researchers who value the ease of use of high-end resources via such gateways."

During the first half of calendar year 2016, 7,274 gateway users consumed 97 million hours of normalized "TeraGrid Service Units" (a CPU measure). Fully 88% of these were delivered by SDSC's *Comet* and *Gordon* supercomputers.

"We anticipate that *Comet* will reach its goal of serving an active research community of more than 10,000 users, mostly via these science gateways," said SDSC Deputy Director Shawn Strande, who also is *Comet's* program manager. "One of our goals for *Comet* is to speed up as many researchers as possible, rather than supporting a handful of heroic calculations, so we configured it to serve as one of the most productive HPC systems available to the academic research community, just as its predecessor, *Trestles*, was."

Watch a video about *Comet* by scanning the QR code on the left or by visiting https://youtu.be/2rX9antVMRw

SDSC Director Michael Norman (third from right), joins *Comet* launch team members in accepting HPCwire's Editors' Choice award for the Center's 'Comet' supercomputer from HPCwire publisher Tom Tabor (far right) at the Supercomputing 2015 (SC15) conference in Austin, TX.
Photo: Cindy Wong, SDSC

Virtualized Clusters Expand Comet's Reach

Comet is the first XSEDE production system to support high-performance virtualization at the multi-node cluster level. Its use of Single Root I/O Virtualization (SR-IOV) means researchers can use their own software environment as they do with cloud computing, but achieve the high performance they expect from a supercomputer.

Comet's virtual cluster interface was used by scientists to confirm the discovery of gravitational waves in the universe by the NSF-funded Laser Interferometer Gravitational-Wave Observatory (LIGO) program, verifying a hypothesis by Albert Einstein 100 years ago. That landmark discovery, announced by the NSF in February 2016, has opened up a new era of exploration for astronomers and astrophysicists. (Read more about this landmark discovery on page 17.)

"We are pioneering the area of virtualized clusters, specifically with SR-IOV," said Philip Papadopoulos, SDSC's chief technical officer. "Virtual sub-clusters can run applications over InfiniBand at near-native speeds, and that marks a huge step forward in HPC virtualization. In fact, a key part of this is virtualization for customized software stacks, which lowers the entry barrier for a wide range of researchers by letting them project an environment they already know onto *Comet*."

The Open Science Grid, a multi-disciplinary research partnership specializing in high-throughput computational services funded by the U.S. Department of Energy and the NSF, recently added high-performance virtualized clusters to its global infrastructure by taking advantage of the new and unique capability of *Comet*.

The integration of *Comet* into the OSG (Open Science Grid) provisioning system was led by a team including UC San Diego Professor Frank Würthwein, an expert in experimental particle physics and advanced computation. Würthwein joined SDSC in early 2015 to help implement a high-capacity data cyberinfrastructure across all UC campuses, as well as connect to key cyberinfrastructure organizations such as OSG. Würthwein was OSG's director during 2005, and has again served as its executive director since February 2015.

SDSC Receives 2015 HPCwire Editors' Choice Award for 'Comet' Supercomputer
Annual Awards Recognize Leaders in the Global HPC Community

SDSC was a recipient of the 2015 HPCwire and Editors' Choice Awards for its new *Comet* supercomputer that entered production early that year as a result of a National Science Foundation grant worth nearly $24 million, including hardware and operating funds. The award, for 'Best Data-Intensive System: End-User Focused', went to SDSC for *Comet's* role as a gateway to discovery that will serve thousands of researchers through simple, domain-specific web interfaces. The award was presented at the 2015 International Conference for High Performance Computing, Networking, Storage and Analysis (SC15), in Austin, Texas.

"The entire *Comet* team thanks HPCwire for honoring us with this award," said SDSC Director Michael Norman, who also is the principal investigator for the *Comet* project. "We hit the ground running following a smooth buildout and launch process, and now *Comet* is serving a wide range of users to help advance scientific discovery."

Gordon
Delivering on Data-intensive Demands

Gordon entered production in early 2012 as one of the 50 fastest supercomputers in the world, and the first one to use massive amounts of flash-based memory. That made it many times faster than conventional HPC systems, while having enough bandwidth to help researchers sift through tremendous amounts of data. *Gordon* is the result of a $20 million NSF grant.

By the end of June 2016, 1,186 research projects using *Gordon* were awarded among 1,138 principal investigators. One of *Gordon's* most data-intensive tasks has been to rapidly process raw data from almost one billion particle collisions as part of a project to help define the future research agenda for the Large Hadron Collider (LHC). Under a partnership between a team of UC San Diego physicists and the Open Science Grid, *Gordon* provided auxiliary computing capacity by processing massive data sets generated by one of the LHC's two large general-purpose particle detectors used to find the elusive Higgs particle. The around-the-clock data processing run on *Gordon* was completed in about four weeks' time, making the data available for analysis several months ahead of schedule.

TSCC Computing "Condo"
Affordable Computing for Campus & Corporate Users

SDSC's contribution to research computing at UC San Diego has been the *Triton Shared Computing Cluster (TSCC)*, a "condo computing" program established in 2013 that continued to exhibit robust growth in the latter half of 2015 and into 2016. Condo computing is a form of shared ownership model in which researchers use funds from grants or other sources to purchase and contribute compute "nodes" (computer servers) to the system. The result is a researcher-owned, shared computing resource of medium- to large-proportions and much larger than could typically be afforded by the average researcher for dedicated use. In early 2015, following a competitive solicitation process, SDSC selected a new technology vendor providing attractive pricing on compute nodes with the latest Intel server processors and other technology. This ignited new interest by researchers in *TSCC* participation, resulting in the addition of 91 compute nodes (more than 2000 compute "cores" or processors) to the cluster, increasing its capability by over 150%. These nodes were purchased by 14 new research groups or labs, doubling the participating groups to 29. (Read more about *TSCC* on page 28.)

Learn more about SDSC's HPC Systems by scanning the QR code on the left or by visiting https://goo.gl/McdvqA

Phil Papadopoulos is SDSC's Chief Technology Officer and chief architect behind SDSC's *Data Oasis* storage system. Papadopoulos also is principal investigator for the Prism@UCSD project to build a campus cyberinfrastructure capable of supporting extreme data-intensive communications.

Data Oasis
Among Academia's Fastest Parallel File Systems

SDSC's *Data Oasis* is a Lustre-based parallel file storage system linked to *Comet*, *Gordon*, and *TSCC*. As a critical component of SDSC's Big Data initiatives, *Data Oasis* currently has about 12 petabytes (PB) of capacity and speeds of up to 200 gigabytes (GB) per second to handle just about any data-intensive project. *Data Oasis* ranks among the fastest parallel file systems in the academic community. Its sustained speeds mean researchers could retrieve or store 240 terabytes (TB) of data—the equivalent of *Comet's* entire DRAM memory—in about 20 minutes, significantly reducing time needed for retrieving, analyzing, storing, or sharing extremely large datasets. In short, *Data Oasis* allows researchers to analyze data at a much faster rate than most other systems, which in turn helps extract knowledge and discovery from these datasets. In early 2015, *Data Oasis* began undergoing significant upgrades, including ZFS, a combined file system originally designed by Sun Microsystems and mated in a new hardware server configuration.

SDSC Cloud
First Ever Large-scale Academic Deployment of Cloud Storage in the World, Now with Compute

The SDSC IT Services team administers one of the first large-scale academic deployments of cloud storage in the world, now with computing capability. UC San Diego campus users, members of the UC community, and UC affiliates are eligible to join the hundreds of users who already benefit from the multi-petabyte, OpenStack's Swift object store. *SDSC Cloud* is the perfect storage choice for researchers with fixed budgets because unlike other cloud providers, *SDSC Cloud* boasts a simplified recharge plan that eliminates secondary fees such as bandwidth costs, charges assessed per request, and regional migration fees. Entering its second year in production is an elastic compute facility, based on Openstack Nova and Ceph. This comprehensive cloud environment allows researchers a testbed and development environment for developing cloud-based services. It is especially attuned to data sharing, data commons, and data analytics services.

Networking & Connectivity

SDSC has helped lay the groundwork and provide expertise in implementing networks that allow fast and unrestricted flow of information between systems and researchers, both on and off the UC San Diego campus.

Pacific Research Platform
Advancing Collaboration Up and Down the West Coast

From biomedical data to particle physics, researchers depend heavily on high-speed access to large datasets, scientific instruments, and computing resources. The NSF recently funded a $5 million, five-year award to UC San Diego and UC Berkeley to establish a high-capacity, data-centric "freeway system" that will give participating universities and other research institutions the ability to move data about 1,000 times faster than what's possible on today's inter-campus shared Internet.

SDSC is an anchor participant in this project, called the Pacific Research Platform, or PRP. The PRP links numerous research universities on the West Coast – including the 10 UC campuses, San Diego State University, Caltech, USC, Stanford, and University of Washington – via the Corporation for Education Network Initiatives in California (CENIC)/Pacific Wave's 100G infrastructure. The initiative also extends to Lawrence Berkeley National Laboratory, the National Energy Research Scientific Computing Center, NASA Ames, and the National Center for Atmospheric Research and several campuses outside of California, including international campuses. Partner networks include the Pacific Northwest Gigapop, the Energy Sciences Network, NASA Research and Engineering Network, the Metropolitan Research & Education Network, StarLight, and the Front Range Gigapop – with a long-term goal of engaging other research and education networks in the U.S. and abroad. (Read more about the PRP on page 33.)
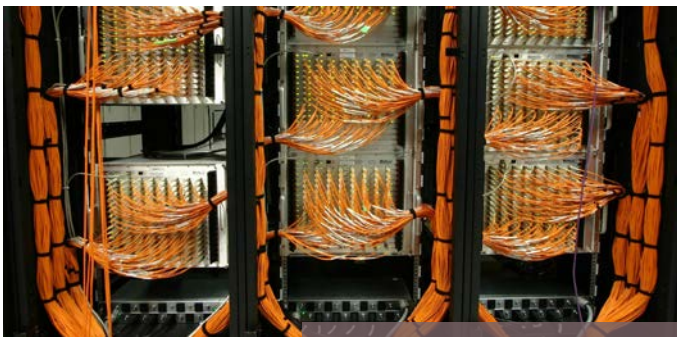

Image: Tim McNew, SDSC

Prism@UCSD
The HOV Lane for Broad-bandwidth Research

Working with campus partners, SDSC helped establish a research-defined, end-to-end networking cyberinfrastructure for the UC San Diego campus that is capable of supporting large data transmissions between facilities that might otherwise hobble the main campus network. This network extends the 'Science DMZ' concept into a Distributed Science DMZ. Called Prism@UCSD and backed by a $500,000 NSF grant, researchers with SDSC and the campus' California Institute for Telecommunications and Information Technology (Calit2) began work on the network in 2013 to support research in data-intensive areas such as genomic sequencing, climate science, electron microscopy, oceanography, and physics. "One can think of Prism as the HOV lane, whereas our very capable campus network represents the other lanes on the freeway," said Philip Papadopoulos, SDSC's chief technology officer and principal investigator for the Prism@UCSD project.

CHERuB
Connecting to the Information Superhighway

Called CHERuB for Configurable, High-speed, Extensible Research Bandwidth, this project is the result of a second $500,000 NSF grant, awarded to UC San Diego's Administrative Computing and Telecommunications (ACT) organization to connect the campus's Science DMZ (PRISM) and SDSC's *Comet* and *Gordon* supercomputers to high-bandwidth national research networks to advance a new range of data-driven research. With 100 gigabit-per-second connectivity, CHERuB supports multi-institutional data transit over networks such as the Internet2's Advanced Layer 2 Service (AL2S), the Department of Energy's ESnet, Pacific Wave/CENIC. CHERuB also is a joint project among those networks called the Advanced Networking Initiative (ANI), the result of a $62 million grant under the American Recovery and Reinvestment Act to build a national 100G "information backbone." Research domains that can benefit from CHERuB include cosmology, atmospheric sciences, electron microscopy, genomic sequencing, oceanography, high-energy physics, and telemedicine – all of which encompass data-intensive research. "CHERuB allows research domains that need very large datasets and data flows to exist and not collide with smaller-sized flows in the everyday Internet," according to SDSC Network Architect and CHERuB co-PI Thomas Hutton.

SCIENCE HIGHLIGHTS
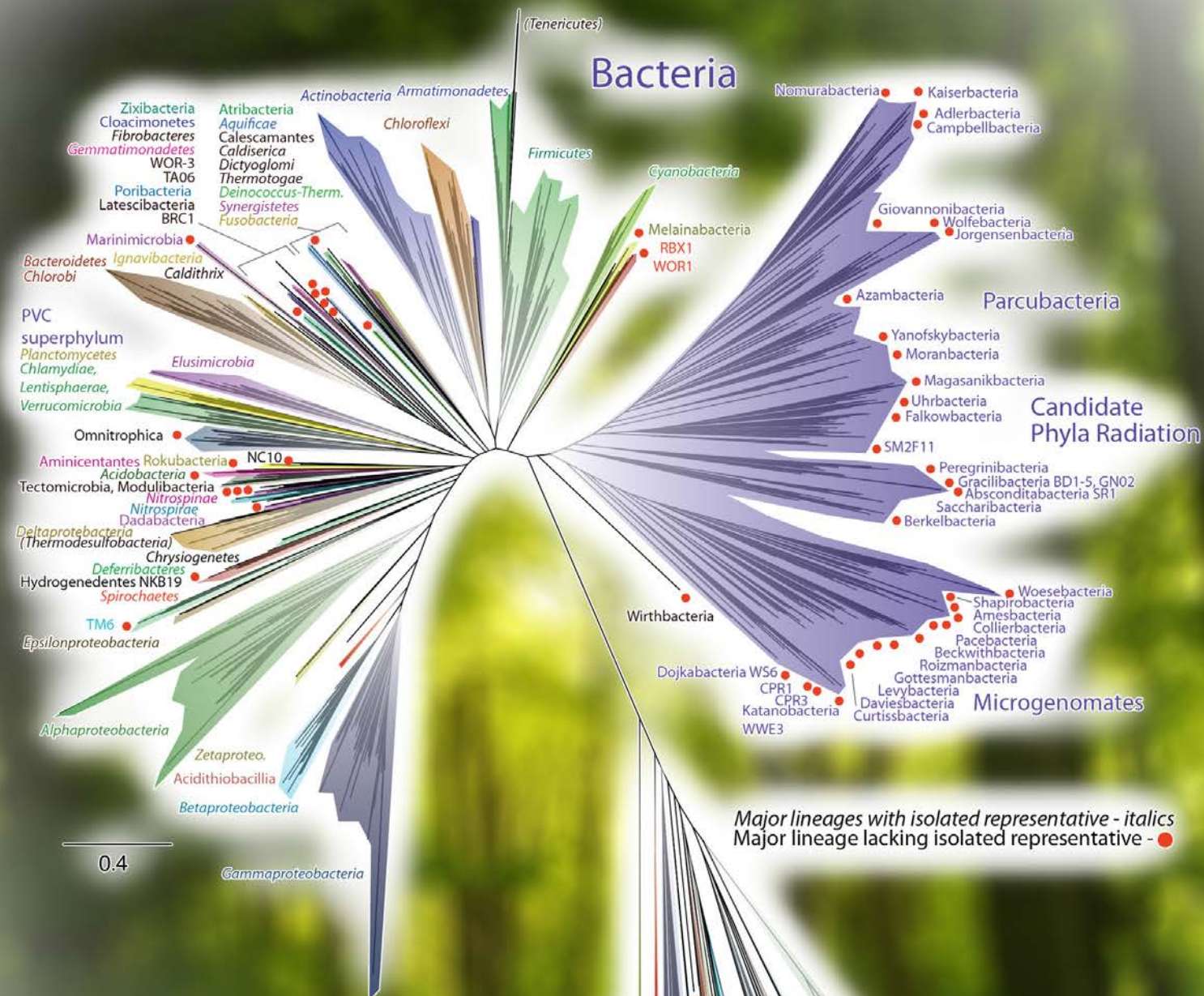
BURGEONING
BIODIVERSITY

SDSC Biology Researcher Mark Miller is Principal Investigator (PI) for the popular CIPRES Science Gateway.

SDSC Supercomputers, CIPRES Science Gateway Help Define New "Tree of Life"

A new "tree of life" depicting more than 1,000 new types of bacteria and Archaea lurking in the Earth's nooks and crannies was revealed this year with the help of SDSC's data-intensive supercomputing resources and a phylogenetics "gateway" called CyberInfrastructure for Phylogenetic RESearch or CIPRES, a website that allows researchers to explore evolutionary relationships among species.

"The CIPRES Gateway allows scientists to conduct their research in significantly shorter times without having to understand how to operate supercomputers," said Mark Miller, principal investigator for the CIPRES gateway and an SDSC biology researcher, adding that to-date, more than 2,500 papers have been published citing the use of this gateway.

The new "tree of life", published in the April issue of *Nature Microbiology* and widely publicized in the general press, reinforced once again that the life we see around us – plants, animals, humans and other so-called eukaryotes – represents a tiny percentage of the world's biodiversity. "The 'tree of life' is one of the most important organizing principles in biology," said Jill Banfield, a UC Berkeley professor of earth and planetary science, policy, and management, and the study's principal investigator. "The new depiction will be of use not only to biologists who study microbial ecology, but also biochemists searching for novel genes and researchers studying evolution and earth history."

Access to supercomputers was a key part of this study, helping researchers to investigate relationships by comparing DNA sequences information between species. This type of analysis is becoming more powerful as the number of DNA sequences available is increasing rapidly, with new, larger data sets requiring higher levels of computational power. Use of the CIPRES gateway initially allowed researchers access to SDSC's *Gordon*, the first high-performance supercomputer to use massive amounts of flash-based SSD (solid state drive) storage. The final trees were generated by SDSC's latest system, *Comet*, a petascale supercomputer designed to transform advanced scientific computing by expanding access and capacity among traditional as well as non-traditional research domains.

"The CIPRES Science Gateway was critical to our work," said Laura Hug, who computed the trees at the UC Berkeley and is now a biology faculty member at the University of Waterloo, Canada. "No run had successfully finished prior to our introduction to CIPRES."

Their investigation, representing the total diversity among all sequenced genomes, produced a tree with branches dominated by bacteria, especially by uncultivated bacteria. A second view of the tree grouped organisms by their evolutionary distance from one another rather than current taxonomic definitions, making clear that about one-third of all biodiversity comes from cultivated bacteria, one-third from uncultivated bacteria, and one-third from Archaea and Eukaryotes.
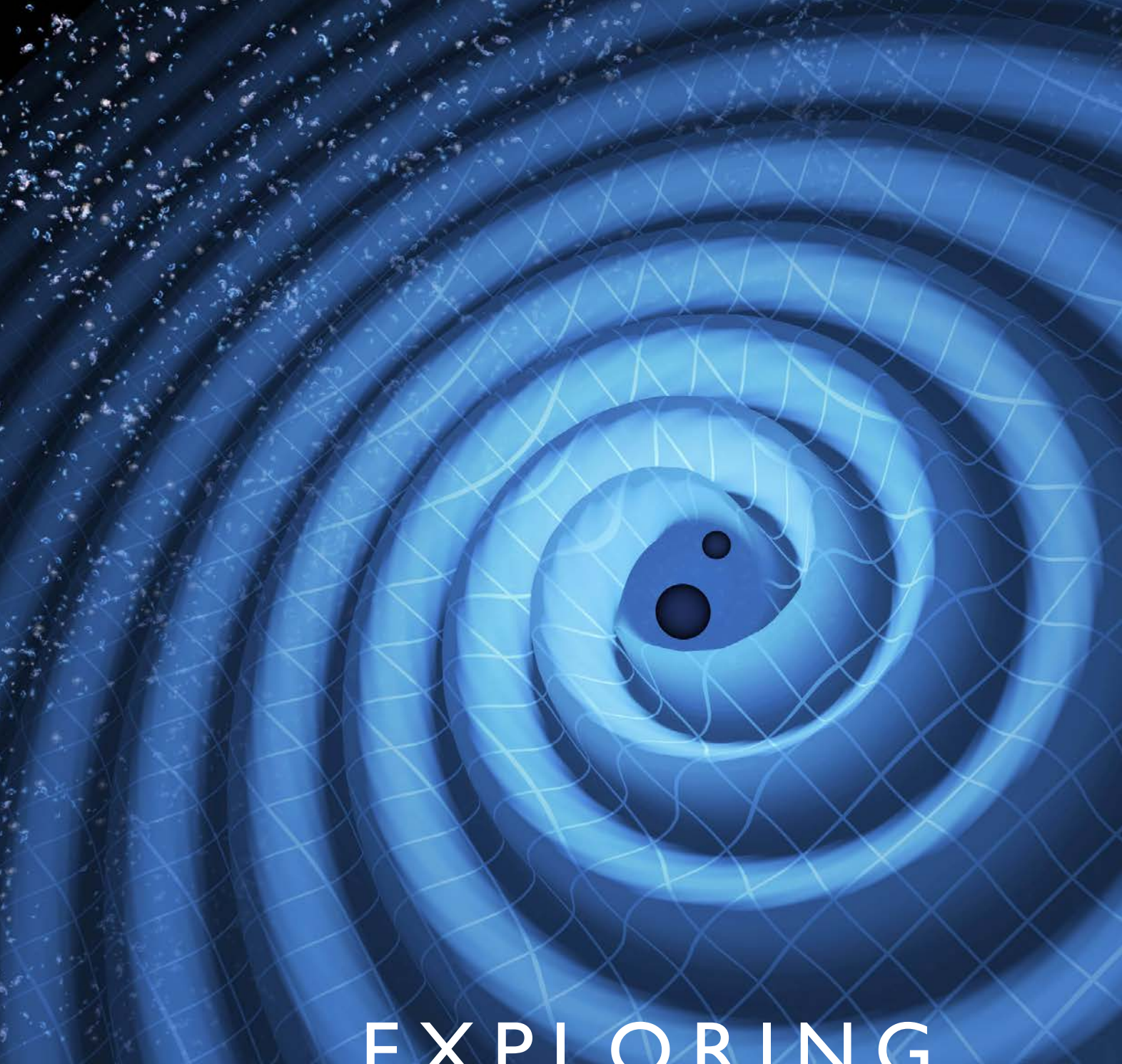
EXPLORING
THE UNIVERSE

This illustration depicts the merger of two black holes and the gravitational waves that ripple outward as the black holes spiral toward each other. In reality, the area near the black holes would appear highly warped, and the gravitational waves would be difficult to see directly.

Image: T. Pyle, LIGO

SDSC's 'Comet' Supercomputer Helps Confirm Gravitational Wave Discovery

SDSC's *Comet* was one of several supercomputers used by researchers to help confirm detection of gravitational waves in the universe as hypothesized by Albert Einstein some 100 years ago. The landmark discovery, announced by the National Science Foundation (NSF) in February 2016, has opened a new era of exploration for astronomers and astrophysicists.

On September 14, 2015 at 4:50:45 a.m. EST, scientists detected gravitational waves by both of the NSF-funded Laser Interferometer Gravitational-Wave Observatory (LIGO) detectors, according to the agency's announcement. The waves reached Earth from the southern hemisphere, passed through the planet, and emerged first at the LIGO interferometer near Livingston, Louisiana, and then, seven milliseconds later and 1,890 miles away at the second LIGO interferometer, in Hanford, Washington.

Scientists soon realized they had the tell-tale "chirp" signature of two black holes merging. At the moment of the collision, about three times the mass of the sun was converted into gravitational waves with a peak power outlet of about 50 times that of the whole visible universe, according to the NSF/LIGO announcement.

The collision of the two black holes happened 1.3 billion years ago, generating waves that have been traveling through the universe ever since. The discovery, however, immediately rippled with full force throughout and beyond the astrophysics research community, confirming Einstein's belief that gravitational waves existed. That belief was an outgrowth of his groundbreaking general theory of relativity, which depicted gravity as a distortion of space and time triggered by the presence of matter.

While the production discovery analyses for the actual gravity wave detection last September were run on clusters at Syracuse University and at the Albert Einstein Institute-Hannover, Germany, SDSC's *Comet* was one of several systems which contributed processing to the analysis and review of the result.

"The detection of gravitational waves represents the pinnacle of human achievement," said SDSC Director Michael Norman, who is also the principal investigator for the *Comet* program. "As an astrophysicist, it is doubly rewarding for me because it underscores the coupled nature of experimental and computational science in solving some the deepest mysteries in the universe."

LIGO researchers consumed almost 630,000 hours of computational time on *Comet*, using the system's new Virtual Cluster interface for the analysis of the data. This 'bare metal' approach – a term that describes an environment in a computer system or network in which a virtual machine is managed like a physical computer, and performs as well as one – means that a virtual computing cluster looks, feels, and performs almost exactly like the physical hardware, thereby reducing the entry barrier for researchers by letting them project an environment they already know onto *Comet*.

Saul Teukolsky, the Hans A. Bethe Professor of Physics and Astrophysics at Cornell University, has been using *Comet* since early 2015 for a project named "Gravitational Waves from Compact Binaries: Computational Contributions to LIGO."

"Nearly all of our work has been LIGO-related," said Teukolsky. "I believe that the theoretical work that I and other researchers do supports the experimental work done by other LIGO researchers in that these wave form simulations produce the kind of templates that they need."

In addition to black hole simulations, Teukolsky's research team also run binary neutron star cases. "LIGO hopes to detect those soon," he said.

MATERIALS
ENGINEERING

Thomas Manz, a chemical engineering professor at New Mexico State University, used SDSC's *Comet* supercomputer to design a cleaner and more energy-efficient catalytic process for producing polyester resins, pulyurethane foams, and plastics.

Making Chemical Production More Eco-Friendly & Energy-Efficient

A new catalytic process, designed with the aid of advanced computational resources including SDSC's *Comet* system, is expected to use less energy and generate fewer by-products than conventional catalysts used widely in the chemical industry for the production of polyester resins such as fiberglass, in addition to polyurethane foams and plastics.

The findings, published online in *Theoretical Chemistry Accounts* by Thomas Manz, a chemical engineering professor at New Mexico State University, suggests that a novel catalyst—a solution-phase bidentate zirconium complex—can efficiently mediate selective oxidation reactions directly with molecular oxygen, without the need for co-reactants. Selectivity refers to the ability to target a specific chemical bond. Greater selectivity results in greater energy efficiency, fewer byproducts and, in turn, more product.

"We believe this is one of the first examples to date of designing a brand new catalytic route and catalyst class from scratch using quantum chemistry calculations, and one that will reduce unwanted co-products that can be environmentally unfriendly," said Manz. "We're also convinced that this new route can result in substantial energy savings when compared with existing processes which use a co-reductant or require an oxidant other than molecular oxygen."

While that paper focuses on ethylene epoxidation, Manz's most recent computations using *Comet* show that their new catalyst can also be used in a two-catalyst process to selectively oxidize propene to pro-pylene oxide using molecular oxygen as the oxidant without co-reductant.

"It's much more difficult to produce propylene oxide than ethylene oxide," said Manz, adding that pro-pylene oxide, which is used to make polyurethanes, polyester resins, and other compounds, is one of the world's major chemical products, with global produc-tion estimated at about 15 billion pounds per year.

New catalyst shown in center with a zirconium (Zr) atom colored cyan, oxygen atoms (red), and organic ligands (green and orange.) Surrounding this is a computed reaction cycle for selective oxidation of ethylene to ethylene oxide over this catalyst. The number beside each arrow is the computed energy change in kilocalories per mole. Image: Thomas Manz and Bo Yang, NMSU.
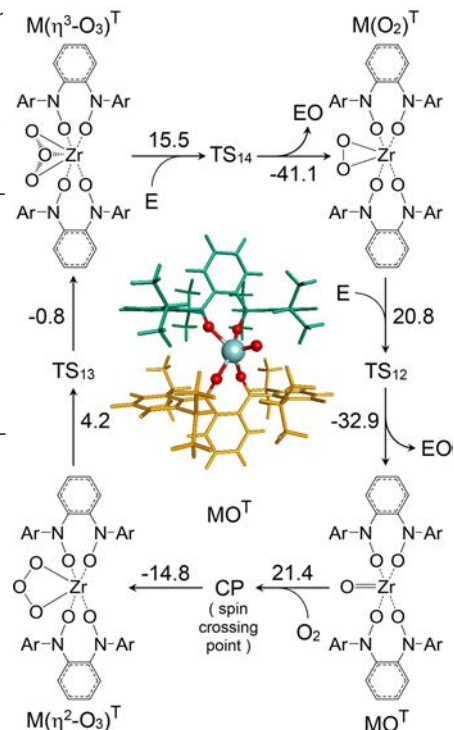
MAKING
RESEARCH MORE
ACCESSIBLE AND
COLLABORATIVE

SDSC Researcher Peter Rose is Site Head of the RCSB Protein Data Bank at SDSC, and as Project Scientist leads the Structural Bioinformatics Laboratory.

SDSC Awarded $1.4 Million NIH Structural Bioinformatics Grant

SDSC is the recipient of a National Institutes of Health (NIH) grant worth almost $1.4 million to make biological structures more widely available to scientists, educators, and students.

The award was granted to Peter Rose, Site Head of the RCSB Protein Data Bank (PDB) West at SDSC and Project Scientist of the Center's Structural Bioinformatics Laboratory; and Andreas Prlić, Technical and Scientific Team lead at the RCSB PDB. The RCSB Protein Bank is the single worldwide repository for the three-dimensional structures of large molecules and nucleic acids that are vital to pharmacology and bioinformatics research.

"Currently, interactive visualization of large complex structures and structural comparisons across the entire Protein Data Bank archive exceeds available network bandwidth and the memory of typical scientists' desktops, laptops, or mobile devices," said Rose, whose research interests also include large-scale "Big Data" mining, machine learning, and visualization of 3D structures and their application in structure-based drug design. "This currently requires dedicated local network and a high-performance computing infrastructure that is not widely available."

Part of the challenge is that the size and complexity of the structures have increased dramatically. For example, the recently determined structure of the HIV-capsid contains about 2.5 million atoms. "In short, interactive visualization of large-scale structural analyses and queries of the archive have become a 'Big Data' or data-intensive challenge," said Rose.

The project aims to make these structures accessible to a wider populace of researchers and students by developing a set of compression algorithms, applications, and workflows that will significantly improve the performance of interactive visualization of three-dimensional structures of large complexes over the internet.

The RCSB PDB website is accessed by about 300,000 unique visitors per month. "The tools that are being developed as part of this grant will directly benefit a large community of scientists and data analysts," said Prlić.
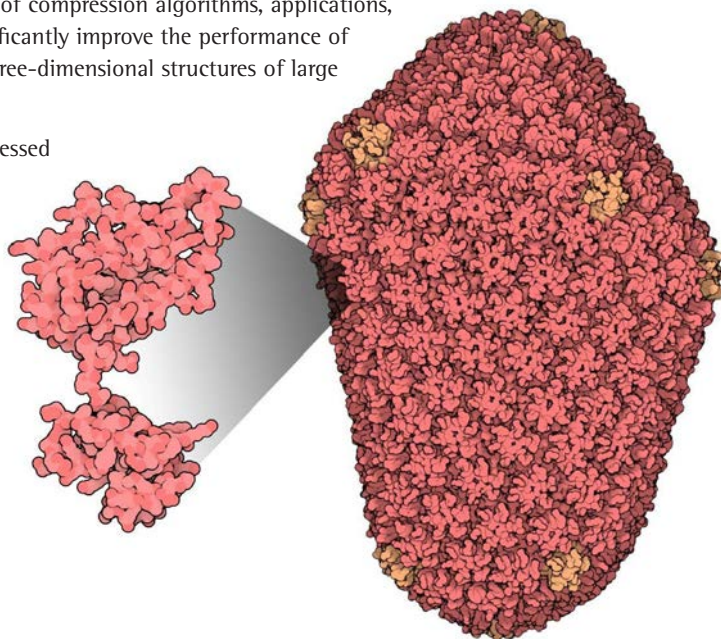
Structure of HIV Capsid
Image credit: RCSB Protein Data Bank

HUMAN HEALTH AND QUALITY OF LIFE

New Class of Anti-infective Drugs
May Overcome Antibiotics Resistance

Antibiotic resistance has been called one of the world's most pressing health problems, with some studies suggesting that without major improvements in drug discovery, by 2050 more individuals will die from drug-resistant bacterial infections than cancer.

To help stem this nightmare scenario, a team of researchers – with the aid of SDSC's data-intensive *Gordon* supercomputer – has identified a class of possible antibiotics with the potential to disable drug-resistant bacteria.

In essence, these new agents were found to attack the bacteria along two fronts: its external lipid cellular wall and its internal factory responsible for generating cellular energy in the form of adenosine triphosphate, or ATP.

In findings published in the online journal *Proceedings of the National Academy of Sciences (PNAS)*, the researchers showed that some FDA-approved anti-infective drugs act in part by disabling a cell's ability to synthesize the production of ATP. These so-called "uncouplers" work by collapsing a cell's 'proton motive force', or a reservoir of protons that powers protein turbines used to synthesize ATP.

Researchers further discovered that a few newly approved anti-tuberculosis drugs worked not only as uncoupler agents, but they also targeted enzymes that maintain a bacteria cell's outer membrane.

"This multiple targeting is expected to be of importance in overcoming the development of drug resistance because targeting membrane physical properties is expected to be less susceptible to the development of resistance," said contributing author J. Andrew McCammon, the Joseph E. Mayer Chair Professor of Theoretical Chemistry and Distinguished Professor of Pharmacology at UC San Diego; and an Investigator with the Howard Hughes Medical Institute.

Using molecular dynamics-based *in silico* screening, the researchers sought to identify other compounds with dual-targeting capabilities. Among other things, the team discovered that the brain cancer drug lead Vacquinol was one such compound, which was of particular interest because that drug also has direct killing activity against the tuberculosis bacteria.

Using *Gordon*, the team tested a series of compounds for uncoupler activity in hopes of finding more potent analogs with activity against the tuberculosis bacteria and S. (staphlococcus) aureus infections. Subsequent experiments revealed several agents with uncoupling capabilities, in addition to others with at least dual targeting capabilities.

"Pure uncouplers without enzyme targets are generally not expected to be good drug leads," said McCammon, also an SDSC Fellow. "But these results indicate that screening for combined enzyme inhibition, plus uncoupler activity, could lead to new antibiotic leads."
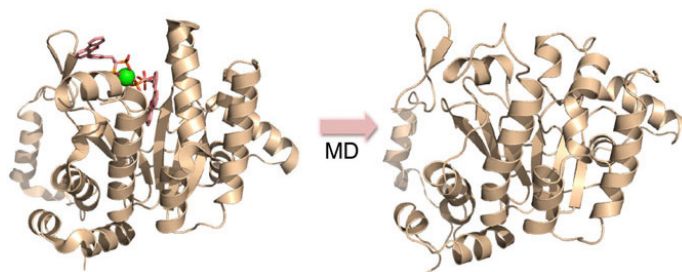
Vacquinol and its analogs as inhibitors for Rv3378c activity and bacterial cell growth. This image shows MD simulation/*in silico* screening approach used to identify Rv3378c inhibitors, with the X-ray structure (Left) and MD snapshot (Right). Image courtesy of J. Andrew McCammon, UC San Diego.

Advancing Our Understanding of Autism

A recent study finding that genetic changes that cause autism are more diverse than previously thought was partly enabled by the use of SDSC's new *Comet* supercomputer as well as the Center's expertise in data-intensive genome analyses. Researchers at the UC San Diego School of Medicine, led by Jonathan Sebat, associate professor of psychiatry and cellular and molecular medicine and director of the school's Beyster Center for Genomics of Psychiatric Disease, used *Comet* to unravel the genetic basis of autism spectrum disorder (ASD) by sequencing the complete genomes of hundreds of volunteers from families with one child affected by ASD, including the parents and typically developing siblings.

Mahidahr Tatineni, director of SDSC's User Support Group, and Wayne Pfeiffer, a SDSC Distinguished Scientist and expert on genome sequencing, collaborated on this project. The study, published this April in *The American Journal of Human Genetics*, highlights several genes that could play a key role in brain development, according to Sebat.

Uncovering a Mechanism to Block a Cancer Pathway

This image shows how platinum (large light grey sphere) in Cisplatin outcompetes copper to bind in a therapeutic way with MEK (mitogen-activated protein kinase) and block a cancer-causing pathway (sulfur = yellow, nitrogen = blue, hydrogen = white spheres). Image credit: Andreas Goetz and Igor Tsigelny, SDSC

Cisplatin is part of the chemotherapy treatment programs for many of the most common types of cancer. Last October, a team of UC San Diego researchers from SDSC, the Moores Cancer Center, and Department of Neurology announced an unexpected way this drug works, by blocking one of the pathways most commonly involved in driving cancer growth.

According to a study published online in the journal *Oncotarget*, the researchers showed for the first time that cisplatin can inhibit MEK1/2 (mitogen-activated protein kinase), an important enzyme in the communication pathway that transmits signals from the cell surface to the interior of the tumor cells.

Uncovering this mechanism is providing insights into how this drug might be combined with others to inhibit this pathway to improve patient care.

"We theorized that cisplatin might be able to bind in a therapeutic way with MEK1/2," said Igor Tsigelny, a research professor with SDSC, Moores Cancer Center, and the Department of Neurosciences.

ADF (Amsterdam Density Functional) software was used to perform highly accurate simulations on SDSC's *Comet* supercomputer and predict chemical structure and reactivity.

"Such simulations are computationally demanding," said co-author Andreas Goetz, an assistant project scientist at SDSC. "Quantum chemistry simulations of structural and electronic properties played a key role in demonstrating at a molecular level that cisplatin is able to replace copper at the binding sites. *Comet* was essential to obtaining timely results."

IMPACT
AND
INFLUENCE

LOCAL IMPACT AND INFLUENCE

Geisel Library, UC San Diego.
Credit: Erik Jepsen/UC San Diego Publications

Providing Leading "Cyberinfrastructure" and Expertise for UC San Diego's Grand Research Challenges

SDSC's dedication to harnessing meaning and value from a tidal wave of "big data" now generated by academic centers, commercial laboratories, government laboratories and observational tools is a major goal for UC San Diego researchers, as well as those in local industry and government. UC San Diego's strategic plan adopted in 2014 highlights four "grand research" themes of historical excellence at the university that can benefit from end-to-end "big data" cyberinfrastructure and educational know-how at SDSC. They include:

- Understanding and Protecting the Planet

- Enriching Human Life and Society

- Exploring the Basis of Human Knowledge, Learning, and Creativity

- Understanding Cultures and Addressing Disparities in Society

Specifically, SDSC's capabilities support the campus' research themes in the following areas of interest:

'Big Data' Science

Data from scientific research is being generated at a breathtaking pace, accelerated by the convergence of two new eras in the way research is now conducted: computational science, and data-intensive science and engineering – otherwise known as "Big Data" Science. SDSC is a pioneer in this field, and data-intensive science is at the core of the Center's new strategic plan.

Brain Research

Much of this work now requires data-intensive computation using the most powerful supercomputers along with the ability to analyze huge volumes of information from this research. SDSC has the high-performance computing and big data resources, along with the technical expertise, to support this grand challenge.

Smart San Diego

Remote sensing networks, combined with the ability to assimilate vast quantities of information, not only will form the basis to forecast devastating earthquakes, but also will save lives and property from the ravages of wildfires. In recent years, SDSC has engaged in collaborative efforts with academic scientists at UC San Diego and elsewhere to develop computer codes and other systems that promise to dramatically reduce time and effort needed to simulate and respond to hazards.

Infancy of the Universe

When scientists at the National Science Foundation's Laser Interferometer Gravitational Wave Observatory (LIGO) detected such waves in late 2015, they used SDSC's *Comet* supercomputer and other advanced computational resources to be sure of what they found. Supercomputers such as *Comet* will continue to provide researchers with valuable insights into such discoveries.

SDSC continued to play collaborative roles during the 2015/16 fiscal year in several other local-level programs, both on campus and beyond.

TSCC
Campus Research Computing Support Grows in 2016

In the modern research university, high-performance computing (HPC) is essential to advancing science in almost every discipline, from areas such as physics and engineering that are traditional users of HPC, to disciplines such as economics and political science that are riding the "big data wave" and emerging as new users. SDSC's contribution to research computing at UC San Diego has been the *Triton Shared Computing Cluster (TSCC)*, a "condo computing" program established in 2013 that continued to exhibit robust growth in the latter half of 2015 and into 2016. Condo computing is a form of a shared ownership model in which researchers use funds from grants or other sources to purchase and contribute compute "nodes" (computer servers) to the system. The result is a researcher-owned, shared computing resource of medium- to large-proportions and much larger than could typically be afforded by the average researcher for dedicated use.

In early 2015, following a competitive solicitation process, SDSC selected a new technology vendor providing attractive pricing on compute nodes with the latest Intel server processors and other technology. This ignited new interest by researchers in *TSCC* participation, resulting in the addition of 91 compute nodes (more than 2000 compute "cores" or processors) to the cluster, increasing its capability by over 150%. These nodes were purchased by 14 new research groups or labs, doubling the participating groups to 29.

Of particular note, a multi-campus collaboration of Professors Mike Gilson (Pharmacy), Andy McCammon (Chemistry), Rommie Amaro (Chemistry) and Ross Walker (Molecular Dynamics) of UC San Diego; Professor David Mobley (Pharmaceutical Sciences) of UC Irvine; and Chia-en Chang (Chemistry) of UC Riverside, received administrative supplements from the National Institutes of Health, totaling over $350,000, to purchase 27 special compute nodes for *TSCC*, equipped with a total of 216 NVIDIA graphics processing units (GPUs), which are well-adapted to computational chemistry and molecular biology. These nodes are now supporting a number of NIH-funded research grants developing and using computational chemistry methods for research related to biomolecular function, computer-aided drug design, protein-ligand binding, and related areas. Although the primary focus of *TSCC* is supporting researchers at UC San Diego, the program also attracted interest from off-campus, with researchers from The J. Craig Venter Institute, Syracuse University, and UC Merced joining the program and contributing nodes that enhanced the overall capacity and capability of the system to the benefit of all participants.

The *TSCC* group participated in the UC San Diego Research IT Spring Showcase, held in May 2016. The showcase featured a poster by researchers Eric Van Nostrand and Gabriel Pratt from the lab of Professor Eugene Yeo (Cellular & Molecular Medicine). The featured work was called "Identification and integrated analysis of RNA binding, protein binding sites and regulatory networks" and relied heavily on *TSCC* for data preparation and analysis. As stated by the researchers, "The ability to easily scale up with parallel processing has enabled us to rapidly process and analyze the over one thousand eCLIP datasets, test analysis improvements, and develop new mapping and binding site identification methods and metrics."

HPWREN technicians install a microwave dish antenna on a remote mountaintop tower. Courtesy of HPWREN

HPWREN

A Wireless Education and Safety Network for Greater San Diego

A goal for many research institutions is to transform the product or services of a government-funded research program into one that benefits society-at-large. One of SDSC's key accomplishments in this regard has been the High Performance Wireless Research and Education Network, or HPWREN.

As a collaborative project between SDSC and the Scripps Institution of Oceanography at UC San Diego, HPWREN functions as an internet-connected "cyberinfrastructure" for research, education, and public safety activities that connects often hard-to-reach areas in remote environments via a system of cameras and weather stations to report local weather and environmental conditions, from severe rainstorms to wildfires and earthquakes. (Please read more about HPWREN in the Focused Solutions and Applications section on page 50.)

To find more information about HPWREN online use a QR code reader or visit http://hpwren.ucsd.edu
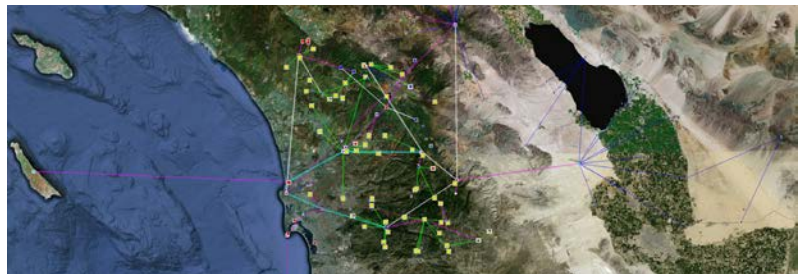
The HPWREN network extends high-speed data communications to remote areas of San Diego, Imperial, and Riverside counties. Courtesy of HPWREN

Ange Mason is SDSC's education program manager, helping to create innovative programs such as the Center's Research Experience for High School (REHS) students.

Educating and Empowering the Next Generation

Through its award-winning TeacherTech program, SDSC has trained more than 1,000 teachers in the San Diego region in science and technology, helping many underserved students span the "digital divide" to the Information Age. SDSC also provides programs to train and educate local high school students in computer science and technology, while fostering opportunities to bridge the gender gap for women in science, technology, education, and math (STEM).

SDSC, UC San Diego Extension Launch Data-Centric Courses

Spring 2016 saw SDSC and UC San Diego Extension join forces to launch a Data Mining Essentials course designed to provide individuals in business enterprises and scientific communities with improved tactics critical to design, build, verify, and test predictive data models. Topics included learning methods for data exploration and preparation, reviewing fundamental machine learning algorithms, and becoming familiar with model evaluation approaches. Participants were given access to a comprehensive set of data mining tools available on SDSC's high-performance computers, and completed the course via online sessions after attending hands-on training sessions at SDSC. More courses will be offered in late 2017 and early 2018. A discounted academic rate is available to UC San Diego graduate students, post-doctoral fellows, faculty, and staff.

SDSC's Annual Research Experience for High School Students

SDSC's Research Experience for High School Students (REHS) summer internship program scores high marks from students and mentors alike. REHS pairs students with SDSC mentors to help them gain experience in an array of computational research areas that include working with multi-scale simulation software and graphics processing unit (GPU) architectures in biomolecular research, exploring molecules associated with Parkinson's and Alzheimer's diseases, and validating neuronal models on supercomputers as part of SDSC's Neuroscience Gateways (NSG) project. Other opportunities focus on developing web pages and content, using social media to spread the word about computationally-based science, and learning how to effectively communicate the benefits of such research to science and society among a broader, less technically minded audience.

SDSC, UC San Diego Health Sciences Help High School Students 'MAP' their Academic Future

SDSC, in collaboration with the UC San Diego Division of Health Sciences, offers a novel program that creates mentoring relationships to enhance high school students' desire to learn, teach, and conduct research in a variety of science, technology, engineering, and mathematics (STEM) fields.

During the 2015/16 fiscal year, the Mentor Assistance Program (MAP) welcomed 77 students from 33 high schools, along with 34 mentors from 24 UC San Diego campus departments. Co-founded by Ange Mason, SDSC's education program manager; and Kellie Church, assistant professor in the university's Department of Reproductive Medicine, the first year of the program concluded with a symposium at SDSC that drew almost 300 guests, including school district administrators, principals, and teachers as well as parents and family members of students.

"The recent symposium was one of the proudest moments of my professional career," said Mason. "Some 90 percent of the students said they would return if given the opportunity, and half of them were offered summer internships, giving them the opportunity to continue their research experience throughout the summer."

"The MAP program was simply phenomenal," said participant Zoe Smith, who attends one of San Diego's High Tech High schools. "My mentor was awesome, my teammate was great, and our final product was a success. I could not have asked for a better experience!"

The second year of the program will start in late 2016.

For more information on SDSC's pre-college education programs use a QR code reader or visit StudentTech at http://education.sdsc.edu/studenttech/

Poster sessions for SDSC's REHS (top) and MAP programs let students share their work with educators and their families.
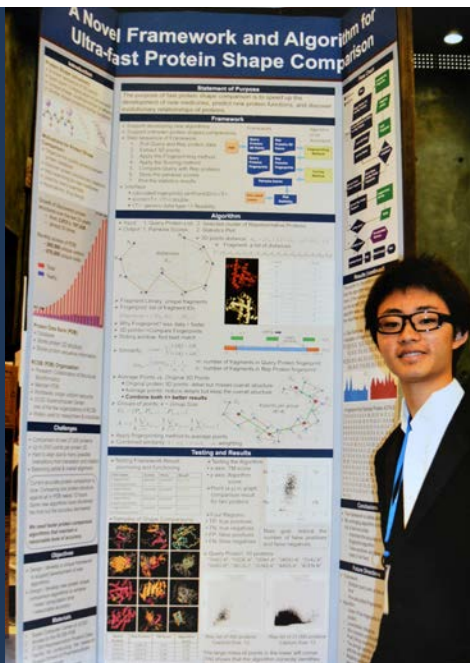
SDSC-Mentored High School Student Wins Top Science Prize

Kevin Wu, a senior at Canyon Crest Academy, won first place in the 2015 Greater San Diego Science and Engineering Fair competition and was invited to the California State Science Fair competition for a project to create a novel framework and algorithm that accelerates the process of protein shape comparison. "I've always been interested in math and computer science, especially the algorithms," said Wu, who started this project earlier as an intern at SDSC under Peter Rose, Principal Investigator at the Structural Bioinformatics Lab at SDSC and the Site Head of the RCSB Protein Data Bank West (RCSB PDB), the single worldwide repository for three-dimensional structures of proteins and nucleic acids that are vital to biology, medicine, and drug discovery. Due to the complex computation involved, accurately comparing one protein against all others in the PDB can take about one to three days. "My goal was to speed that up," said Wu. The result? Wu's "fingerprinting" algorithm cut that comparison process to just minutes using an approximate method before resorting to more extensive calculations on a much smaller subset of structures.

STATE IMPACT AND INFLUENCE

Aligning with Principles and Partnerships

UC@SDSC: Data-Enabled Science Based on Collaboration, Innovation, & Education

Founded as a national supercomputer center in 1985, SDSC's mission has evolved over recent years to include, among other things, the fostering of data-rich research collaborations across the University of California system. As part of that effort, SDSC has aligned itself with three core UC principles when it comes to system-wide research investments:

- Act as one system of multiple campuses to enhance UC's influence and advantage

- Promote efficient inter-campus collaborations and system-wide economies of scale

- Serve the State of California

SDSC's portfolio of high-performance computing resources, along with its 'big data' expertise and outreach programs, are essential to stimulate collaboration across the UC system. Partnerships range from finding ways to better predict the impact of wildfires, earthquakes or droughts, to tracking infectious disease behaviors via social media, or creating a data superhighway to move huge amounts of information rapidly and securely.
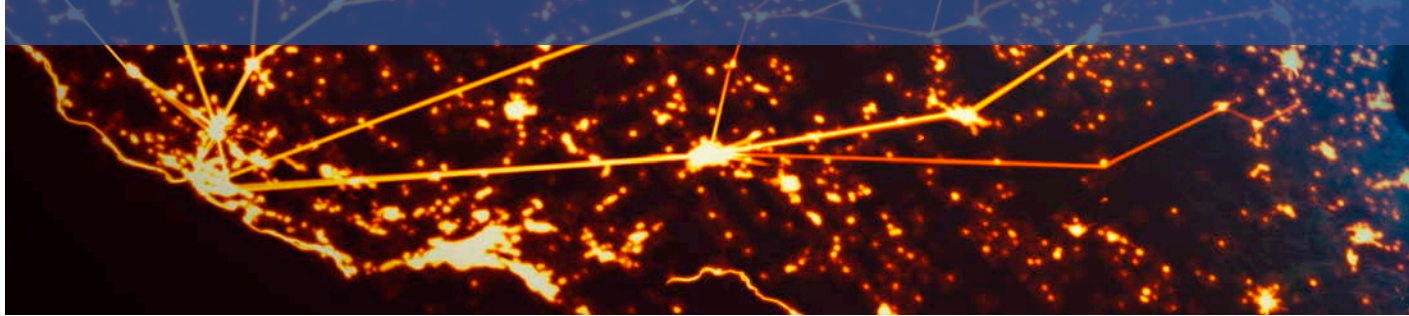
In 2014, SDSC launched an initiative called UC@SDSC – an engagement strategy that highlights collaboration, innovation, and education while promoting the Center's resources and technical expertise as a valuable asset to the entire UC system. An External Advisory Board consisting of Vice Chancellors, Deans, and Distinguished Professors from all 10 UC campuses and leadership from three national laboratories was formed to provide guidance and recommendations, resulting in numerous new collaborations. Some highlights for the latest fiscal year include:

HPC@UC

Launched in April 2016 as one of the latest UC engagement initiatives, HPC@UC provides UC researchers access to SDSC's high-performance computing resources, including *Comet* and *Gordon*, and expertise to make efficient use of them. The collaboration is being offered in partnership with the UC Vice Chancellors of Research as well as campus CIOs. HPC@UC is intended to broaden the base of UC researchers who use advanced computing while seeding promising computational research. The response to HPC@UC has been overwhelmingly positive. To date, SDSC has awarded more than three million core-hours to researchers at UC Santa Barbara, UC Irvine, UC Santa Cruz, and UC Los Angeles.

To read more about HPC@UC, use a QR code reader or visit www.sdsc.edu/collaborate/hpc_at_uc.html

Pacific Research Platform

To meet the needs of researchers in California and beyond, the National Science Foundation awarded a five-year grant to fund the Pacific Research Platform (PRP). The PRP's data sharing architecture, with end-to-end 10-100 gigabits per second (Gb/s) connections, will enable region-wide virtual co-location of data with computing resources and enhanced security options.

Led by Calit2 Director Larry Smarr, Calit2 Researcher Tom DeFanti; SDSC's Frank Würthwein and Phil Papadopoulos; John Graham (UC San Diego); Camille Crittenden (UC Berkeley); John Hess (CENIC), and Eli Dart (ESnet); the PRP supports a broad range of data-intensive research projects that will have wide-reaching impacts on science and technology worldwide. Cancer genomics, human and microbiome 'omics integration, biomolecular structure modeling, galaxy formation and evolution, telescope surveys, particle physics data analysis, simulations for earthquakes and natural disasters, climate modeling, virtual reality and ultra-resolution video development are just a few of the projects that are benefiting from the use of the PRP. The PRP will be extensible across other data-rich domains as well as other national and international networks, potentially leading to a national and eventually global data-intensive research cyberinfrastructure.

The initiative also extends to Lawrence Berkeley National Laboratory, the National Energy Research Scientific Computing Center, NASA Ames, and the National Center for Atmospheric Research and several campuses outside of California. Partner networks include the Pacific Northwest Gigapop, the Energy Sciences Network, NASA Research and Engineering Network, the Metropolitan Research & Education Network, StarLight, and the Front Range Gigapop. The project uses CENIC's California Research and Education Network (CalREN), and integrates Science DMZs, developed by ESnet as secure network enclaves for data-intensive science and high-speed data transport, creating a secure big-data network that will enable researchers worldwide to collaborate with each other.

To read the full press release about the PRP use a QR code reader or visit https://goo.gl/asXPx7

Health CyberInfrastructure UC Engagement Efforts

SDSC's Health Cyberinfrastructure (CI) Division is participating in a multi-million dollar project with the City of Hope, a comprehensive cancer care center, along with UC Irvine and other universities and health organizations, to create a cyberinfrastructure to aid research by developing a secure, cloud-based data management platform. The Health CI Division is assisting the City of Hope's consolidation of its California Teachers Study (CTS) data, including datasets and accompanying documentation. The work entails the design, build, and implementation of a dedicated cloud-based data management platform within the division's Sherlock Cloud infrastructure.

To read the full press release about the Health CI Division's participation with the City of Hope, use a QR code reader or visit https://goo.gl/yoeDTJ

Additional Collaborations

Molecular Dynamics

Ross Walker, director of the Walker Molecular Dynamics Lab at SDSC and an adjunct professor with UC San Diego's Department of Chemistry and Biochemistry, is further optimizing his AMBER Molecular Dynamics (MD) software suite to provide the fastest possible performance for MD simulations on commodity hardware under a NSF proposal called "Sustained Innovation in Acceleration of Molecular Dynamics on Future Computational Environments: Power to the People in the Cloud and on Accelerators", along with funding from Intel and NVIDIA, and a National Institutes of Health (NIH) supplement proposal called "A Next-Generation Intercampus Computational Resource for Drug Discovery Innovation." The NIH component of this project is focused on a co-designed, multi-node graphics processing unit (GPU) system from Exxact Corp. called the UC Southern California BioSciences Research resource, which is managed as part of SDSC's *Triton Shared Computing Cluster (TSCC)*, expanding the reach beyond UC San Diego to include other UC campuses. The collaboration currently includes UC Irvine and UC Riverside, in addition to UC San Diego researchers including Andrew McCammon and Rommie Amaro.

Social Media Analysis

A group of masters students, under the guidance of SDSC's Amarnath Gupta, has been studying the effects of political debates on social media political conversations using Twitter data, news network surveys, and exit polls. The collaboration with UCLA's UC Institute for Prediction Technology (UCIPT) is a multi-campus program to accelerate innovations that leverage social technologies to predict human behaviors and outcomes with collaborative input from UC Irvine and UC San Diego. The students are performing topic modeling on hashtags and comparing usage in tweets pre- and post-debate, as well as comparing exit polls and surveys of voters' decisive election issues. The project is using AsterixDB database technology, developed by researchers at UC Irvine and UC Riverside.

Seismic Modeling

In early 2016, SDSC opened a second Intel Parallel Computing Center, with the focus on earthquake research including detailed computer simulations of major seismic activity that can be used to better inform and assist disaster recovery and relief efforts. Yifeng Cui, director of the High Performance GeoComputing Laboratory at SDSC, is heading the new center. In addition, the Keck Foundation-funded Collaboratory for Interseismic Simulation and Modeling (CISM) project, launched within the last year, is being led by the Southern California Earthquake Center (SCEC), with Cui as a co-PI. The goals of the collaboration, which also includes UC Riverside, is to port, tune, and verify the CISM software RSQSim, a physics-based rupture simulator, on large supercomputing systems.
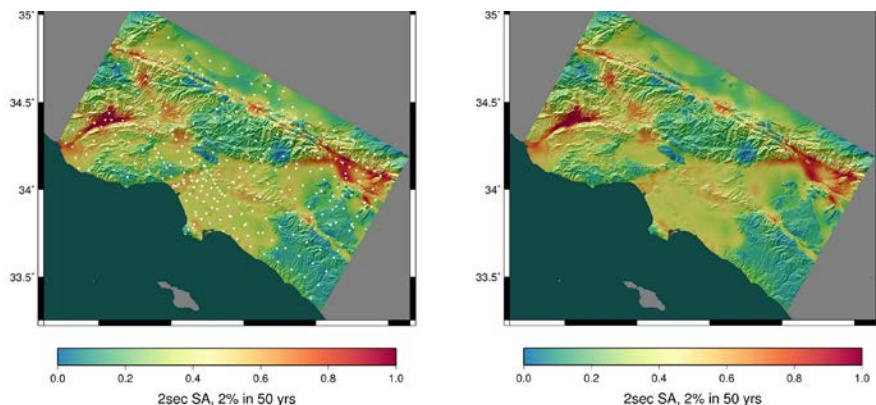
CyberShake Study 15.4 hazard map of Los Angeles basin with 336 sites marked with white triangles (left). Maps display the shaking intensity expected with a 2% probability in 50 years. Warm colors represent areas of high hazard. Map credit: SCEC, using a GPU-based wave propagation code ported and optimized at SDSC

New Ocean Current Simulations Alter View of Climate Change Impacts
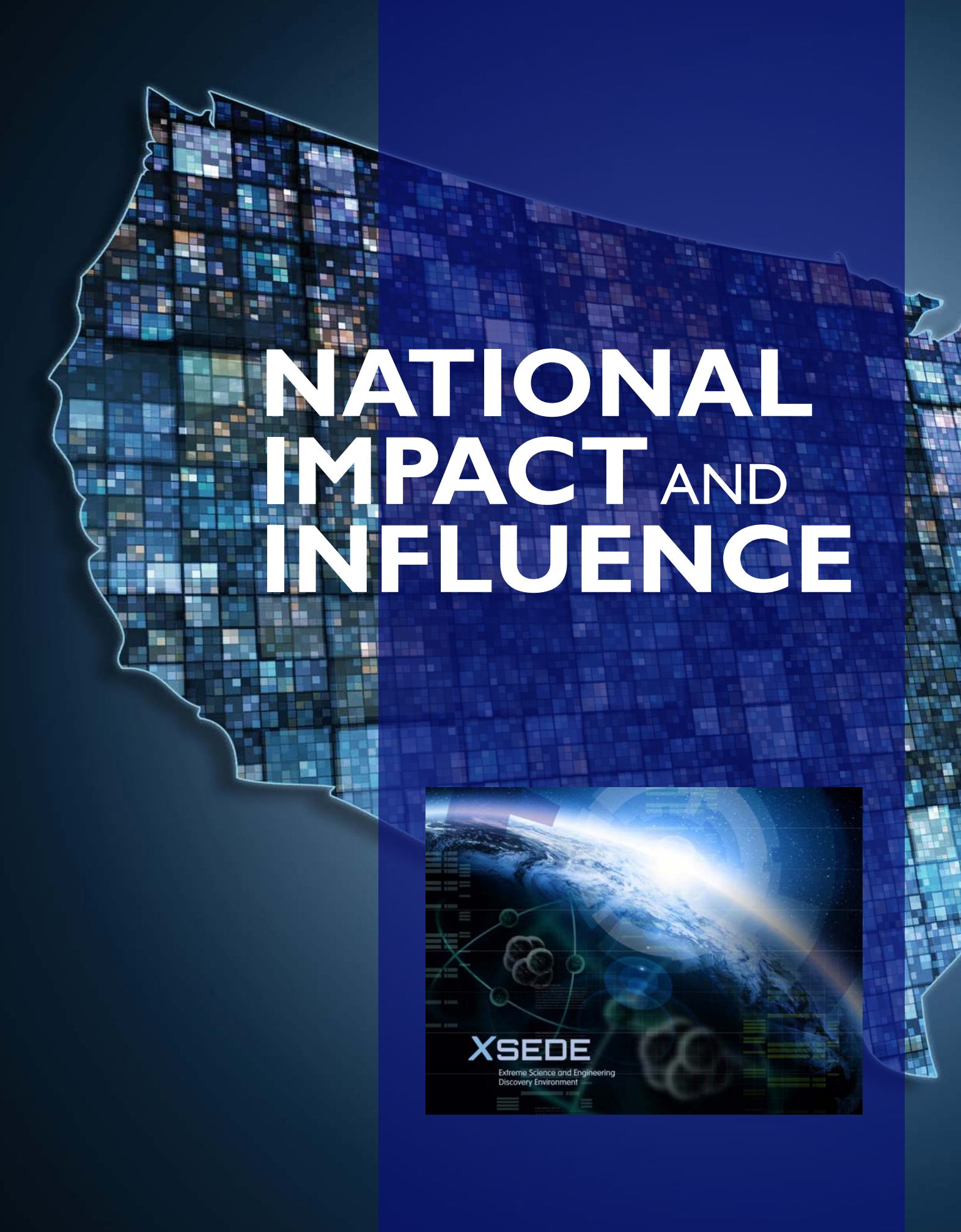
A "more realistic" computer model, created with the aid of SDSC's *Gordon* supercomputer, paints a new picture of global warming's impact on the complex processes that drive ocean mixing in the vast eddies swirling off the California coast. The new model, published in the July issue of *Nature Geosciences*, more accurately describes how global warming affects the winds and circulation in the California Current system that runs parallel to the coast, and how previous models, based largely on offshore winds generated by global warming, misrepresented what was actually happening. Information like this is critical – the California Current provides a habitat for one of the most populous and species-diverse regions in the world, while providing a bounty of seafood and tourism jobs for the state.

For years, computer models suggested that global warming would increase coastal upwelling, implying an increase in net productivity of nutrients for marine life. However, these earlier models failed to zoom in on other features, including coastline shape and topography as well as sea surface temperatures. The latest simulations and images show that the orography and meandering coastline are actually slackening winds closest to the California shoreline along the entire U.S. West Coast, and in particular from San Francisco to south of Newport Beach. More specifically, the simulations and observations show that near-shore reduction of winds not only is weakening upwelling along the coast, but it's also reducing the number and velocity of swirling eddies which typically would quench the flow of nutrients for biological production.

"*Gordon* helped and is helping us to develop numerical coupled atmospheric and oceanic simulations that are very costly in terms of resources," said Lionel Renault, an atmospheric and oceanic sciences researcher with UCLA, and lead author of the paper.

To read the full press release, use a QR code reader or visit https://goo.gl/kH3UY7

NATIONAL IMPACT AND INFLUENCE

XSEDE
Extreme Science and Engineering
Discovery Environment

SDSC's Advanced Computing Resources and Partnerships at the National Level

I n late 2015, SDSC celebrated its 30th anniversary, using this milestone to highlight the Center's accomplishments while paying tribute to the founding directors who charted the course for SDSC to be one of the first four supercomputer centers opened by the National Science Foundation (NSF).

SDSC has a long history of projects and partnerships that have benefited science and society at the national level across numerous domains, from brain research to tracking cosmological events at the furthest reaches of the universe. Together these projects have generated thousands of published papers and presentations, underscoring SDSC's contributions across the country as well as in other parts of the world.

"The launch of *Comet* marks yet another stage in SDSC's leadership in the national cyberinfrastructure ecosystem," said James Kurose, Assistant Director of the National Science Foundation's Computer and Information Science and Engineering (CISE) Directorate, during SDSC's 30th anniversary celebration last fall. "Through this launch and the extraordinary computing capabilities of SDSC, the Center will continue to expand the frontiers of science and engineering, allowing researchers to open new windows into phenomena as vast as the Universe and as small as nanoparticles."

SDSC Director Michael Norman noted that SDSC's mission has expanded over its three decades to encompass much more than advanced computation, including a host of services related to the voluminous amount of digitally based information generated daily, and systems designed to analyze, store, and share that data.

"In recent years the research community has moved into a new era of scientific endeavor based on computational science, now best described as data-intensive science," said Norman. "The term 'big data' became the shorthand description for this, or, for academia, 'data science and engineering.' This convergence of computational science with data science and engineering rests on an inherent reliance of interdisciplinary collaborations, which is needed to solve the grand research challenges of our times."

Some of SDSC's key national activities and partnerships include:

The NSF's Extreme Science and Engineering Discovery Environment (XSEDE) Program

XSEDE, a partnership of 19 institutions, enters its next five years under a renewed NSF agreement, continuing to provide academic researchers with the most advanced collection of integrated advanced digital resources and services in the world. As the only supercomputer center participant on the West Coast, SDSC provides advanced user support and expertise for XSEDE researchers across a variety of applications. In 2015, XSEDE forged a new alliance with the Open Science Grid (OSG), a multi-disciplinary research partnership specializing in high-throughput computational services funded by the U.S. Department of Energy and the NSF. The partnership allows OSG scientists to use XSEDE resources, such as *Comet*, to further their research. The integration of *Comet* into the OSG provisioning system was led by a team including UC San Diego Professor Frank Würthwein, an expert in experimental particle physics and advanced computation. Würthwein joined SDSC in January 2015 as the Center's lead for distributed high-throughput computing. (Read how SDSC's *Comet* supercomputer helped confirm gravitational wave discovery on Page 17)

The Open Science Grid Consortium

The OSG, a partnership of more than 100 institutions across universities and national laboratories, coordinates a sharing environment for compute and storage resources, networks, and software, as well as new innovations. SDSC is currently one of lead institutions in the OSG, and SDSC Director Michael Norman is a member of the OSG Council, the governing body that defines policy and strategic direction. OSG operates services that allow for transparent computation across more than 150 computing clusters worldwide, including National Grid Initiatives in Europe, Asia, and the Americas. UC San Diego Physics Professor Frank Würthwein joined SDSC in 2015 and currently serves as OSG's Executive Director.

Making Data-intensive, High-Performance Computing Resources Available

SDSC's new *Comet* supercomputer is accessible via the XSEDE allocation process to U.S. researchers as well as those affiliated with U.S.-based research institutions. *Comet* entered service in early 2015 with the mission of expanding access and capacity among traditional as well as non-traditional research domains. (see page 8 for more details.) *Gordon*, the first high-performance supercomputer to use large amounts of flash-based memory, entered service in 2012 following a $20 million NSF award. Billed as the "largest thumb drive in the world," *Gordon* has been an ideal resource for data mining and exploration, where researchers have to churn though tremendous amounts of data just to find a small amount of valuable information.

Frank Würthwein is the Distributed High-Throughput Computing lead at SDSC and also serves as OSG's Executive Director.

Sid Karin, SDSC's first director, was on hand to mark the Center's 30th anniversary to accept two awards: Founding Director and Founding Fathers. Karin thanked SDSC Director Michael Norman as well as a larger group, which "without their help, support, and involvement, there was no chance of SDSC being successful whatsoever" when he approached the NSF three decades ago about a cooperative agreement to establish a supercomputer center on the northwest end of the UC San Diego campus.

When SDSC first opened its doors, the Center chose to focus on a number of disciplines that emphasized both its service activities as well as its research and development activities, according to Karin. "Among those disciplines were what we now call 'big data', as well as computational biology," he said. "Looking back, we were very early adopters in these two areas, and I'm pleased to see how our early visions of those disciplines have been expanded with SDSC's current research activities and current supercomputers, which have been configured very specifically to deal with big data challenges as well as bioinformatics."

In noting the remarkable advances made in high-performance computing over the last three decades, Karin recalled SDSC's first supercomputer, a Cray XMP-48. "It was serial number seven – which made it one of the seven most powerful supercomputers in the world. Today, a typical smart phone is about as powerful as that XMP-48 was when we started operations in late 1985. I understand that today, *Comet* is about two million times as powerful as that first Cray system. I find that quite remarkable, and at the same time gratifying that more scientific disciplines are realizing how advanced computation can accelerate their research."

NSF West Big Data Innovation Hub (WBDIH)

In November 2015, the National Science Foundation established four regional Big Data Innovation Hubs throughout the U.S. The Western region is comprised of 13 states with Montana, Colorado, and New Mexico marking the eastern boundary. Principal investigators for the WBDIH include SDSC Director Michael Norman; Michael Franklin, the Thomas M. Siebel Professor of Computer Science and Chair of the Computer Sciences Division at UC Berkeley; and Ed Lazowska, the Bill & Melinda Gates Chair in Computer Science & Engineering at the University of Washington. The Hub's purpose is to connect, educate, incubate and facilitate multi-state, multi-sector partnerships in the area of big data innovation. Thematic spokes include Managing Natural Resources & Hazards, Metro Data Science, and Precision Medicine with supporting thematic rings that include Big Data Technology and Data-Enabled Scientific Discovery & Learning. (Please see westbigdatahub.org for more details.)

Supporting the National BRAIN Initiative through the Neuroscience Gateway

Charting brain functions in unprecedented detail could lead to new prevention strategies and therapies for disorders such as Alzheimer's disease, schizophrenia, autism, epilepsy, traumatic brain injury, and more. The BRAIN Initiative (Brain Research through Advancing Innovative Neurotechnologies), launched by President Barack Obama in 2013, is intended to advance the tools and technologies needed to map and decipher brain activity, including advanced computational resources and expertise. In 2015, the NSF and the United Kingdom's Biotechnology and Biological Sciences Research Council (BBSRC) awarded funding for a new Neuroscience Gateways project led by SDSC. That project, which will contribute to the national BRAIN initiative, is a collaboration between UC San Diego, with SDSC's Amit Majumdar as the principal investigator (PI) and Subhashini Sivagnanam as co-PI; Yale University, with Ted Carnevale as PI; and University College London, with Angus Silver as PI. (Learn about other Science Gateways involving SDSC staff starting on Page 41.)

National Data Service (NDS)

SDSC has taken a leadership role in the burgeoning National Data Service, a U.S. consortium of research computing centers, governmental agencies, libraries, publishers and universities. NDS builds on the data archiving and sharing efforts already underway within scientific communities and links them together with a common set of tools. NDS's unique value is not in new tools, but in making it easier to use such tools together through dependency management and increased interoperability between cyberinfrastructure services. NDS is an emerging vision for how scientists and researchers across all disciplines can find, reuse, and publish data and provides a platform and sandbox for developers creating data services. SDSC maintains this commitment through leadership on the Technical Advisory Committee, seed storage and compute capacity to enable pilot projects. (Learn more about the NDS at www.nationaldataservice.org.)

Big Data Specialization Courses

Under a new partnership with Coursera, SDSC launched a series of MOOCs (massive open online courses) as part of a Big Data Specialization on the Coursera platform. Consisting of five courses and a final Capstone project, this specialization provides valuable insight into the tools and systems used by big data scientists and engineers. In the final Capstone project, learners apply their acquired skills to a real-world big data problem. To date, the courses have reached more than 350,000 students in every populated continent – from Uruguay to the Ivory Coast to Bangladesh. A subset of students pay for a certificate of completion. The Big Data Specialization course, launched in September 2015, will be offered at least until September 2018.

To learn more about SDSC's Big Data Specialization courses offered through the Coursera platform, use a QR code reader or visit www.coursera.org/specializations/bigdata

Visualization of 3-D Cerebellar Cortex model generated by researchers Angus Silver and Padraig Gleeson from University College London. The NeuroScience Gateway was used for simulations.

NSF Awards $15 Million to Create Science Gateways Community Institute
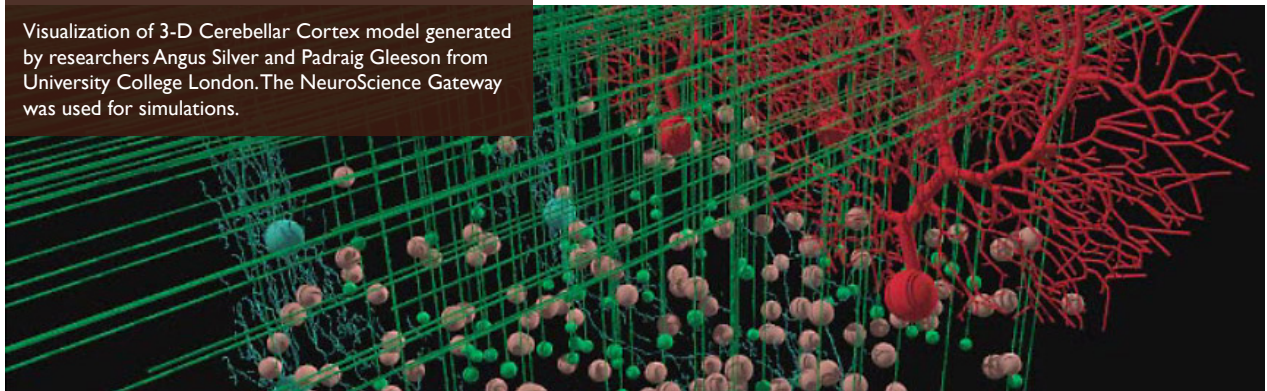
Nancy Wilkins-Diehr is an associate director of SDSC and co-PI of XSEDE (Extreme Science and Engineering Discovery Environment) as well as director of XSEDE's Extended Collaborative Support Services. Her XSEDE responsibilities include providing user support for Science Gateways as well as education, outreach, and training.

Multi-Partner Project Led by **SDSC** to Meet Researchers' Needs across **All Science Domains**

The National Science Foundation this summer awarded a collaborative team led by SDSC a $15 million, five-year grant to establish a Science Gateways Community Institute (SGCI) to accelerate the development and application of highly functional, sustainable science gateways that address the needs of researchers across the full spectrum of NSF directorates.

The Institute's goal is to increase the number, ease of use, and effective application of gateways for the greater research and engineering community, resulting in broader gateway use and more widespread engagement in science by professionals, citizen scientists, students, and more. The project officially got underway this summer.

A science gateway is a community-developed set of tools, applications, and data services and collections that are integrated through a web-based portal or suite of applications. Such gateways provide scientists access to many of the tools used in cutting-edge research – telescopes, seismic shake tables, supercomputers, sky surveys, undersea sensors, and more – and connect often diverse resources in easily accessible ways that save researchers and institutions time and money.

"Gateways foster collaborations and the exchange of ideas among researchers and can democratize access, providing broad access to resources sometimes unavailable to those who are not at leading research institutions." said Nancy Wilkins-Diehr, SDSC associate director and principal investigator for the project.
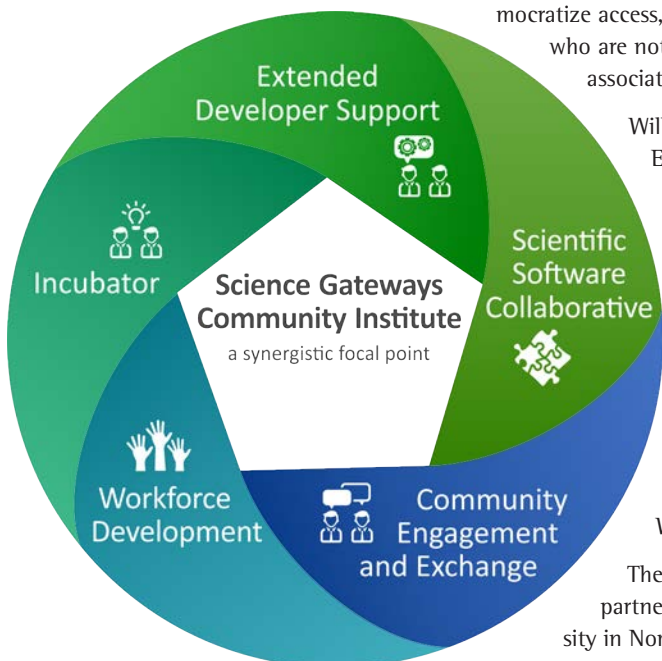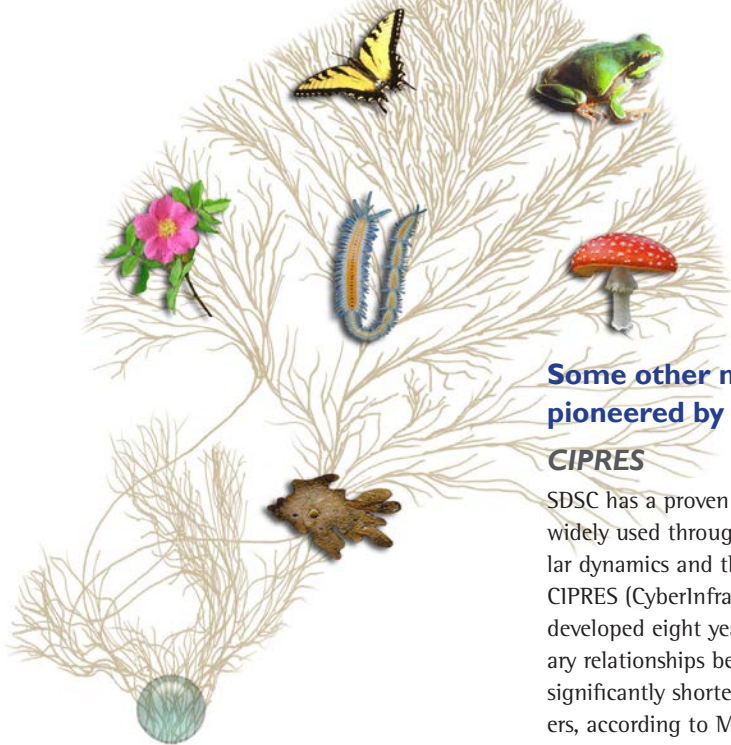
Wilkins-Diehr also is co-PI of the NSF-funded Extreme Science and Engineering Discovery Environment (XSEDE) program. "In XSEDE, we have observed tremendous growth in terms of the number of gateway users, the number of processing hours used on HPC resources and the number of published research papers using gateways in the last couple of years," she said. "We see the services offered by SGCI dovetailing nicely with those offered by XSEDE."

The primary focus of XSEDE's Extended Collaborative Support (ECSS) program, for example, is to support developers of existing gateways with their back-end connections to XSEDE resources. "SGCI frees us up to offer services developing front ends, both for projects that use supercomputers and those that do not," said Wilkins-Diehr.

The new SGCI award brings together expertise from a wide range of partner universities and institutions including Elizabeth City State University in North Carolina; Indiana University; University of Notre Dame; Purdue University; the Texas Advanced Computing Center (TACC) at the University of Texas, Austin; and the University of Michigan at Ann Arbor.

"We're excited about the opportunity to build community – nationally and internationally," added Wilkins-Diehr. "Sharing expertise about basic infrastructure allows developers to concentrate on the novel, the challenging, and the cutting-edge development needed by their specific user community."

Extended Developer Support

Incubator

Science Gateways Community Institute
a synergistic focal point

Scientific Software Collaborative

Workforce Development

Community Engagement and Exchange

Some other notable Science Gateways pioneered by SDSC researchers

CIPRES

SDSC has a proven track record in leading the creation of science gateways, which are widely used throughout numerous domains from astronomy and biophysics to molecular dynamics and the neurosciences. One of the most popular science gateways is the CIPRES (CyberInfrastructure for Phylogenetic RESearch) gateway, a web-based portal developed eight years ago by SDSC researchers that allows scientists to explore evolutionary relationships between species. Using gateways, scientists can conduct their research in significantly shorter times without having to understand how to operate supercomputers, according to Mark Miller, principal investigator for the CIPRES gateway and an SDSC biologist. Typically, about 200 CIPRES jobs are running simultaneously on SDSC's *Comet* supercomputer, and another 100 on *Gordon*. "The scheduling policy on *Comet* allows us to make big gains in efficiency because we can use anywhere between one and 24 cores on each node," said Miller. "When you are running 200 small jobs 24/7, those savings really add up in a hurry." CIPRES was used extensively to help create a new tree of life under a recent UC Berkeley-led study that added more than 1,000 new types of bacteria and Archaea (see page 15). In mid-2016, CIPRES reached a milestone by supporting its 2,500th scientific publication.

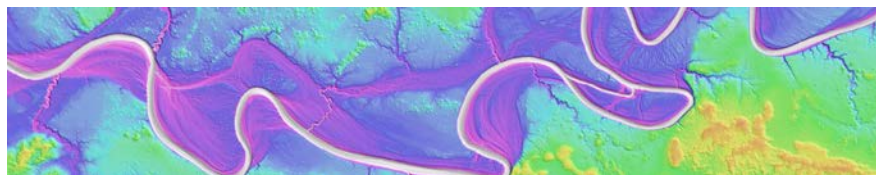For more information on CIPRES, use a QR code reader or visit www.phylo.org

Image of the Black Rock Desert Wilderness, Nevada. Image: Barrett Salisbury, Arizona State University/OpenTopography. Data: Lidar data acquisition by the National Center for Airborne Laser Mapping for Yo Matsubara, University of Virgina, Charlottesville.

OpenTopography

Initiated in 2009 with funding from the NSF, OpenTopography provides easy access to earth science-oriented, high-resolution topographical data and processing tools for a broad spectrum of research communities. A collaboration between UC San Diego, Arizona State University and UNAVCO, OpenTopography employs sophisticated cyberinfrastructure that includes large-scale data management, high-performance computing (HPC), and service-oriented architectures, providing researchers with efficient web-based access to large, high-resolution topographic datasets. Currently, OpenTopography data holdings comprise 200 lidar point cloud datasets and 95 raster datasets including the highly popular Satellite Radar Topography Mission (SRTM) global dataset. There has been a steady increase in data holdings due to partnerships and collaborations with various organizations with the academic NSF domain and beyond. OpenTopography has been averaging almost 400 new user registrations per month since the start of 2016, with more than 12,000 registered users and numerous others accessing data and running jobs as guests.

For more information on OpenTopography use a QR code reader or visit www.opentopography.org

SciGaP

Under a 2013 NSF award totaling $5 million for a collaborative five-year project, SDSC researchers are helping to develop and build a Science Gateway Platform (SciGaP) as a service to advance scientific discovery by providing researchers improved access to a variety of hosted or cloud services. The project is led by Indiana University's (IU) Marlon Pierce and Suresh Marru. The SciGaP project will create a set of hosted infrastructure services that gateway providers can easily adopt to build new gateways, according to Amit Majumdar, director of SDSC's Data Enabled Scientific Computing (DESC) group. These services will provide the basic features that any gateway requires, such as tools to connect high-performance computers and data resources across the country. Majumdar and Mark Miller of SDSC are leading SDSC's participation in the project. Also participating in the project is Borries Demeler from The University of Texas Health Science Center at San Antonio (UTHSCSA).

(More information about SDSC's Neurosciences Gateway is on page 39.)
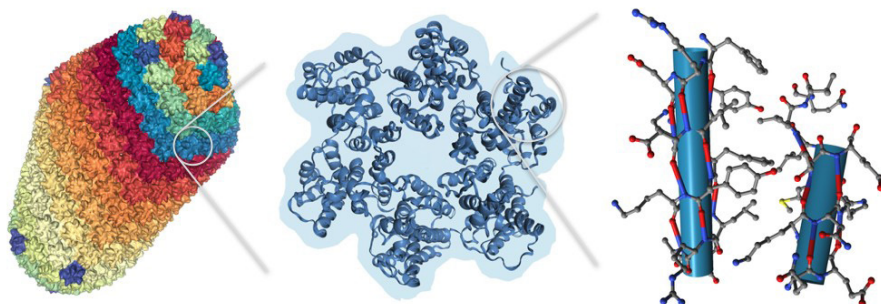
Multi-scale visualization of a ~2.5 million atom HIV-1 capsid structure (PDB ID 3J3Q) in a web browser. Courtesy of the RCSB PDB.

Protein Data Bank Surpass 120,000 Structures

By mid-2016, the Protein Data Bank (PDB), the single worldwide repository for three-dimensional structures of large molecules and nucleic acids, archived more than 120,000 structures after doubling its size in just six years. Co-located at SDSC in conjunction with UC San Diego's Skaggs School of Pharmacy and Pharmaceutical Sciences, and Rutgers, The State University of New Jersey, the Research Collaboratory for Structural Bioinformatics (RCSB) PDB serves as the U.S. data center for the archive. In addition, RCSB PDB supports online access to these structures to help researchers understand many facets of biomedicine, agriculture, and ecology, from protein synthesis and biological energy to fighting disease. The RCSB PDB website is being used by about 315,000 unique visitors per month.

"SDSC has provided safe haven for the RCSB PDB since it arrived at UC San Diego in the late 1990s," said SDSC Director Michael Norman. "It was the project that initially got us involved in data science, and it remains an important element in our 'big data' strategy."

In late 2015, researchers at SDSC were awarded a three-year National Institutes of Health (NIH) grant worth almost $1.4 million to make biological structures more widely available to scientists, educators, and students.

(Please see Page 21 for more details about this important award.)

FOCUSED SOLUTIONS and APPLICATIONS

FOCUSED
SOLUTIONS AND
APPLICATIONS

SDSC's wide range of expertise in advanced computation and 'big data' science has resulted in collaborative projects that extend beyond UC San Diego to both the local and national communities. These partnerships also bring together researchers across academia, industry, and government to collectively advance scientific discovery ranging from deepening our understanding of how the human mind works, to finding solutions to intractable diseases such as cancer in addition to natural and man-made disasters and emerging threats around the world.

For Industry…
Helping Build San Diego's Growing Biotech Sector

Ron Hawkins is director of Industry Relations for SDSC and manages the Industry Partners Program, which provides member companies with a framework for interacting with SDSC researchers and staff to develop collaborations.

The revolution in DNA sequencing technology and the resulting downstream applications such as precision medicine, agricultural genomics, and microbiomics are creating massive amounts of molecular sequence data that must be stored and analyzed.

Cost per Genome

The decreasing cost of DNA sequencing has rapidly outpaced computing capabilities, resulting in a backlog of computing required to analyze massive data sets. (Graphic courtesy of National Human Genome Research Institute)

This trend is adding new energy and growth to San Diego's already-vibrant biotech sector. In addition to benefiting established companies such as Illumina, a major developer of DNA sequencing instruments and supplies, entrepreneurs and investors are creating startups such as Human Longevity, Inc., a genomics and cell therapy company; and Edico Genome, a core technology company developing computational accelerators for genomics analysis. Located in the heart of this activity, on the Torrey Pines Mesa, SDSC has become a linchpin for many biotech companies due to its expertise in big data, data-intensive computing, and data science. SDSC's computational scientists and bioinformatics experts have worked with biotech firms, both large and small, to set up and tune
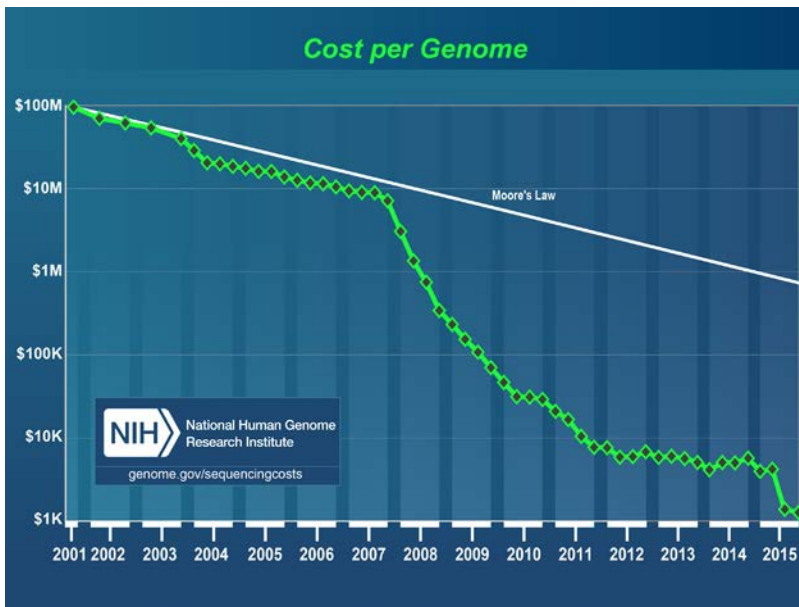
genomics analysis pipelines on SDSC's *Comet* and *Gordon* supercomputers, and to develop and host secure storage systems for large-scale genomics data sets.

As a member of Biocom, the leading local trade organization for San Diego life sciences companies, SDSC experts participated in Biocom's "Big Data Strategy" committee designed to help Biocom's hundreds of member companies formulate approaches to deal with and leverage 'big data' in their businesses. SDSC also helped organize and participate in Biocom's 'Big Data Executive Summit' held in May 2016, and SDSC Chief Data Science Officer Ilkay Altintas was a keynote speaker at Xconomy's 'Big Data Meets Big Biology' summit held in March 2016.

As part of SDSC's partnership with Biocom, the Center inaugurated semi-annual workshops for Biocom members on how to use its high-performance computing and storage resources. SDSC also provided Biocom members with discounted opportunities to participate in SDSC's data science training sessions. In partnership with storage technology company HGST, and high-performance computing solutions partner AdvancedHPC, SDSC hosted the 'Big Data Resources for Biotech Research' workshop in May 2016, which was well-attended by both campus researchers and local companies.

All of these activities served to further highlight SDSC's already visible role in San Diego's biotech ecosystem, and to further cement SDSC's partnerships in advancing life sciences research for the social benefit and economic development of Southern California. In concert with SDSC's strategic initiative in biomedical computing, the Center expects to grow its industrial outreach and partnerships in the biotech sector during late 2016 and through 2017.

Raymond de Callafon is director of the SyGMA laboratory and a professor at
UC San Diego's Department of Mechanical and Aerospace Engineering

Electric Grid Monitoring Laboratory Joins Industry in Advancing Electric Power Grid Technology

A new research laboratory housed at SDSC will help engineers from academia and industry harness the power of control theory to help improve the way electric power grids are operated in San Diego and beyond. Called SyGMA – for 'Synchrophasor Grid Monitoring and Automation' – the new industry-sponsored facility is supported by San Diego Gas & Electric (SDG&E), National Instruments, and OSIsoft, a manufacturer of application software for real-time data infrastructure solutions.

The SyGMA lab will collect and analyze data from a device called a synchrophasor, which monitors conditions on power transmission lines. This includes current, voltage, and specific properties of electricity such as frequency, phase angle, and real and reactive power flow.

"The emerging synchrophasor network on the 'smart' grid is an embodiment of the much-discussed 'Internet of Things (IoT),'" said Raymond de Callafon, a professor with the Department of Mechanical and Aerospace Engineering at UC San Diego's Jacob School of Engineering, and director of the new laboratory. "In addition to developing new systems and methods for monitoring and controlling electric grid stability, the SyGMA Lab will advance our understanding of how to collect, store, and analyze data from the vast sensor networks that will eventually comprise the IoT."

For more information on SyGMA, use a QR code reader or visit http://sygma.sdsc.edu

For Data-Intensive Applications…
SDSC Centers of Excellence

SDSC's Centers of Excellence are part of a broader initiative to assist researchers across many data-intensive science domains, including those who are relatively new to computational and data-enabled science. These centers represent key elements of SDSC's wide range of expertise, from big data management to the analysis and advancement of the internet.

SDSC Chief Data Science Officer Ilkay Altintas also directs the WorDS Center and is principal investigator for the WIFIRE project, a university-wide collaboration funded by the National Science Foundation (NSF) to create a cyberinfrastructure to effectively monitor, predict, and mitigate wildfires.

Workflows for Data Science (WorDS) Center

Called WorDS, for 'Workflows for Data Science', this center of excellence leverages more than a decade of experience within SDSC's Scientific Workflow Automation Technologies Laboratory, which developed and validated scientific workflows for researchers working in computational science, data science, and engineering.

"Our goal with WorDS is to help researchers create their own workflows to better manage the tremendous amount of data being generated in so many scientific disciplines, while letting them focus on their specific areas of research instead of having to solve workflow issues and other computational challenges as their data analysis progresses from task to task," said Ilkay Altintas, SDSC's chief data science officer and director of WorDS.

The WorDS Center is funded by a combination of sponsored agreements and recharge services. Expertise and services include:

- world-class researchers and developers well-versed in data science and scientific computing technologies

- research on workflow management technologies that resulted in the collaborative development of the popular Kepler Scientific Workflow System

- development of data science workflow applications through a combination of tools, technologies, and best practices

- hands-on consulting on workflow technologies for big data and cloud systems, i.e., MapReduce, Hadoop, Yarn, Spark and Flink; and

- technology briefings and classes on end-to-end support for data science.
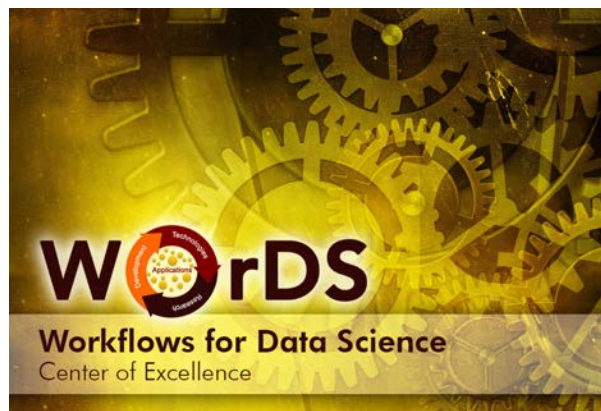
For more information on WorDS use a QR code reader or visit http://words.sdsc.edu

Sandeep Chandra is director of SDSC's Health Cyberinfrastructure (CI) and head of the Sherlock center of excellence.

Sherlock

Sherlock manages healthcare information technology and data services for academia and government that includes compliant cloud hosting, cyber security, data management, application development, and visualization. Formed in 2013 under an alliance between SDSC and several business partners, Sherlock's portfolio of services is offered to federal, state, and local governments, as well as the University of California system and universities nationwide.

Sherlock offers four major products which comply with HIPAA and FISMA regulations for dealing with sensitive information:

- Sherlock Analytics provides a platform for analyzing large, disparate data sets using best-of-breed Business Intelligence (BI) tools.

- Sherlock Case Management is a commercial off-the-shelf Customer Relationship Management (CRM) platform tailored to provide user interfaces, data interfaces, and workflows needed to meet unique project/business requirements.

- Sherlock Cloud is managed cloud hosting that provides both HIPAA- and FISMA-compliant services in accordance with hundreds of National Institutes of Standards and Technology (NIST) controls governing system access, information control, and management processes.

- Sherlock Data Lab helps transform digital data into meaningful information using a hybrid approach to data warehousing.

"Data management, technology, and policy challenges, especially in the health sector, can be overwhelmingly complex and confusing," said Sandeep Chandra, Sherlock's director. "Our expertise spans many IT disciplines that allows us to develop and deploy specific services, providing a solid, secure foundation for a wide range of initiatives, including healthcare fraud."

For more information on Sherlock use a QR code reader or visit http://sherlock.sdsc.edu

KC Claffy leads SDSC's CAIDA group, which has played a leading role in internet research in response to industry, government, and academic needs by providing tools and analyses to promote a scalable global internet infrastructure.

For more information on CAIDA use a QR code reader or visit www.caida.org/home/about/annualreports

For more about CLDS details and events use a QR code reader or visit http://clds.sdsc.edu

The WorDS Center and Sherlock join two other SDSC centers of excellence specializing in big data management across multiple disciplines, as well as internet topologies.

Internet Research for Cybersecurity and Sustainability

The Center for Applied Internet Data Analysis (CAIDA), formed in 1997 and based at SDSC, is a commercial, government, and research collaboration aimed at promoting the engineering and maintenance of a robust and scalable global internet infrastructure. Led by KC Claffy, CAIDA engages in internet research and infrastructure projects that span mapping of connectivity, monitoring security and stability, developing and evaluating future internet architectures, and economics and public policy. During 2015, CAIDA continued a collaboration with the Massachusetts Institute of Technology to map the mesh of interconnection in the internet, including methods to detect and localize the congestion to specific points in networks. This year CAIDA made progress on projects that monitor security and stability aspects of the global internet: traffic interception events (hijacks), macroscopic outages, and compliance with best practices that mitigate the risk of denial-of-service attacks. To support these and other community projects, CAIDA released BGPstream, a software framework for processing large amounts of historical and live inter-domain routing data. CAIDA's future internet architecture research includes tracking and modeling IPv6 evolution and contributing to development of a new information-centric network architecture, Named Data Networking (NDN). Inspired by the challenges of trying to make TCP/IP effectively serve the needs of emerging communication environments, notably the Internet of Things (IoT) and its cellular industry cousin '5G', NDN offers a radical change in the internet's basic protocol architecture with a new protocol based on data location rather than the host's location.

Center for Large-scale Data Systems Research

CLDS was established in 2012 as an industry-university partnership to study and address technical as well as technology management-related challenges facing information-intensive organizations in the era of big data. CLDS specializes in developing applicable concepts, frameworks, analytical approaches, case analyses and systems solutions to big data management, with a related goal of developing a set of benchmarks for providing objective measures of the effectiveness of hardware and software systems dealing with data-intensive applications. Based at SDSC to leverage the Center's resources and large-scale compute and storage resources, CLDS initiatives include the Big Data Benchmarking Community effort and the How Much Information? research program. As an industry-university collaboration, CLDS encourages participation by industry and welcomes industry sponsorship of projects. Center research is available via a variety of venues, including working papers, research briefings, multi-company forum workshops, and sponsor conferences.

For Hazards…

High Performance Wireless Research and Education Network (HPWREN): Supporting Science & Public Safety

HPWREN functions as an internet-connected cyberinfrastructure for research, education, and public safety activities. The project supports a fixed, high-bandwidth wireless data network using backbone nodes, typically on mountaintops, to connect often hard-to-reach areas in remote environments via a system of sensors, cameras, and weather stations to report local weather and environmental conditions, from severe rainstorms to wildfires and earthquakes. A key feature of HPWREN is imaging—its network currently includes more than 64 fixed mountaintop cameras positioned in 16 remote locations across San Diego, Riverside, and Imperial counties to support public safety operations.

Originally funded by the National Science Foundation and later transitioned to a sustainable public-private, membership-based funding model, HPWREN is now a joint collaboration between SDSC and UC San Diego's Scripps Institute of Oceanography (SIO), supported by a partnership that includes UC San Diego's California Institute for Telecommunications and Information Technology (Calit2) Qualcomm Institute, the Institute of Geophysics and Planetary Physics, Caltech's Palomar Observatory, San Diego County, San Diego Gas & Electric, Seismic Warning Systems, and a number of other government and non-profit organizations.

In 2015, Seismic Warning Systems joined HPWREN as a new Tier 1 (highest membership level) member. Seismic Warning Systems was founded in 2000 to provide systems and services for early seismic detection and automated system response to commercial, industrial, and government customers. This partnership allows Seismic Warning Systems to leverage HPWREN's high-speed wireless data transmission network to relay immediate earthquake notifications from its Regional Earthquake Warning Systems to first responders, critical infrastructures, families, schools, and more.

One of HPWREN's antenna-mounted observation cameras captures a lightning strike. Courtesy of HPWREN/SDSC.

For more information on HPWREN use a QR code reader or visit http://hpwren.ucsd.edu

HPWREN antenna by the Samuel Oschin telescope dome. Photo courtesy of Palomar/Caltech

During the 2015-16 fiscal year, HPWREN continued its long-time support to another Tier 1 member, Caltech's Palomar Observatory, by replacing and upgrading one of the key microwave links on Palomar Mountain in North San Diego County, and adding a second link to provide greater redundancy and communications bandwidth. HPWREN was recognized by the observatory for its support of key astronomical studies such as the Palomar-QUEST survey and the Palomar Transient Factory (PTF). The high-bandwidth communications capability provided by HPWREN enabling high-resolution astronomical images to be brought off the mountain for rapid analysis is expected to feature prominently in future projects such as the Zwicky Transient Factory (ZTF).

HPWREN also supports the Area Situational Awareness for Public Safety Network (ASAPnet), an extension of the HP-WREN infrastructure for the benefit of public safety communities, especially firefighters in San Diego County. ASAPnet consists of a wireless internet data communications overlay supporting rural fire stations and other firefighter assets, in addition to environment-observing cameras and other sensors.

In 2016, ASAPnet inaugurated the AlertSoCal system, a network of mountaintop cameras operated by researchers at SIO, expanding Southern California's state-of-the-art earthquake and weather monitoring systems to better detect fires in real time before they spread. New AlertSoCal 4K high-definition cameras will augment the existing HPWREN cameras. AlertSo-Cal provides firefighters and the public with a virtual fire look-out tower equipped with real-time and on-demand time-lapse imagery up to 12 hours in the past to spot the first signs of fire ignition. The unprecedented view in these remote regions and within the wildland-urban interface can aid fire crews with critical information on fire evolution in its early stages to support safer operations and more timely evacuations of residents.

WIFIRE: An Advanced Cyberinfrastructure to Help Firefighters Get a Jump on Wildfires

HPWREN/ASAPnet is a participant in the WIFIRE project which catalogs and integrates data related to dynamic wildfire models from sensors, satellites, and scientific models, creating visual programming interfaces for scalable wildfire models.

WIFIRE, funded by a $2.65 million multi-year grant from the NSF, is providing researchers from UC San Diego and the University of Maryland support to create a cyberinfrastructure to better monitor, predict, and mitigate wildfires in the future. UC San Diego participants in WIFIRE include researchers from SDSC, the California Institute for Telecommunications and Information Technology's (Calit2) Qualcomm Institute, and the Mechanical and Aerospace Engineering (MAE) department at the university's Jacobs School of Engineering.

"WIFIRE is designed to be scalable to users with different skill-levels using specialized web interfaces and user-specified alerts for environmental events broadcasted to receivers before, during, and after a wildfire," said Ilkay Altintas, SDSC's chief data science officer and WIFIRE's principal investigator. "This approach allows many big data sets to be subjected to user-specified data processing algorithms to generate threshold alerts and map visualizations within a very short time. Integration of this data into both rapidly available fire image data and models will better enable situational awareness, responses, and decision-making. We are currently testing the effectiveness and accuracy of our tools with a number of fire departments, including our own campus emergency services and the Los Angeles Fire Department."

For more information on WIFIRE use a QR code reader or visit http://wifire.ucsd.edu

Simulating "The Big One"

SDSC Researcher Yifeng Cui heads SDSC's High Performance GeoComputing Laboratory, and is an adjunct professor at San Diego State University. He is an internationally recognized expert in the development of large-scale earthquake simulation software and helped to establish the Southern California Earthquake Center (SCEC) as a world leader in advancing high-performance computing activities in earthquake system science.

California comprises about three-quarters of the nation's long-term earthquake risk, and Southern California accounts for nearly 50 percent of the national annualized earthquake loss, according to Federal Emergency Management Agency (FEMA) data.

In early 2016, Intel opened a second parallel computing center at SDSC to focus on earthquake research, including detailed computer simulations of major seismic activity that can be used to better inform and assist disaster recovery and relief efforts. Cui is the principal investigator for the new center.

Yifeng Cui is named SDSC's annual "Pi Person of the Year" for his research on computationally challenging problems in seismic research. (Read more on Page 4.)

The new Intel center continues the interdisciplinary collaboration between SDSC and the Southern California Earthquake Center (SCEC), one of the largest open research collaborations in geoscience. In addition to UC San Diego, the new center also includes researchers from UC Riverside, the University of Southern California, and San Diego State University.

"These are exciting times for computational seismology, as Intel's new technology and this additional support will help accelerate earthquake modeling research," said Cui. "The primary goal of the software development is to integrate the modernized code elements into a software ecosystem used by SCEC for production size of physics-based seismic hazard analysis in the U.S. and elsewhere."

Watch a Magnitude 7.8 earthquake simulation by scanning the QR code above or by visiting https://youtu.be/znWqQ7wa8-w
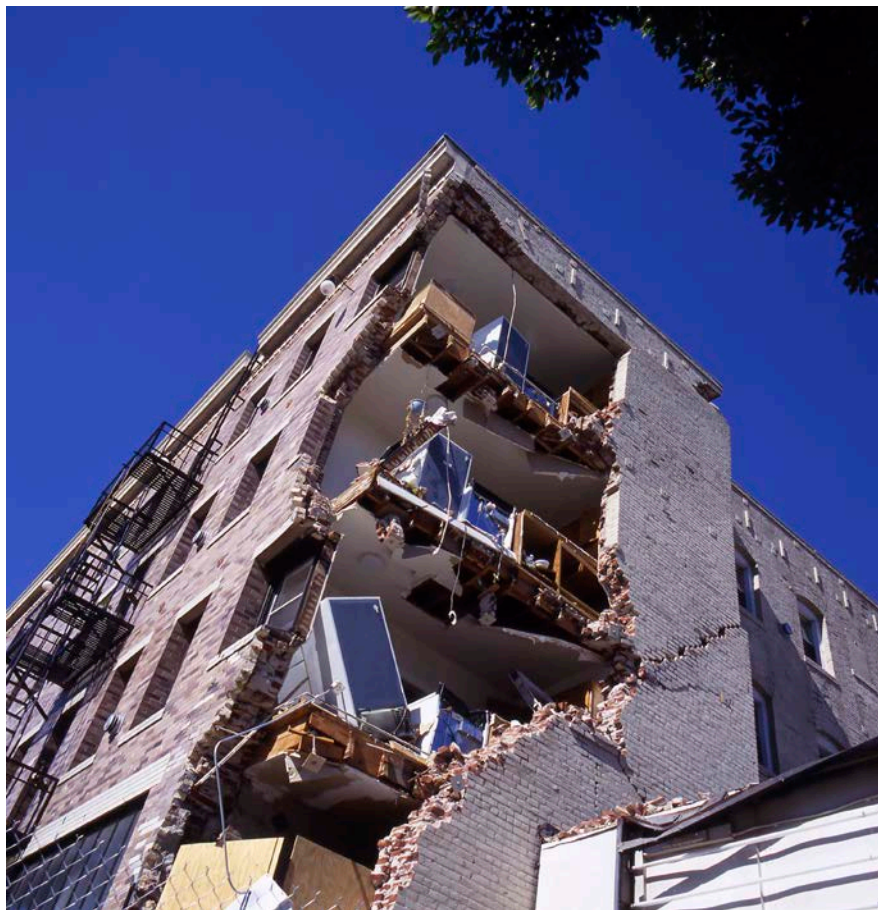
Proposal Success Rate

|	FY12	FY13	FY14	FY15	FY16
Proposals Submitted	118	84	77	91	73
Proposals Funded	55	37	39	38	33
Success Rate	46%	45%	50%	42%	45%

(Information updated from 2014/15 Annual Report)

In perhaps the most competitive landscape for federal funding in the last two decades, SDSC's overall success rate on federal proposals remains at about 45% compared to a national average of about 21% for computer science and engineering proposals at the National Science Foundation.

Number of Sponsored Research Awards
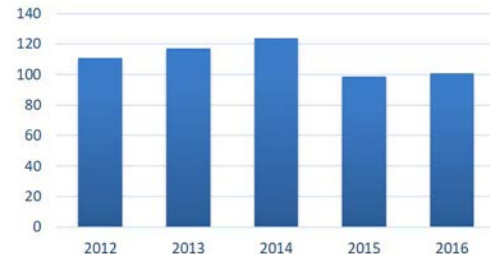
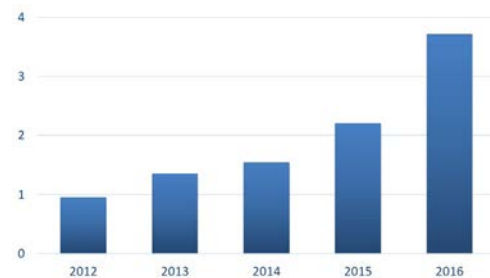*(Information updated from 2014/15 Annual Report)*

Sponsored Project Expenditures ($M)

— Contract and Grant Funds — Total Direct Expenditures — Indirect Costs Recovery
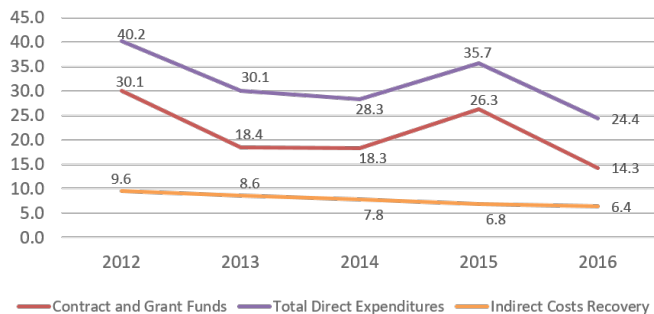
(Information updated from 2014/15 Annual Report)

Apart from the extraordinary research impact of SDSC collaborations and partnerships, a quick look at the fiscal impact of these collaborations is impressive. During its 28-year history, SDSC revenues have exceeded $1 billion, a level of sustained funding matched by few academic research units in the country. At the close of the 2016 fiscal year SDSC had 86 NSF-funded projects totaling $89 million.

Revenue/Support from Industry ($M)

(Information updated from 2014/15 Annual Report)

Geographical Distribution of National Users of SDSC HPC Resources

A total of 2,348 unique users from around the world accessed SDSC's HPC resources (*Gordon* and *Comet*) during FY2015-16. Of these users, 2,304 were based in the United States. The adjacent map displays a geographic disbursements of users from different cities across the U.S.

On *Gordon*, a total of 106,121,878 service units (SUs) were used during the fiscal period. On *Comet*, a total of 305,270,046 SUs were used during the same period.

ORGANIZATION & LEADERSHIP

SDSC Org Chart

MIKE NORMAN
Director

SHAWN STRANDE
Deputy Director

ILKAY ALTINTAS
Chief Data Science Officer

CHAITAN BARU
Assoc. Director Data Science and Engineering (on assignment to NSF)

PHIL PAPADOPOULOS
Chief Technology Officer

AMARNATH GUPTA
Assoc. Director Academic Personnel

WINSTON ARMSTRONG
Chief Information Security Officer

FRANK WÜRTHWEIN
Distributed High-Throughput Computing Lead

NIEVES RANKIN
Business Services

AMIT MAJUMDAR
Data-Enabled Scientific Computing

ILKAY ALTINTAS
Cyberinfrastructure Research, Education & Development

JAN ZVERINA
External Relations

CHRISTINE KIRKPATRICK
IT Systems and Services

PHIL PAPADOPOULOS
Cloud and Cluster Software Development

SANDEEP CHANDRA
Health Cyberinfrastructure

RON HAWKINS
Industrial Relations & Marketing

Executive Team

Ilkay Altintas
Chief Data Science Officer

Chaitanya Baru
Associate Director, Data Science and Engineering

Sandeep Chandra
Division Director, Health Cyberinfrastructure

Ronald Hawkins
Director, Industry Relations

Christine Kirkpatrick
Division Director, IT Systems and Services

Amit Majumdar
Division Director, Data Enabled Scientific Computing

Michael Norman
SDSC Director

Philip M. Papadopoulos
Division Director, Cloud & Cluster Software Development

Nieves Rankin
Division Director, Business Services

Shawn Strande
Deputy Director

Frank Würthwein
Lead, Distributed High-Throughput Computing
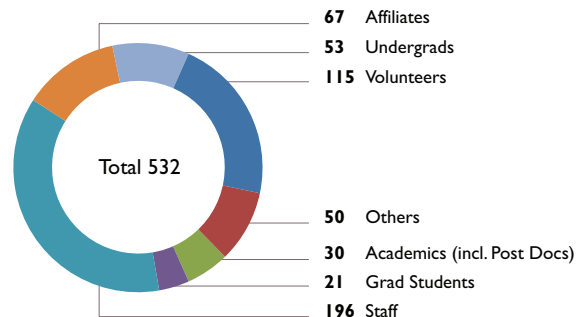
Jan Zverina
Division Director, External Relations

UC External Advisory Board

Kimberly S. Budil

Michael Carey

Trish Damkroger

Paul Dodd

Adams Dudley

Lise Getoor

Ralph Greenspan

Eamonn Keogh

Juan C. Meza

Peter Nugent

John Sarrao

Tim Sherwood

Paul Weiss

Tarek I. Zohdi

Executive Committee

UC SAN DIEGO

Sandra Brown

Mark Ellisman

Michael Holst

J. Andrew McCammon

John Orcutt

Al Pisano (chair)

Tajana Rosing

Nicholas Schork

Brian Schottlaender

Robert Sullivan

Susan Taylor

Gabriel Wienhausen

SDSC

Ilkay Altintas

Chaitanya Baru

Sandeep Chandra

Ronald Hawkins

Christine Kirkpatrick

Amit Majumdar

Michael Norman

Philip M. Papadopoulos

Nieves Rankin

Shawn Strande

Frank Würthwein

Jan Zverina

SDSC Census FY2016

Total 532

67 Affiliates
53 Undergrads
115 Volunteers
50 Others
30 Academics (incl. Post Docs)
21 Grad Students
196 Staff

RESEARCH EXPERTS

SDSC Computational Scientists

Laura Carrington, Ph.D.
Director, Performance, Modeling, and Characterization Lab, SDSC
Principal Investigator, Institute for Sustained Performance, Energy, and Resilience (DoE)
HPC benchmarking, workload analysis
Application performance modeling
Energy-efficient computing
Chemical engineering

Dong Ju Choi, Ph.D.
Senior Computational Scientist, SDSC
HPC software, programming, optimization
Visualization
Database and web programming
Finite element analysis

Pietro Cicotti, Ph.D.
Senior Computational Scientist, SDSC
Architecture
Runtime systems
Performance tools
Analysis, optimization, and modeling

Yifeng Cui, Ph.D.
Director, Intel Parallel Computing Center, SDSC
Director, High-performance GeoComputing Laboratory, SDSC
Principal Investigator, Southern California Earthquake Center
Senior Computational Scientist, SDSC
Adjunct Professor, San Diego State University
Earthquake simulations
Parallelization, optimization, and performance evaluation for HPC
Multimedia design and visualization

Andreas Goetz, Ph.D.
Co-Director, CUDA Teaching Center
Co-Principal Investigator, Intel Parallel Computing Center
Quantum Chemistry
Molecular Dynamics
ADF and AMBER developer
GPU accelerated computing

Amit Majumdar, Ph.D.
Division Director, Data Enabled Scientific Computing, SDSC
Associate Professor, Department of Radiation Medicine and Applied Sciences, UCSD
Algorithm development
Code optimization
Code profiling/tuning
Science Gateways
Nuclear engineering

Michael Norman, Ph.D.
Director, San Diego Supercomputer Center
Distinguished Professor, Physics, UCSD
Director, Laboratory for Computational Astrophysics, UCSD
Computational astrophysics

Dmitri Pekurovsky, Ph.D.
Member, Scientific Computing Applications group, SDSC
Optimization of software for scientific applications
Performance evaluation of software for scientific applications
Parallel 3-D Fast Fourier Transforms
Elementary particle physics (lattice gauge theory)

Wayne Pfeiffer, Ph.D.
Distinguished Scientist, SDSC
Supercomputer performance analysis
Novel computer architectures
Bioinformatics

Robert Sinkovits, Ph.D.
Director, Scientific Computing Applications, SDSC
High performance computing
Software optimization and parallelization
Structural biology
Bioinformatics
Immunology
Relational databases

Mahidhar Tatineni, Ph.D.
User Support Group Lead, SDSC
Research Programmer Analyst
Optimization and parallelization for HPC systems
Aerospace engineering

Igor Tsigelny, Ph.D.
Research Scientist, SDSC
Research Scientist, Department of Neurosciences, UCSD
Computational drug design
Personalized cancer medicine
Gene networks analysis
Molecular modeling/molecular dynamics
Neuroscience

Rick Wagner, Ph.D. candidate
High-performance Computing Systems Manager
Large-scale Linux-based high-performance computing clusters
Cyberinfrastructure systems architecture and design
Computational astrophysics

Ross Walker, Ph.D.
Director, Walker Molecular Dynamics Lab
Co-Director, CUDA Teaching Center
Director, Intel Parallel Computing Center
Adjunct Professor, Department of Chemistry and Biochemistry, UCSD
Molecular dynamics
Quantum chemistry
GPU accelerated computing

Nancy Wilkins-Diehr, M.S.
Associate Director, SDSC
Co-Principal Director, XSEDE at SDSC
Co-Director for Extended Collaborative Support, XSEDE
Science gateways
User services
Aerospace engineering

Frank Würthwein, Ph.D.
Distributed High-Throughput Computing Lead, SDSC
Executive Director, Open Sckeince Grid
Professor of Physics, UCSD
High-capacity Data Cyberinfrastructure
High-energy Particle Physics

SDSC Data Scientists

Ilkay Altintas, Ph.D
Chief Data Science Officer, SDSC
Director, Workflows for Data Science (WorDS) Center of Excellence
Lecturer, Computer Science and Engineering @ UCSD
Assistant Research Scientist, SDSC
Scientific workflows
Big Data applications
Distributed computing
Reproducible science
Kepler Scientific Workflow System

Michael Baitaluk, Ph.D.
Assistant Research Scientist, SDSC
Principal Investigator, Biological Networks, SDSC
Scientific data modeling and information integration
Gene networks
Systems and molecular biology
Bioinformatics

Chaitan Baru, Ph.D.
SDSC Distinguished Scientist
Director, Center for Large-scale Data Systems Research (CLDS), SDSC
Associate Director, Data Science and Engineering, SDSC
Assoc. Director, Data Initiatives, SDSC
Data management
Large-scale data systems
Data analytics
Parallel database systems

Hans-Werner Braun, Ph.D.
Research Scientist Emeritus, SDSC
Internet infrastructure, measurement/analysis tools
Wireless and sensor networks
Internet pioneer (PI, NSFNET backbone project)
Multi-disciplinary and multi-intitutional collaborations

Amit Chourasia, M.S.
Senior Visualization Scientist, SDSC
Lead, Visualization Group
Principal Investigator, SEEDME.org
Visualization and computer graphics
Ubiquitous Sharing Infrastructure

KC Claffy, Ph.D.
Director/PI, CAIDA (Center for Applied Internet Data Analysis), SDSC
Adjunct Professor, Computer Science and Engineering, UCSD
Internet data collection, analysis, visualization
Internet infrastructure development of tools and analysis
Methodologies for scalable global Internet

Alberto Dainotti, Ph.D.
Assistant Research Scientist, CAIDA (Center for Applied Internet Data Analysis)
Internet measurements
Traffic analysis
Network security
Large-scale internet events

Amogh Dhamdhere, Ph.D.
Assistant Research Scientist, CAIDA (Center for Applied Internet Data Analysis)
Internet topology and traffic
Internet economics
IPv6 topology and performance
Network monitoring and troubleshooting

Amarnath Gupta, Ph.D.
Associate Director, Academic Personnel, SDSC
Director of the Advanced Query Processing Lab, SDSC
Co-principal Investigator, Neuroscience Information Framework (NIF) Project, Calit2
Bioinformatics
Scientific data modeling
Information integration and multimedia databases
Spatiotemporal data management

Mark Miller, Ph.D.
Principal Investigator, Biology, SDSC
Principal Investigator, CIPRES Gateway, SDSC/XSEDE
Principal Investigator, Research, Education and Development Group, SDSC
Structural biology/crystallography
Bioinformatics
Next-generation tools for biology

Dave Nadeau, Ph.D.
Senior Visualization Researcher, SDSC
Data mining
Visualization techniques
User interface design
High-dimensionality data sets
Software development
Audio synthesis

Philip M. Papadopoulos, Ph.D.
Chief Technology Officer, SDSC
Division Director, Cloud and Cluster Software Development, SDSC
Associate Research Professor (Adjunct), Computer Science, UCSD
Rocks HPC cluster tool kit
Virtual and cloud computing
Data-intensive, high-speed networking
Optical networks/OptIPuter
Prism@UCSD

Andreas Prlić, Ph.D.
Technical and Scientific Team Lead, RCSB Protein Data Bank
Bioinformatics
Structural biology
Computational biology
Protein Data Bank

Peter Rose, Ph.D.
Site Head, RCSB Protein Data Bank West
Principal Investigator, Structural Bioinformatics Laboratory
Structure-based drug design
Bioinformatics
Computational biology
Protein Data Bank

David Valentine, Ph.D.
Research Programmer, Spatial Information Systems Laboratory, SDSC
Spatial and temporal data integration/analysis
Geographic information systems
Hydrology
Spatial management infrastructure

Ilya Zaslavsky, Ph.D.
Director, Spatial Information Systems Laboratory, SDSC
Spatial and temporal data integration/analysis
Geographic information systems
Hydrology
Spatial management infrastructure

Andrea Zonca, Ph.D.
HPC Applications Specialist
Data-intensive computing
Data visualization
Cosmic microwave background
Python development