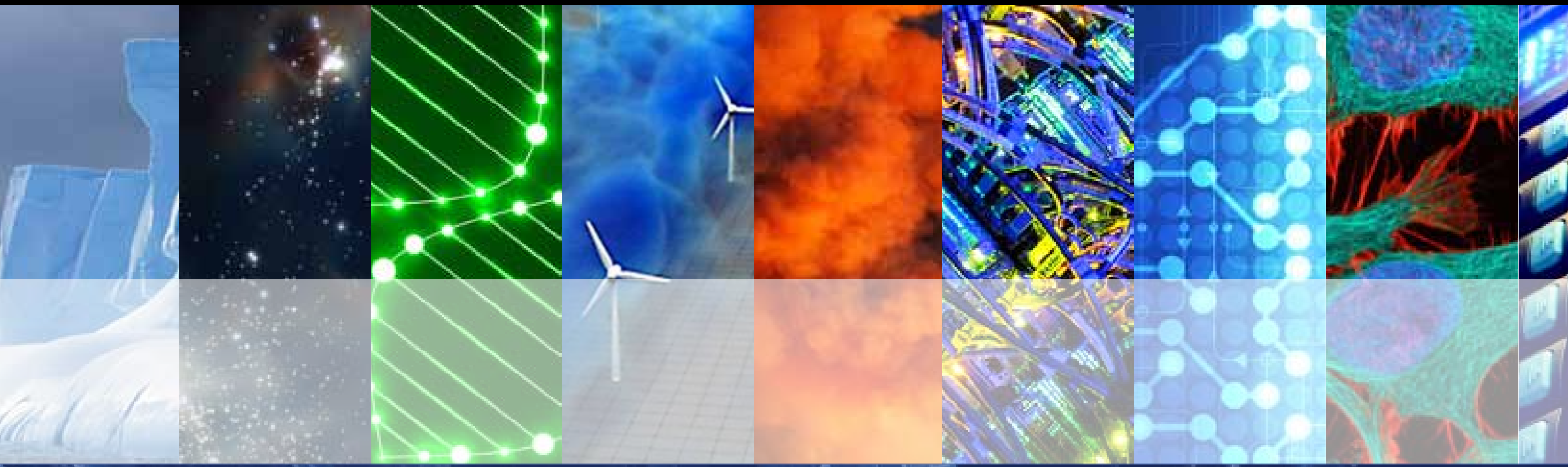
SAN DIEGO SUPERCOMPUTER CENTER at UC SAN DIEGO



GATEWAY to DISCOVERY

Tackling Today's Grand Research Challenges

SDSC

ANNUAL REPORT 2014/15

SAN DIEGO SUPERCOMPUTER CENTER

As an Organized Research Unit of UC San Diego, SDSC is considered a leader in data-intensive computing and cyberinfrastructure, providing resources, services, and expertise to the national research community, including industry and academia. Cyberinfrastructure refers to an accessible, integrated network of computer-based resources and expertise, focused on accelerating scientific inquiry and discovery. SDSC supports hundreds of multidisciplinary programs spanning a wide variety of domains, from earth sciences and biology to astrophysics, bioinformatics, and health IT. SDSC's *Comet* joins the Center's data-intensive *Gordon* cluster, and are both part of XSEDE (eXtreme Science and Engineering Discovery Environment), the most advanced collection of integrated digital resources and services in the world.

SDSC INFORMATION

Michael L. Norman, Director

San Diego Supercomputer Center
University of California, San Diego
9500 Gilman Drive MC 0505
La Jolla, CA 92093-0505
Phone: 858-534-5000

info@sdsc.edu
www.sdsc.edu

Jan Zverina
Director, Communications and Media Relations
jzverina@sdsc.edu
858-534-5111

GATEWAY TO DISCOVERY
SDSC Annual Report 2014/15

(PDF version available online at the SDSC website)

EDITOR:
Jan Zverina

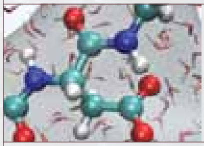
CO-EDITOR:
Warren Froelich

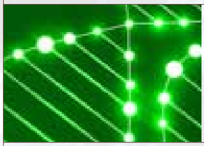
CONTRIBUTORS:
Diane Baxter, Lynne Friedmann, Julie Gallardo, Ron Hawkins, Richard Moore, Nieves Rankin, Wayne Pfeiffer, Susan Rathbun, Bob Sinkovits, Nancy Wilkins-Diehr, Nicole Wolter, Jan Zverina

CREATIVE DIRECTOR:
Ben Tolo

DESIGN:
Ben Tolo

All contents are provided as of the end of CY2014. Any opinion, conclusions, or recommendations in this publication are those of the author(s) and do not necessarily reflect the views of NSF, other funding organizations, SDSC, or UC San Diego. All brand names and product names are trademarks or registered trademarks of their respective holders.





DIRECTOR'S LETTER

Computational Convergence & Collaboration

While this annual report highlights many of our accomplishments during 2014, 2015 marks the year that the San Diego Supercomputer Center celebrates its 30th anniversary. It's gratifying to know that as we commemorate this milestone, SDSC is well positioned as a valuable resource to the entire research community in the years ahead. In fact, our mission has expanded, as you'll read below.

This year is highlighted by the launching of a new supercomputer, called *Comet*, meeting emerging requirements often referred to as the 'long tail' of science—the idea that the large number of modest-sized computationally-based research projects still represent, in aggregate, a tremendous amount of research that can yield scientific advances and discovery. Built and supported from a 2013 grant from the National Science Foundation (NSF), *Comet* provides an innovative petascale system designed to transform advanced computing by expanding access and capacity among both traditional and non-traditional research domains.

Comet joins our data-intensive *Gordon* supercomputer as a popular resource at the local, national, and even global levels. Some of the latest projects using *Gordon*, as well as our other systems, are presented in this report.

Supercomputers are only part of the equation, however. SDSC also has distinguished itself by developing and implementing one of the fastest parallel file storage systems in all of academia, *Data Oasis*, and *SDSC Cloud*, the first-ever large-scale academic deployment of cloud storage. We're also building a new range of high-speed networks across the UC San Diego campus and beyond, such as Prism@UCSD and CHERuB, capable of moving massive amounts of data—a real benefit to researchers involved in genome sequencing, climate science, and other data-intensive domains. Taken together, SDSC has worked with its partners to establish and co-manage a state-of-the-art cyberinfrastucture that is enabling scientific research like never before.

At SDSC, it's about human productivity, not just hardware productivity. It's about helping people to more easily conduct productive and robust research, from the inner workings of the brain to the outer reaches of the universe. That requires a high level of expertise on our part as science undergoes one of the most dramatic shifts since Galileo first peered through a telescope at the heavens.

For example, our new *Comet* supercomputer is particularly well-suited to supporting a large number of Science Gateways to serve a wide range of researcher needs. Such Gateways, which are a community-developed set of tools, applications, and data integrated through a web-based portal, increase human productivity by connecting researchers to many of the resources used in cutting-edge research—telescopes, seismic shake tables, sky surveys, and more—allowing them to focus on their science goals without having to know how supercomputers or other data cyberinfrastructures work. We expect *Comet* to be the leading Science Gateways cluster in the NSF's XSEDE portfolio, just like its predecessor, *Trestles*. In fact, one of the most popular gateways, the CIPRES phylogenetics portal, was developed right here at SDSC.

The NSF recognizes that a new era of science has dawned, one that revolves around cyberinfrastructure of data-enabled science and engineering. In alignment with that, SDSC's staff of 250 researchers and other professionals is making a major push toward developing integrated high-performance compute/data management environments across a variety of application domains.

UC@SDSC Engagement Program

Early on, SDSC recognized that rumbling sound as the coming deluge of digitally-generated data from every corner of academia, industry, and government. While the brightest minds have been busy trying to understand the depth and breadth of 'big data' from its management to monetization, SDSC has focused on establishing new programs, partnerships, and processes with one overriding goal in mind: Reduce the barriers to conducting the highest levels of scientific research by offering data science solutions and applications.

One such initiative is called UC@SDSC, which is in effect an expansion of SDSC's overall mission as I mentioned earlier. We are working with all other University of California campuses to identify specific collaborations where we can apply SDSC's resources, services, and expertise in data-enabled science and engineering.

Built upon three fundamental tenets—Collaboration, Education, and Innovation—we've already put in place several key initiatives that are helping UC@SDSC get off to a strong start. We formed an External Advisory Board composed of some of the brightest minds across the UC system to foster additional research collaborations between SDSC and the rest of the campus network. Our inaugural meeting was a great success, and the board is already helping us develop additional initiatives that will be of benefit UC-wide.

A related initiative is the recent appointment of UC San Diego Physics Professor Frank Würthwein as head of SDSC's Distributed High-Throughput Computing Group. Frank is an expert in high-energy particle physics and advanced computation, and his appointment paves the way for him and his team to pioneer a shared data and compute platform across the entire UC system anchored at SDSC. More about Frank's work can be found on page 36.

On the education and outreach front, we have launched SDSC's first UC Graduate Student Summer Fellowship program, providing opportunities for graduate students throughout the system to learn about SDSC's expertise and utilize the Center's wide range of resources to advance their own research. This novel program, to begin this summer, is funded by SDSC to specifically foster stronger ties with other UC campuses. Participants will gain exposure to a more diverse range of career options, gain hands-on computational experience, and add computational research scientists as essential mentors who will help them succeed in their careers.

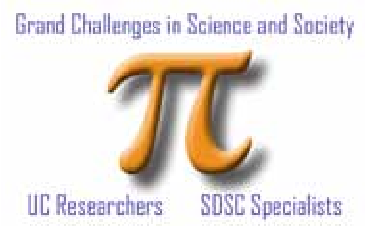
These initiatives underscore SDSC's commitment to strengthening its engagement efforts. I invite you to read more about UC@SDSC starting on page 34 of this report.

π Person of the Year

In keeping with our "people first" theme, this year's report also includes a profile and Q&A on our second annual "π Person of the Year." Why π? Because we believe that a well-rounded researcher resembles the mathematical symbol by having one leg in cyberinfrastructure technology, one leg in a scientific domain, and a bar across the top that represents his or her ability to bridge these two distinct communities with applied research.

Please join me in congratulating Wayne Pfeiffer as our latest π Person of the Year. Wayne is one of SDSC's Distinguished Scientists, and a bioinformatics researcher who has been with SDSC from the beginning—in fact he helped write the proposal that founded SDSC almost 30 years ago. Wayne is currently engaged in leading-edge research in genome sequencing, and has participated in data-intensive global projects such as the one that studied the genes of a woman who led a healthy life up until the age of 115.

In fact, one can view SDSC as a π symbol when it comes to our UC-wide engagement efforts. The two 'legs' can be considered SDSC specialists and UC researchers, connected by the top bar that represents the grand challenges of science and society.



I'm also pleased to note that SDSC has launched a completely redesigned website which reflects our strong focus on collaboration, education, and innovation while highlighting the people behind the research. We've also started a new e-newsletter called *SDSC Innovators*, providing yet another way for us to communicate the wide range of activities going on at SDSC. Like the website, SDSC Innovators focuses on the innovators behind our research and resources.

I hope that in reading through these pages, you come away with a good sense of just how much is going on at SDSC as we welcome this new era of data-enabled science and engineering, and remain focused on providing a high level of expertise to benefit researchers who require advanced computation. Please keep in touch as we celebrate our 30th anniversary later this year!

Michael L. Norman
SDSC Director

Wayne Pfeiffer

SDSC π Person of the Year

Smarter Science *for* Society

SDSC Distinguished Scientist Wayne Pfeiffer was recently named SDSC's second "Pi Person of the Year" for his research on computationally challenging problems in bioinformatics, particularly ones in genomics that can benefit from access to data-intensive supercomputers. Named after the π symbol, this award recognizes researchers who have one 'leg' in a science domain and the other in cyberinfrastructure technology.

Pfeiffer's latest research includes analyses of genomic variation within species, or even individuals and phylogenetic analyses of evolution among groups of genetically related species. With a background in computational physics and a Ph.D in engineering science from the California Institute of Technology, he began his career at General Atomics, focused on fission and fusion energy. As the longest serving SDSCer, Pfeiffer helped co-found SDSC with Sid Karin in 1985.

What parts of your career have been the most fun?

Pfeiffer: In my 30s, I did nuclear fusion research. The physics problems were really challenging, in fact so challenging that eventually I decided to do something more practical. From my experience in fusion, I had remotely used supercomputers at Lawrence Livermore National Laboratory, and helped Sid Karin write a proposal to found the supercomputer center here in San Diego. I stepped down from line management over 10 years ago and in recent years have been doing bioinformatics research.

How did you get into bioinformatics?

Pfeiffer: Six years ago I learned that SDSC Biologist Mark Miller developed software that let other biologists do phylogenetic analyses of DNA sequence data on an SDSC computer cluster via a browser interface rather than logging onto the cluster. The CIPRES portal, or gateway, looked promising but each analysis was running on only a single core. To broaden the gateway's appeal, I helped implement the analyses done by the three most popular phylogenetic codes on more powerful clusters at our sister supercomputer centers in Illinois and Texas, where each analysis could run in parallel on several cores simultaneously. This allowed more computationally demanding analyses and significantly increased usage of the gateway.

Next, I developed a hybrid parallel version of one of the codes, called RAxML, which allowed it to run on tens of cores at a time and further improved its performance. This was done via an e-mail collaboration with Alexis Stamatakis, lead developer of the code in Germany. Soon afterward, SDSC launched *Trestles* and subsequently *Gordon*, allowing us to move the CIPRES gateway analyses back to SDSC from Illinois and Texas. I have since helped make several more phylogenetic analysis codes available on our supercomputers. CIPRES has been much more successful than Mark and I ever imagined.

Could you describe some of your recent bioinformatics projects?

Pfeiffer: More than three years ago I joined a collaboration led by Henne Holstege of VU Amsterdam in which we used whole-genome sequencing to determine how many somatic mutations had accumulated in the DNA of a Dutch woman over her 115-year life. We looked for mutations in her white blood cells that were not present in her brain cells immediately after she died and reasoned that those mutations had built up from the much larger number of cell divisions in her blood as compared to her brain.

Based upon our analysis we estimated that our subject, whom we called W115, had about 450 somatic mutations in the non-repetitive genome of the white blood cells studied, corresponding to an average of four mutations per year. However, we found only four mutations that mapped to regions in genes that code for proteins, whereas most were in genomic regions predicted to have neither adverse nor favorable impact on genetic fitness. The message here is that one can have lots of somatic mutations and a long life, provided the mutations do not affect genetic fitness. Our results were published last year in *Genome Research*.

I have also been collaborating with Josephine Braun and Carmel Witte of the San Diego Zoo on a project to determine whether birds in the Zoo's collection obtain *Mycobacteria* infections from each other or from the environment. This involves looking for mutations in each bacterial sample that are not present in the others using a software pipeline that I developed and run on *Gordon*.

Any thoughts about your future at SDSC?

Pfeiffer: SDSC is as exciting a place to work as it ever has been, which is why I am still here. The Center remains a magnet for researchers needing access to powerful computational resources, and there are many good years still ahead!

Wayne Pfeiffer approaches Mount Humphreys in the Sierra Nevada.
Image credit: Betsy Pfeiffer.

SDSC's GATEWAYS TO DISCOVERY

Historically, SDSC has provided advanced computational resources and the expertise required to apply those resources to accelerate research and discovery at the local, state, and national levels while spanning academia, industry, and government.

As the Center marks its 30th anniversary in 2015, SDSC is committed to remaining at the forefront of not only high-performance computing (HPC), but in fostering new processes and partnerships to make scientific research more efficient and robust in what has become an ever-increasing collaborative community.

"Originally, HPC was for modeling and simulations that generated lots of data as its product," said SDSC Director Michael Norman. "Now it's almost the opposite, with lots of data coming in for analysis from elsewhere, such as mass spectrometers, and genome sequencing. How modern science is being conducted has informed us about what a contemporary supercomputer center should be doing and the kind of resources it should be deploying."

SDSC received back-to-back grants from the National Science Foundation (NSF) for its data-intensive *Gordon* supercomputer that came on line in 2012, followed by *Comet*, a petascale cluster designed to offer capacity and virtualized computing in which users will have more control over the software environment. *Comet* entered operations in May 2015.

"Together, these machines are redefining the very nature of scientific computing," said Norman. "But for us, it's also about human productivity, not just hardware productivity."

SDSC's Comet Launch Team. SDSC Director Mike Norman (yellow shirt) is Comet's PI.

Comet

HPC for the 99 Percent

Comet is SDSC's newest 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. The result of an NSF award currently valued at $21.6 million including hardware and operating funds, *Comet* is capable of an overall peak performance of two petaflops, or two quadrillion operations 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* is really all about providing high-performance computing to a much larger research community—what we call 'HPC for the 99 percent'—and serving as a gateway to discovery," said Norman, the project's principal investigator. "Comet has been specifically configured to meet the needs of underserved researchers in domains that have not traditionally relied on supercomputers to help solve problems, as opposed to the way such systems have historically been used."

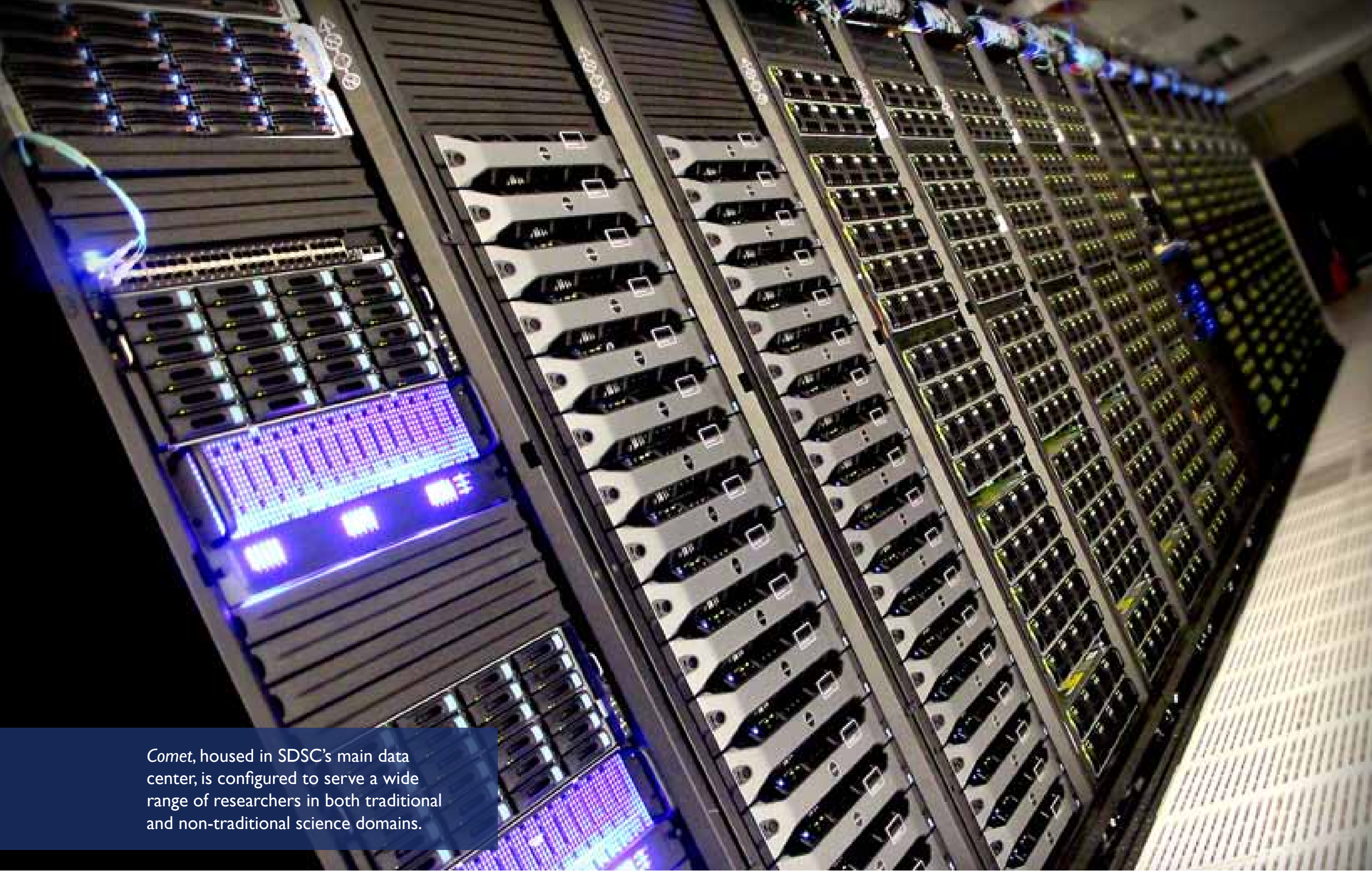
Comet is configured to provide a solution for emerging research requirements often referred to as the 'long tail' of science, which describes the idea that the large number of modest-sized, computationally-based research projects still represents, in aggregate, a tremendous amount of research and resulting scientific impact and advance.

"One of our key strategies for *Comet* has been to support modest-scale users across the entire spectrum of NSF communities, while also welcoming research communities that are not typically users of more traditional HPC systems, such as genomics, the social sciences, and economics," said Norman.

Comet is a Dell-integrated cluster using Intel's Xeon® Processor E5-2600 v3 family, with two processors per node and 12 cores per processor running at 2.5GHz. Each compute node has 128 GB (gigabytes) of traditional DRAM and 320 GB of local flash memory. Since *Comet* is designed to optimize capacity for modest-scale jobs, each rack of 72 nodes (1,728 cores) has a full bisection InfiniBand FDR interconnect from Mellanox, with a 4:1 over-subscription across the racks. There are 27 racks of these compute nodes, totaling 1,944 nodes or 46,656 cores.

In addition, *Comet* is slated to have four large-memory nodes, each with four processors and 1.5 TB (terabytes) of memory, as well as 36 GPU nodes, each with four NVIDIA GPUs (graphic processing units). The GPUs and large-memory nodes are for specific applications such as visualizations, molecular dynamics simulations, or de novo genome assembly.

Comet users will also have access to 7.6 PB (petabytes) of Lustre-based high-performance storage, with 200 GB/s (gigabytes per second) bandwidth to the cluster. It is based on an evolution of SDSC's *Data Oasis* storage system, with Aeon Computing as the primary storage vendor. *Data Oasis* will feature a new 100 Gb/s (gigabits per second) connectivity to Internet2 and ESNet, allowing users to rapidly move data to SDSC for analysis and data sharing, and to return data to their institutions for local use.



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.

By mid-2015, *Comet* will be the first XSEDE production system to support high-performance virtualization at the multi-node cluster level. *Comet's* use of Single Root I/O Virtualization (SR-IOV) means researchers can use their own software environment as they do with cloud computing, but can achieve the high performance they expect from a supercomputer.

"The variety of hardware and support for complex, customized software environments will be of particular benefit to Science Gateway developers," said Nancy Wilkins-Diehr, co-PI of the XSEDE program and SDSC's associate director. "We now have more than 30 such Science Gateways running on XSEDE, each designed to address the computational needs of a particular community such as computational chemistry, atmospheric science or the social sciences."

SDSC team members will work closely with communities and enable them to develop the customized software stacks that meet their needs by defining virtual clusters. With significant advances in SR-IOV, virtual clusters will be able to attain near-native hardware performance in both latency and bandwidth, making them suitable for MPI-style parallel computing.

Comet replaces *Trestles*, which entered production in early 2011 to provide researchers not only significant computing capabilities, but to allow them to be more computationally productive.

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 employ massive amounts of flash-based memory. That made it many times faster than conventional HPC systems, while having enough bandwidth to handle extremely large datasets. The result of a five-year, $20 million NSF grant, *Gordon* has 300 trillion bytes of flash memory and 64 I/O nodes, making the system ideal for researchers who need to sift through tremendous amounts of data.

In effect, *Gordon* is designed to do for scientific research what Google does for web searches.

By the end of 2014, 1098 research projects using *Gordon* were awarded among 762 principal investigators. Among the more noteworthy projects since *Gordon's* launch is its use in a global bird genome study published in late 2014 that has researchers rethinking how avian lineages diverged after the

extinction of the dinosaurs. The computations were done with the assistance of SDSC Distinguished Scientist Wayne Pfeiffer using a new code called ExaML (Exascale Maximum Likelihood) to infer phylogenetic trees. The four-year project, called the Avian Genome Consortium was published in the journal *Science* in December 2014. (Read more on page 23)

Another 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

In mid-2013, SDSC launched the *Triton Shared Computing Cluster (TSCC)* after recognizing that UC San Diego investigators could benefit from an HPC system dedicated to their needs, with near-immediate access and reasonable wait times when generally compared to national systems. Following an extensive study of successful research computing programs across the country, SDSC selected the "condo computing" model as the main business model for *TSCC*. Condo computing is a shared ownership model in which researchers use equipment purchase funds from grants or other sources to purchase and contribute compute "nodes" (servers) to the system. The result is a researcher-owned computing resource of medium to large proportions.

In 2014, *TSCC* had 230 users participating in the program, across 14 labs or groups, for a total of 170 nodes (approximately 3,000 processors) that is expected to increase over time, and 80+ teraflops of computing power. Participating researchers/labs cover a diverse array of fields, including engineering, computational chemistry, genomics, oceanography, high-energy physics, and more.

Also during 2014, *TSCC* became part of UC San Diego's Integrated Digital Infrastructure (IDI) program. IDI is the UC San Diego Chancellor's initiative to advance and streamline the delivery of cutting-edge IT services to campus faculty, researchers, and students in the lab. Working with all the IT providers on campus, IDI directs researchers to high-end and big data-friendly services to support research and instruction, including high-speed network connections, high-performance computing, colocation facilities, storage, tools, and training. (Read more about *TSCC* on page 48)

Learn more about the **TSCC program** by scanning the QR code on the left or by visiting http://goo.gl/wwaqL5

Learn about services provided by the **IDI program** by scanning the QR code on the right or by visiting http://idi.ucsd.edu

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

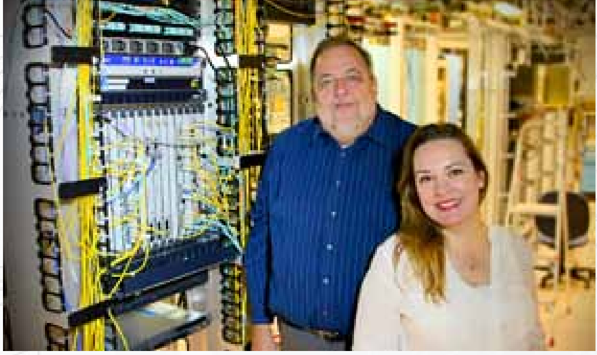
SDSC's *Data Oasis* is a Lustre-based parallel file storage system linked to *Trestles*, *Gordon*, and *TSCC*, and recently to *Comet*. As a critical component of SDSC's Big Data initiatives, *Data Oasis* currently has about 12 PB (petabytes) of capacity and speeds of up to 200 GB/s to handle just about 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 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 under a partnership between SDSC, Aeon Computing, and Intel.

SDSC Cloud

First Ever Large-scale Academic Deployment of Cloud Storage in the World

The SDSC IT Services team administers one of the first large-scale academic deployments of cloud storage in the world. 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 3 PB of raw space, which is organized into object-based cloud storage by OpenStack's Swift platform.

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.



SDSC Network Architect Thomas Hutton and Christine Kirkpatrick, SDSC's division director of IT Systems and Services, stand beside one of the campus' Juniper MX960 routers. The system, which has speeds as fast as 100 Gb/s, provides the main Internet connection to UC San Diego, SDSC, and the UCSD Medical Center.

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.

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.

Currently in production, the project is serving researchers with full functionality. "One can think of Prism as the HOV lane, whereas our very capable campus network represents the other lanes on the freeway," says Philip Papadopoulos, principal investigator on the Prism@UCSD project and SDSC's chief technology officer.

CHERuB

Connecting to the Information Superhighway

In January 2014, SDSC and UC San Diego's Administrative Computing and Telecommunications (ACT) organization were awarded a second $500,000 NSF grant to connect the campus's Science DMZ (PRISM) and SDSC's *Gordon* supercomputer to high-bandwidth national research networks to advance a new range of data-driven research. Called CHERuB for Configurable, High-speed, Extensible Research Bandwidth, the project provides 100 Gb/s (gigabit per second) connectivity, the new high-end for wide-area research networks.

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 and CENIC as well as 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 100Gb "information backbone."

The CHERuB link places UC San Diego among research universities and institutions having the highest available connectivity, with a capacity 10 times greater than existing modern data networks. CHERuB is the missing piece that will connect UC San Diego's Prism distributed science DMZ network to even faster national networks to advance scientific research. Examples of research domains that will benefit from CHERuB include cosmology, atmospheric sciences, electron microscopy, genomic sequencing, oceanography, high-energy physics, and telemedicine, all of which can encompass data-rich research and whose advancements rely on multi-site or inter-institutional activities. "The CHERuB network allows research domains that need very large datasets and dataflows 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

GLOBAL ENERGY CONSERVATION

WIND-FARM COMPUTER SIMULATIONS UNLOCK INCREASED POWER GENERATION

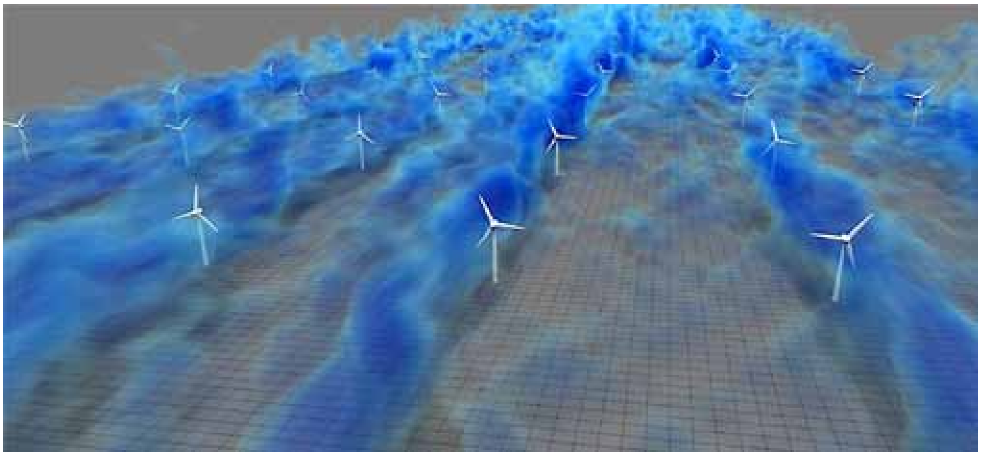
As wind energy becomes more important around the globe as a source for clean, renewable power, researchers at Johns Hopkins University (JHU) have used SDSC's *Trestles* supercomputer to run high-resolution computer simulations that take into account how the air flows within and above a wind-farm in unprecedented detail.

The study, published recently in the *Journal of Renewable and Sustainable Energy*, challenges conventional wisdom that believes the highest power output comes when wind-farm turbines are arranged in a checkerboard pattern. The simulations done using *Trestles* show that the highest power actually results when the lateral offset of turbines is such that they are just outside each other's wakes. Moreover, these detailed computer simulations also take into account the total interaction between the wind-farm and the atmosphere.

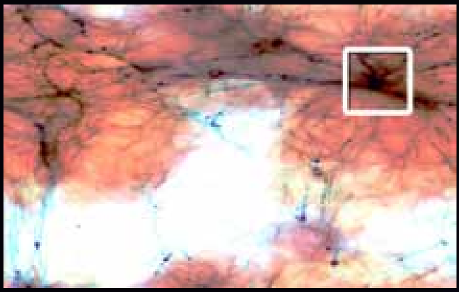
"The effect of spacing and relative positioning of the turbines on the wind-farm is crucial for good wind-farm design," said Richard Stevens, who conducted the research with Charles Meneveau and Dennice Gayme at JHU. "Wind-farm designers typically rely on simpler computer models that predict the wake effects caused by the turbines. While such models work well for smaller wind-farms, they become less accurate for larger wind-farms, where the wakes interact with one another as well as with the atmospheric wind."

After comparing the results from very detailed wind-farm simulations with existing industry models, the research team developed a new model, called the "coupled wake boundary layer model", which can more effectively predict wind-farm performance. These new results are described in a recent paper also published in the *Journal of Renewable and Sustainable Energy.*

"The benefit of high-fidelity computer simulations such as the performed using *Trestles* is that the flow in the wind-farm can be studied in detail to get insight about the main physical mechanisms that are important for the performance of very large wind-farms," said JHU's Meneveau.



Detailed simulations, which require the use of advanced supercomputers, allow researchers to study the flow in wind-farms in great detail. This image shows a visualization of the flow in a very large wind-farm obtained from a high-resolution simulation. The blue regions indicate the low velocity wind-speed regions (wakes) formed behind the turbines. Visualization by David Bock (National Center for Supercomputing Applications) and XSEDE (Extreme Science and Engineering Discovery Environment) as part of XSEDE's Extended Collaborative Support Services.



A view of the entire simulation volume that shows the large scale structure of the gas distribution in filaments and clumps. The red regions are heated by stellar UV light coming from the galaxies, highlighted in white. These galaxies are over 1000 times less massive than the Milky Way and contributed nearly one-third of the UV light during reionization. The field of view of this image is 400,000 light years across, when the universe was only 700 million years old. Image credit: John Wise, Georgia Institute of Technology.

ILLUMINATING THE HEAVENS

SDSC's Trestles Cluster Helps Shed Light on How the Faintest Galaxies Lit the Early Universe

John H. Wise, an astrophysicist at Georgia Tech, used SDSC's Trestles supercomputer to create simulations of the early universe in a collaboration with SDSC Director Michael Norman.

Using SDSC's *Trestles* supercomputer, researchers at the Georgia Institute of Technology and SDSC last year published a study showing how light from tiny galaxies more than 13 billion years ago played a larger role than previously thought in creating conditions in the universe as we know it today.

Ultraviolet (UV) light from stars in these faint dwarf galaxies helped strip interstellar hydrogen of electrons in a process called reionization, according to a study published in July 2014 by *Monthly Notices of the Royal Astronomical Society*. The epoch of reionization began about 200 million years after the Big Bang, and astrophysicists agree that it took about 800 million more years for the entire universe to become reionized. That marked the last major phase transition of gas in the universe, and it remains ionized today.

Astrophysicists, however, are not in 'universal' agreement when it comes to determining which type of galaxies played major roles in this epoch. While most have focused on larger, more luminous galaxies, this latest research, using detailed computer simulations, indicates that scientists should also focus on the smallest ones.

Specifically, these new simulations show that these tiny galaxies–despite being 1000 times smaller in mass and 30 times smaller in size than the Milky Way–contributed nearly 30 percent of the UV light during this process. Videos from the simulations, done using *Trestles* and additional systems at NASA, are at:

(left) Use QR code reader or visit https://goo.gl/x3o7sl

(right) Use QR code reader or visit https://goo.gl/HAaV19

Reionization experts often ignored these dwarf galaxies because they didn't think they formed stars. It was assumed that UV light from nearby galaxies was too strong and suppressed these tiny neighbors.

"It turns out they did form stars, usually in one burst, around half a billion years after the Big Bang," said John H. Wise, a Georgia Tech assistant professor in the School of Physics who led the study. "The galaxies were small, but so plentiful that they contributed a significant fraction of UV light in the reionization process."

"That such small galaxies could contribute so much to reionization is a real surprise," said SDSC Director Michael Norman, also a Distinguished Professor of Physics at UC San Diego and one of the study's co-authors. "Once again, the supercomputer is teaching us something new and unexpected, something that will need to be factored into future studies of reionization."

The research team expects to learn more about these faint galaxies when the next generation of telescopes is operational. For example, NASA's James Webb Space Telescope, scheduled to launch in 2018, will be able to see them.

In addition to Wise and Norman, the research team included Vasiliy G. Demchenko and Martin T. Halicek (Center for Relativistic Astrophysics, Georgia Institute of Technology); Matthew J. Turk (Department of Astronomy, Columbia University); Tom Abel (Kavli Institute for Particle Astrophysics and Cosmology, Stanford University); and Britton D. Smith (Institute of Astronomy, University of Edinburgh).

Creating New RESOURCES *for* RESEARCHERS

SDSC CREATES A NEW DATA SHARING INFRASTRUCTURE

As research has become increasingly collaborative and global, storing and sharing data across a wide range of domains likewise has become increasingly vital. In 2014, SDSC received a three-year, $1.3 million award from the National Science Foundation (NSF) to develop a web-based tool that helps scientists seamlessly share and access preliminary results and transient data from research on a variety of platforms, including mobile devices. Called SeedMe – short for 'Swiftly Encode, Explore and Disseminate My Experiments' – the new award is part of the NSF's Data Infrastructure Building Blocks (DIBBs) program that encourages development of robust and shared data-centric cyberinfrastructure capabilities to promote interdisciplinary and collaborative research.

The SeedMe project builds on the foundation made possible by an $810,000 NSF grant awarded under a separate program in late 2012 to develop a web-based architecture that focused on sharing and accessing visualization-related outputs.

"The earlier SeedMe project highlighted the need for additional capabilities and several feature requests from our current users," said Amit Chourasia, a senior visualization scientist at SDSC and principal investigator for the project. "Now we are using lessons learned from that project to make it more general and applicable to a broader research areas and researchers, while allowing SeedMe to be developed as a modular web-based, feature-rich cyberinfrastructure."

"SeedMe provides an essential yet missing component in current high-performance computing as well as cloud computing infrastructures," added SDSC Director Michael Norman, co-PI on the project. "We view this as an important building block project because computational simulations have become an indispensable tool across a very diverse array of science and engineering investigations."

Current methods for sharing and assessing transient data and preliminary results are cumbersome, labor intensive, and largely unsupported by useful tools and procedures, according to the SDSC researchers.

"Each research team is forced to create their own scripts and *ad hoc* procedures to push data from system to system, and user to user," said Chourasia. "These efforts often rely on email and other means, despite the ubiquity of much more flexible dynamic web-based technologies and the impressive display and interaction abilities of today's mobile devices.

(left) Amit Chourasia is a senior visualization scientist at SDSC and Visualization Services Group lead for the Center. Chourasia also is principal investigator for the SEEDME.org project.

(right) Andreas Goetz is co-director of the CUDA Teaching Center and co-principal investigator of the Intel Parallel Computing Center at SDSC.

SeedMe's goal is to convert a slow, manual, serial, error prone, repetitive, and redundant sharing and assessment process into a streamlined, automatable process that could be easily integrated into existing scientific applications and procedures."

SDSC DEVELOPS MULTI-SCALE SIMULATION SOFTWARE FOR CHEMISTRY RESEARCH

SDSC researchers recently developed software that greatly expands the types of multi-scale QM/MM (mixed quantum and molecular mechanical) simulations of complex chemical systems for scientists to design new drugs, better chemicals, or improved enzymes for biofuels production.

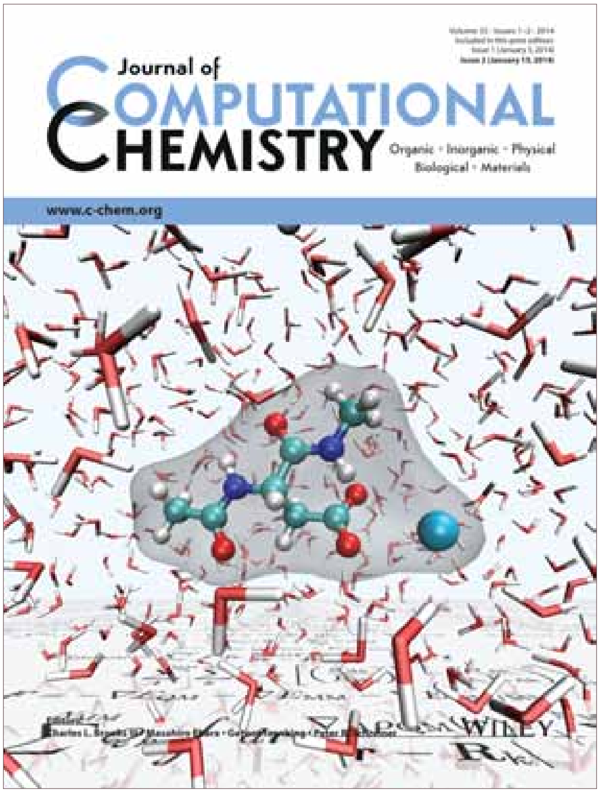
A paper outlining the research, titled 'An Extensible Interface for QM/MM Molecular Dynamics Simulations with AMBER', was featured on the cover of the *Journal of Computational Chemistry* in early 2014.

Multi-scale QM/MM computational methods are crucial to advance the understanding and solution to various challenges in the chemical sciences, ranging from drug design to renewable energies. This was recognized with the award of the 2013 Nobel Prize in chemistry for the development of multi-scale models of complex chemical systems.

In QM/MM simulations, an accurate but computationally complex and thus time-consuming quantum mechanical model, is used to identify important features of the electronic structure of a chemically relevant region. This is required, for example, to describe photo-physical processes or chemical reactions in the active site of enzymes. Effects of the surrounding environment are then included with a computationally less complex classical MM model.

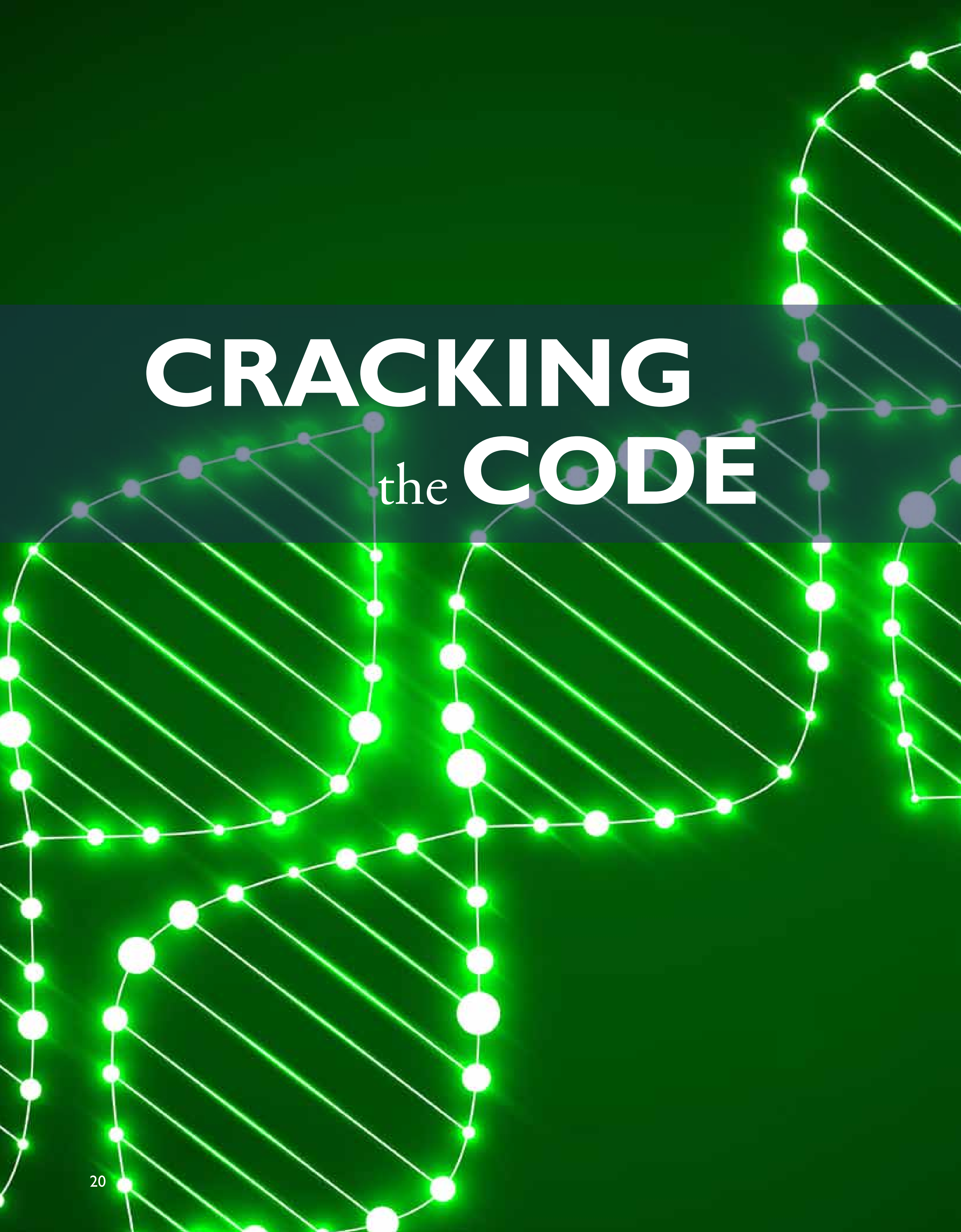
"Our software enables QM/MM simulations with a variety of advanced quantum mechanical models, and by integrating it with the popular AMBER molecular simulation package, which is used by hundreds of academic and industrial research labs, we can reach a very large user base," said Andreas W. Goetz, a research scientist with SDSC and the paper's lead author. "We're looking forward to many exciting applications that will help scientists in computational chemistry and biophysics understand and predict the behavior of molecular systems at a fundamental level."

Authors of the new study include SDSC's Goetz and Ross C. Walker, an SDSC research professor and adjunct associate professor in UC San Diego's Department of Chemistry and Biochemistry. Matthew A. Clark, who developed part of the software during his internship with Walker and Goetz, contributed to the research as part of SDSC's Research Experience for High School Students (REHS) program and later as an undergraduate research intern.



Calcium Binding
The cover shows a calcium ion coordinating to aspartate in aqueous solution, used by Andreas Goetz, Matthew Clark, and Ross Walker to demonstrate features of a new interface to electronic structure programs for *ab initio* wave function theory and DFT-based QM/MM simulations with the AMBER software package. Data exchange between the programs is implemented by means of files and system calls or the message passing interface. The QM/MM equations governing the implementation are visible on the surface that extends to the horizon.

CRACKING the CODE



(left) Wayne Pfeiffer is an SDSC Distinguished Scientist. His current research focuses on computationally challenging problems in bioinformatics, including genomic analyses.

(right) SDSC Biology Researcher Mark Miller is PI of the popular CIPRES Science Gateway.

SDSC RESOURCES AND EXPERTISE USED IN GENOMIC ANALYSES OF A 115 YEAR-OLD WOMAN

A team of researchers investigating the genome of a healthy supercentenarian since 2011 has found many somatic mutations – permanent changes in cells other than reproductive ones – that arose during the woman's lifetime.

At the time of her death at the age of 115, the subject woman called W115 by the researchers, was the second oldest person in the world and showed no signs of vascular disease or dementia. By donating her body to science, she allowed researchers to study her organs and genome.

Analyses required numerous computations, some of which were performed by SDSC Distinguished Scientist Wayne Pfeiffer under a National Institutes of Health grant. The initial analyses identified thousands of putative somatic mutations, many of which were incorrect because of sequencing errors, according to Pfeiffer. Filters were subsequently developed to select the mutations most likely to be somatic.

"The message here is that one can have lots of somatic mutations and still live long, provided the mutations do not affect genetic fitness," said Pfeiffer.

Led by Erik Sistermans and Henne Holstege from the VU University Medical Center in Amsterdam, the researchers recently published their findings in the journal *Genome Research* as reported by *GenomeWeb*. SDSC's *Triton* compute cluster, since superseded by the *Triton Shared Computing Cluster (TSCC)*, assisted in the research.

While previous studies have examined mutations that arise in certain disease conditions such as leukemia, Sistermans said that it was not well known how many mutations might appear in the genomes of healthy cells, according to the *GenomeWeb* report.

The researchers hypothesized that white blood cells, which divide frequently, would have many more somatic mutations than brain cells, which seldom divide. Thus the whole genomes of W115's blood and brain cells were sequenced using SOLiD technology from Life Technologies. Analyses were then done to look for mutations present in the blood cells but not the brain cells.

Two types of mutations were considered: single nucleotide variants (SNVs) and short insertions or deletions (indels). Filtering of the latter was particularly compute-intensive and was done at SDSC. Thousands of core hours were consumed, and some steps required more than 64 GB (gigabytes) of shared memory. After filtering, many of the highly likely and moderately likely somatic mutations were tested by targeted sequencing using newer Ion PGM sequencers, also from Life Technologies.

Based on these validation tests, the researchers estimated that there were about 450 somatic mutations in the non-repetitive genome of the white blood cells studied, corresponding to an average of four mutations per year. These mutations, they noted, were not present in the breast cancer that W115 had at age 100 or in the gastric tumor she had at the time of her death. About 95% of the somatic mutations were SNVs rather than indels.

"Of 376 highly likely somatic SNVs, only four mapped to regions in genes that code for proteins, whereas most were in genomic regions predicted to have neither adverse nor favorable impact on genetic fitness," said Pfeiffer.

"It is important to note that white blood cells differ from most other cells in the body and are especially prone to acquiring somatic mutations," said SDSC Researcher Mark A. Miller, who helped interpret the results. "Large numbers of white blood cells are generated from relatively few hematopoietic stem cells. Because white blood cells divide continually throughout a person's lifetime, it is possible for non-harmful somatic mutations to accumulate."

The researchers concluded that there is a significant somatic mutation background among white blood cells, even in healthy blood.

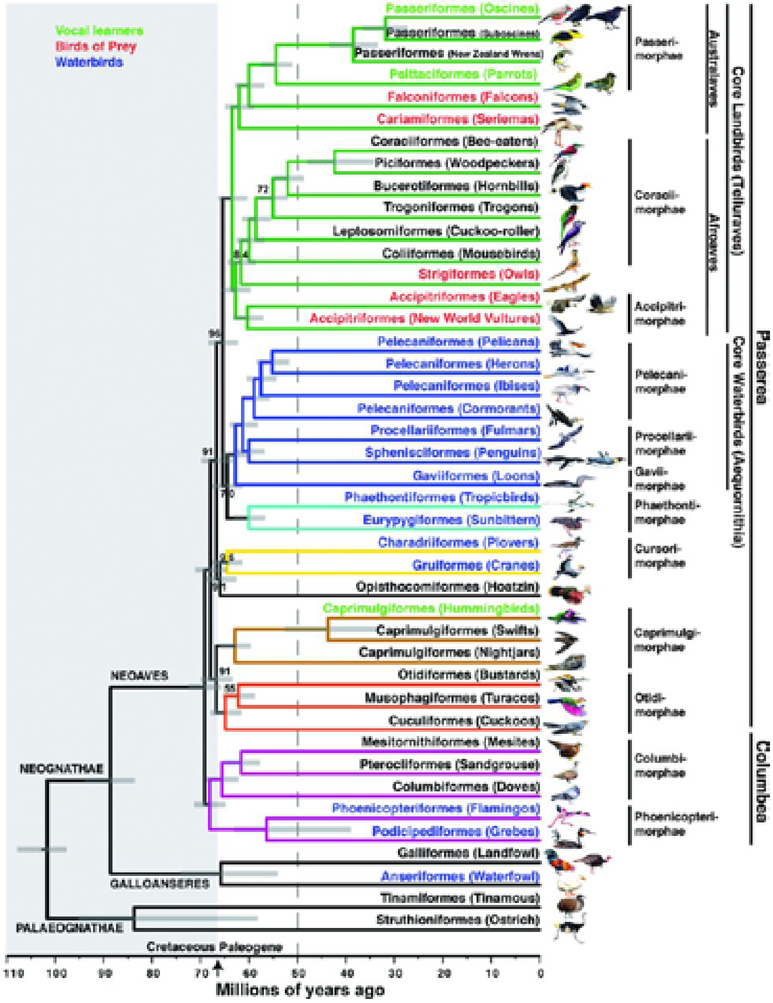
"These mutations accumulate in clones that comprise only some of the white blood cells, which makes their detection more difficult," explained Pfeiffer. "For W115, about 64% of the white blood cells comprised a dominant clone, which was where the mutations were found, while about 44% of the white blood cells were in a second clone subsidiary to the larger one. The remaining cells were presumably in smaller clones that were below our detection limit."

SDSC's *Gordon* supercomputer (below) coupled with newly developed Exascale Maximum Likelihood (ExaML) code played a major role in creating the most reliable tree of life for birds to date. The new avian family tree (below right) clarifies how modern birds emerged following the mass extinction of the dinosaurs some 66 million years ago.
Image credit: Erich D. Jarvis, HHMI.

RESEARCHERS RETHINK HOW OUR FEATHERED FRIENDS EVOLVED

A recently published global genome study that used SDSC's data-intensive *Gordon* supercomputer has researchers rethinking how avian lineages diverged after the extinction of the dinosaurs.

The four-year project, called the Avian Genome Consortium and published in the journal *Science* in late 2014, resulted in a new family "tree" for nearly all of the 10,000 species of birds alive today by comparing the entire DNA codes (genomes) of 48 species as varied as parrot, penguin, downy woodpecker, and Anna's hummingbird.



The massive undertaking, started in 2011, involved more than 200 researchers at 80 institutions in 20 countries, with related studies involving scientists at more than 140 institutions worldwide.

The genome-scale phylogenetic analysis of the 48 bird species considered approximately 14,000 genes. This presented computational challenges not previously encountered by researchers in smaller-scale phylogenomic studies based on analyses of only a few dozen genes. The inclusion of hundreds of times more genetic data per species allowed the researchers to realize the existence of new inter-avian relationships.

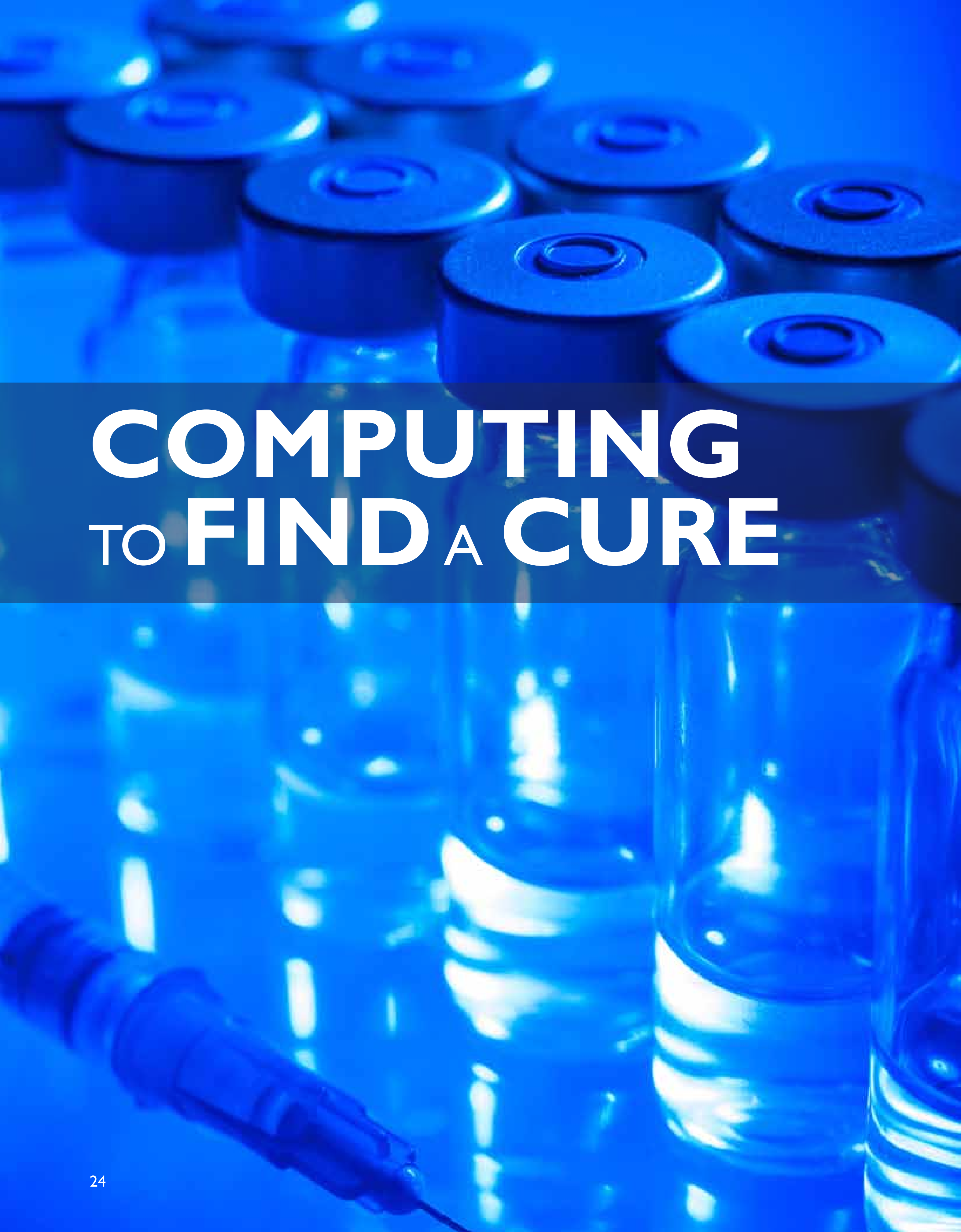
"Characterization of genomic biodiversity through comprehensive species sampling has the potential to change our understanding of evolution," wrote Erich Jarvis, associate professor of neurobiology at the Howard Hughes Medical Institute at Duke University and the study's principal investigator, in an introduction to a special issue of the journal *Science* containing eight papers from the study. An additional 20 papers generated by the study were simultaneously published in other journals.

"For 50 species, more than 10 to the power of 76 possible trees of life exist. Of these, the right one has to be found," said Andre J. Aberer, with the Heidelberg Institute for Theoretical Studies (HITS), in a news release at the time of the study's publication in *Science*. "For comparison: About 10 to the power of 78 atoms exist in the universe."

Many of the computations were done on *Gordon* by Aberer with the assistance of SDSC Distinguished Scientist Wayne Pfeiffer. They ran a new code called ExaML (Exascale Maximum Likelihood) to infer phylogenetic trees using *Gordon* soon after it debuted in 2012 as one of the 50 most powerful supercomputers in the world.

More than 400,000 core hours of computer time were consumed on *Gordon*. "After doing initial analyses on our institutional cluster, we rapidly realized that comprehensive analysis of the more challenging data sets being considered would require supercomputer resources," said Aberer. "Access to *Gordon* was thus invaluable for achieving results in a timely manner."

COMPUTING TO FIND A CURE



As a senior research scientist at SDSC, Julia Ponomarenko was UC San Diego's principal investigator for the Immune Epitope Database (IEDB).

SDSC ASSISTS IN THE SEARCH FOR EBOLA IMMUNE RESPONSE TARGETS

The effort to develop therapeutics and a vaccine against the deadly Ebola virus disease (EVD) requires a complex understanding of the microorganism and its relationship within the host, especially the immune response. Adding to the challenge, EVD can be caused by any one of five known species within the genus Ebolavirus (EBOV), in the Filovirus family.

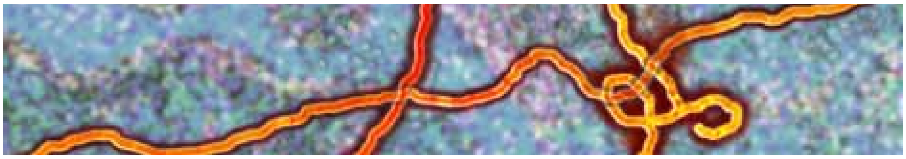
Researchers at SDSC and the La Jolla Institute for Allergy and Immunology (La Jolla Institute) have been assisting the scientific community by running, since August 2014, high-speed online publications of analyses of EBOV-related epitope data being curated in the Immune Epitope Database (IEDB) and predicting epitopes—the viral molecules recognized by the human immune system—IEDB Analysis Resource. Sebastian Maurer-Stroh, of Bioinformatics Institute, a member of A*STAR's Biomedical Sciences Institutes, also assisted with analysis of the latest outbreak sequences of Ebola proteins.

"We showed in our recent publication in *PLOS Currents Outbreaks* that protective mAb (nonoclanal antibody comprising therapeutic cocktail, such as ZMab, ZMapp, and MB-003 epitopes were highly conserved in the Ebolavirus glycoprotein sequences spanning all Ebola virus lineages since 1976, with only one immunodominant epitope of mAb 13F6-1-2 acquiring two novel mutations in the 2014 outbreak that might potentially change the antibody specificity and neutralization activity," said Julia Ponomarenko, a senior research scientist at SDSC and UC San Diego's principal investigator for the IEDB.

These results brought to UC San Diego the National Science Foundation's Ebola Rapid Response Research (RAPID) award for 2015. (Ponomarenko since transferred the award to Sergei Pond, an assistant professor in the Divisions of Infectious Diseases and Biomedical Informatics in UC San Diego's Department of Medicine in early 2015 when Ponomarenko was appointed head of the Bioinformatics core at the Center for Genome Regulation in Barcelona, Spain).

While outbreaks of EVD have occurred in Africa in the past, 2014's epidemic, caused by Zaire Ebolavirus, has been characterized by its unprecedented breadth and rapid spread.

"Clearly, research related to development of therapeutics and a vaccine against EVD is an urgent need, as well-engineered vaccines don't exist at this time," said Alessandro Sette, with the Division of Vaccine Discovery at the La Jolla Institute. "Our analysis is aimed at assisting the clinical and scientific communities in fine evaluation of laboratory results with the express intent of improving therapeutic targets or new vaccine development."



Igor F. Tsigelny is a research professor at the Department of Neurosciences, San Diego Supercomputer Center, and Moores Cancer Center. He is a world-reknowned expert in structural biology, molecular modeling, bioinformatics, and structure-based drug design.

HOMING IN ON ALZHEIMER'S DISEASE

UC San Diego researchers analyzing peptides using SDSC's data-intensive *Gordon* supercomputer have found new ways to elucidate the creation of the toxic oligomers associated with Alzheimer's disease, creating new targets for future drug development.

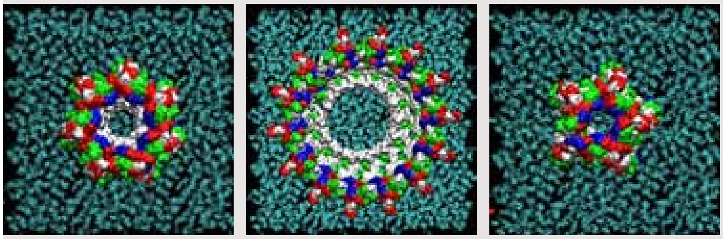
The cross-campus team surveyed all the possible ways to look at the dynamics of conformational changes of these peptides, and the possibility that they might organize into the oligomers theorized to be responsible for the degenerative brain disease. Their findings were published in the February 14, 2014 issue of the *Journal of Alzheimer's Disease*.

Igor Tsigelny, a research scientist with SDSC, the UC San Diego Moores Cancer Center, and the Department of Neurosciences, focused on the small peptide called amyloid-beta, which pairs up with itself to form dimers and oligomers.

"Our research has identified amino acids for point mutations that either enhanced or suppressed the formation and toxicity of oligomer rings," said Tsigelny, the study's lead author. "Aggregation of misfolded neuronal proteins and peptides may play a primary role in neurodegenerative disorders, including Alzheimer's disease."

Tsigelny also noted that recent improvements in computational processing speed have allowed him and other researchers to use a variety of tools, including computer simulations, to take new approaches to examining amyloid-beta, which has proven too unstable for traditional approaches such as x-ray crystallography.

The researchers investigated the single and dimer forms of the peptide with a combination of computational methods including molecular dynamics, molecular docking, molecular interactions with the membrane, as well as mutagenesis, biochemical, and electron microscopy studies. They then looked at how those dimers interacted with additional peptides and which larger structures resulted. The researchers found that depending on their configurations, some dimers did not lead to any further oligomerization, and some formed toxic oligomers implicated in the development of Alzheimer's disease.



Single amyloid-beta monomers can pair up to form a variety of dimers that can aggregate into larger peptide rings that reside on cell membranes such as those pictured. This process has been implicated in the development of Alzheimer's disease. This visualization shows the possible rings which have the most favorable energies of interactions with the membrane. The residues are colored white to represent apolar or hydrophobic areas, green for the polar or hydrophilic areas, blue to show a positive charge, and red to show a negative charge. Image credit: Igor Tsigelny, SDSC/UCSD.

"Remarkably, we showed a greater diversity in amyloid-beta dimers than previously described," said Eliezer Masliah, professor of pathology and medicine at UC San Diego, and a member of the research team. "Understanding the structure of amyloid-beta dimers might be important for the design of small molecules that block formation of toxic oligomers."

Based on their results, the researchers were able to identify key amino acids that altered the formation and toxicity of oligomer rings. "Our data is only theoretical, but there is a good chance the oligomers we have been modeling exist for real," noted Masliah.

According to the researchers, their work implicates a more dynamic role for the amyloid-beta dimers than previously thought. It also suggests that the way dimers form and then grow into larger structures is a rapidly changing process.

Masliah and Tsigelny's collaborators included UC San Diego researchers Yuriy Sharikov, Valentina Kouznetsova, Jerry Greenberg, Wolfgang Wrasidlo, Tania Gonzalez, Paula Desplats, Sarah E. Michael, Margarita Trejo-Morales, and Cassia Overk.

IMPACT AND INFLUENCE

LOCAL IMPACT AND INFLUENCE

In 2014, UC San Diego adopted a new strategic plan as part of a broader initiative to help assure the university's standing as one of the nation's top academic-based research institutions. That strategic plan includes four "grand research" themes that focus on areas in which UC San Diego has deep expertise:

- **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**

As part of its efforts to forge closer ties across campus as a resource and service provider, SDSC last year established or joined several new partnerships to accelerate research within several of these campus-wide themes, as well as foster innovative education and training programs. On the research front, SDSC's dedication to harnessing 'big data' and finding ways to extract meaning and value from voluminous amounts of digitally-based information has been a key attraction for UC San Diego researchers, as well as those in local industry and government.

Geisel Library, UC San Diego

INTEGRATED DIGITAL INFRASTRUCTURE

SDSC is a major participant in UC San Diego's Integrated Digital Infrastructure (IDI) program, an evolution of the former Research CyberInfrastructure (RCI) program. IDI is a campus-centric initiative designed to support the campus strategic plan and researcher needs by:

- Creating user-driven teams of technical specialists to work with researchers to take full advantage of the IDI resources.
- Establishing a 'big data freeway' at UC San Diego that connects campus researchers working with massive data sets across campus, including many storage facilities in SDSC's colocation facility, and beyond. A lynchpin in this is PRISM@UCSD, a research network that is connecting campus labs to central facilities at the California Institute for Telecommunications and Information Technology (Calit2) and SDSC. In addition, the CHERuB program is enabling 100 Gb/s (gigabits per second) connectivity for the UC San Diego campus to CENIC and other networks.
- Running the Regional Datacenter facility at SDSC, providing energy-efficient colocation space for UC research needs.
- Operating the *Triton Shared Computing Cluster (TSCC)* for campus and UC users, with both condo and on-demand access modes. (See page 48 for more information.)
- Supporting curation efforts and the Research Data Library, a resource to provide a full-spectrum of curation of, and long-range storage, for valuable research data collections.

SDSC will play a key role in these IDI initiatives as a provider of resources and expertise. The Center leads both the *TSCC* and co-location programs, while SDSC Director Michael Norman is PI (principal investigator) of the CHERuB/Administrative Computing & Telecommunications (ACT) collaboration and SDSC Chief Technical Officer Philip Papadopoulos is PI of the PRISM/Calit2 partnership.

SDSC is involved with several other collaborations, both on and off campus, in areas ranging from early education to programs that help ensure the safety of residents throughout greater San Diego.

A wireless education and safety network for science and society

HPWREN, the High-Performance Wireless Research and Education Network, provides high-speed Internet access to field researchers from disciplines including geophysics, astronomy, and ecology, educational opportunities through connections to learning centers in several communities, and advanced warning and monitoring systems to firefighters in distant sections of San Diego County. HPWREN consists of a collaboration of researchers at SDSC, the Scripps Institution of Oceanography's Institute (SIO) of Geophysics and Planetary Physics, UC San Diego's California Institute for Telecommunications and Information Technology (Calit2) Qualcomm Institute, Caltech's Palomar Observatory, San Diego Gas & Electric, and various San Diego firefighting agencies including CAL FIRE.

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 to 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).

Facilitating local & state-wide collaborations

SDSC helps coordinate UC CRO (University of California Collaborative Research Opportunities), an effort to facilitate collaboration between SDSC researchers and other UC faculty in preparing research proposals leveraging computational and data science expertise.

© 2015 University of California

SCALING UP COMPUTATIONAL BIOLOGY

Rommie Amaro is director of UC San Diego's National Biomedical Computation Resource (NBCR).

The UC San Diego's National Biomedical Computation Resource (NBCR) in 2014 received $9 million in funding from the National Institutes of Health (NIH), permitting the NBCR to continue connecting biomedical scientists with supercomputing resources and emerging information technologies.

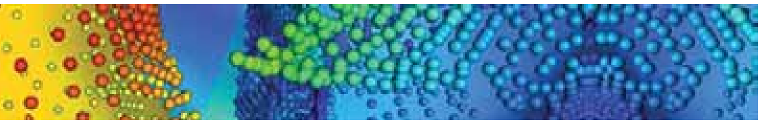
The five-year grant involves faculty from SDSC, UC San Diego's Physical Sciences, School of Medicine, and Jacobs School of Engineering, as well The Scripps Research Institute (TSRI), a private, non-profit research organization.

Biomedical computation, which applies physical modeling and computer science to the field of biomedical sciences, is often a less expensive alternative to traditional experimental approaches, and can speed the rate at which discoveries are made for a host of human diseases and biological processes.

"Our main effort remains focused on making connections across diverse scales of biological organization," explained NBCR Director Rommie Amaro. "As scientists, we ultimately need to connect across three or four scales in order to model and understand complex biological phenomena from the molecular level minutia all the way up to the whole organ."

The project also involves creating scientific workflows to manage the size and diversity of biomedical data. Ilkay Altintas, SDSC's Director of Workflows for Data Science (WorDS) center of excellence at SDSC and an affiliate of university's Qualcomm Institute, is developing workflow tools for NBCR that can help researchers tackle tough research problems such as reproducibility, which arise when research is conducted across scales.

Also crucial to NBCR's mission is the work of SDSC Chief Technical Officer and Program Director Phil Papadopoulos, who created a user-driven, software-based framework for research teams to share significant quantities of data—rapidly, securely and privately—across geographic distance and computing systems.



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

MASTERING DATA SCIENCE AND ENGINEERING

At the higher education level, SDSC is playing a key role in a newly established master's degree (MAS) program in Data Science and Engineering, intended for working professionals with a broad educational background and/or training in computer science, engineering, or mathematics.

Started in the fall of 2014, the Data Science and Engineering master's degree program is being taught by world-renowned professors and researchers from the Department of Computer Science and Engineering (CSE) in the UC San Diego Jacobs School of Engineering, in collaboration with SDSC, where students engage with the Center's technical staff and participate in a capstone project leveraging SDSC's unique resources and expertise.

"Each day the world is becoming not only more digitized, but how we share data covering all aspects of life is easier and more immediate than ever before," said SDSC Director Michael Norman. "Managed well, this data can be used to unlock new sources of economic value and provide fresh insights into scientific discovery."

The Data Science and Engineering program is structured to let students complete the MAS degree in two years. The curriculum consists of seven required courses (three foundational courses and four core courses), two electives chosen from six course options, and a two-quarter capstone project course, for a total of 38 units.

To read the full press release,
use a QR code reader or visit
http://goo.gl/5w5fr4

CALLING ALL GIRLS!

As associate director of education at SDSC, Diane Baxter has been focused on introducing computational sciences and computational thinking skills to students and teachers in regional, national, and international settings.

SDSC's StudentTECH and TeacherTECH programs have long been recognized within the local community as promoting computer science at the K-12 levels with a variety of programs that include summer internships, year-round seminars and workshops, and a wide variety of "train the teacher" programs.

New for 2014 is GirlTECH San Diego, a non-profit collaborative community program launched in partnership with other local universities and industry support groups, to encourage and educate young women to learn and apply computing skills. To date the partnership includes UC San Diego, San Diego State University, the University of San Diego, and Point Loma Nazarene University.

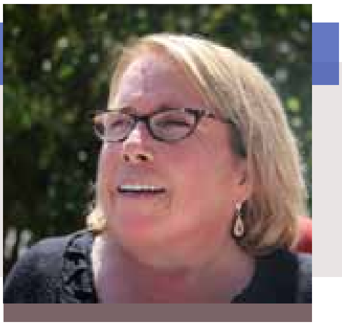
> *"Some young women lack interest because they don't realize that computing will empower them in any field they pursue, and those with an interest in computing don't necessarily pursue greater skill development because it's either not available at their school or because they lack self-confidence to participate in what is perceived as a male-oriented geek environment."*
>
> --Diane Baxter, Associate Director of Education, SDSC.

The GirlTECH program was created in response to several recent statistics, including these figures published in a U.S. Department of Workforce Readiness publication:

- **Some 90 percent of high schools in the U.S. do not teach computer science. In many other countries, computer science courses are required.**
- **Women hold less than 25 percent of STEM (science, technology, engineering, and math) jobs.**
- **Only 19 percent of students enrolled in Advanced Placement Computer Science courses are female.**
- **Software jobs outnumber qualified applicants three-to-one. The gap is one million jobs over 10 years, and these are some of the highest paying jobs.**

GirlTECH San Diego extends the institutions' teaching curriculums beyond the classroom and provides a stimulating environment for students seeking more hands-on involvement, while providing a future source of computer science employees to technology-oriented companies both locally and nationally.

To read the full press release, use a QR code reader or visit http://goo.gl/OHni6y



FROM MENTORING THE NEXT GENERATION...

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.

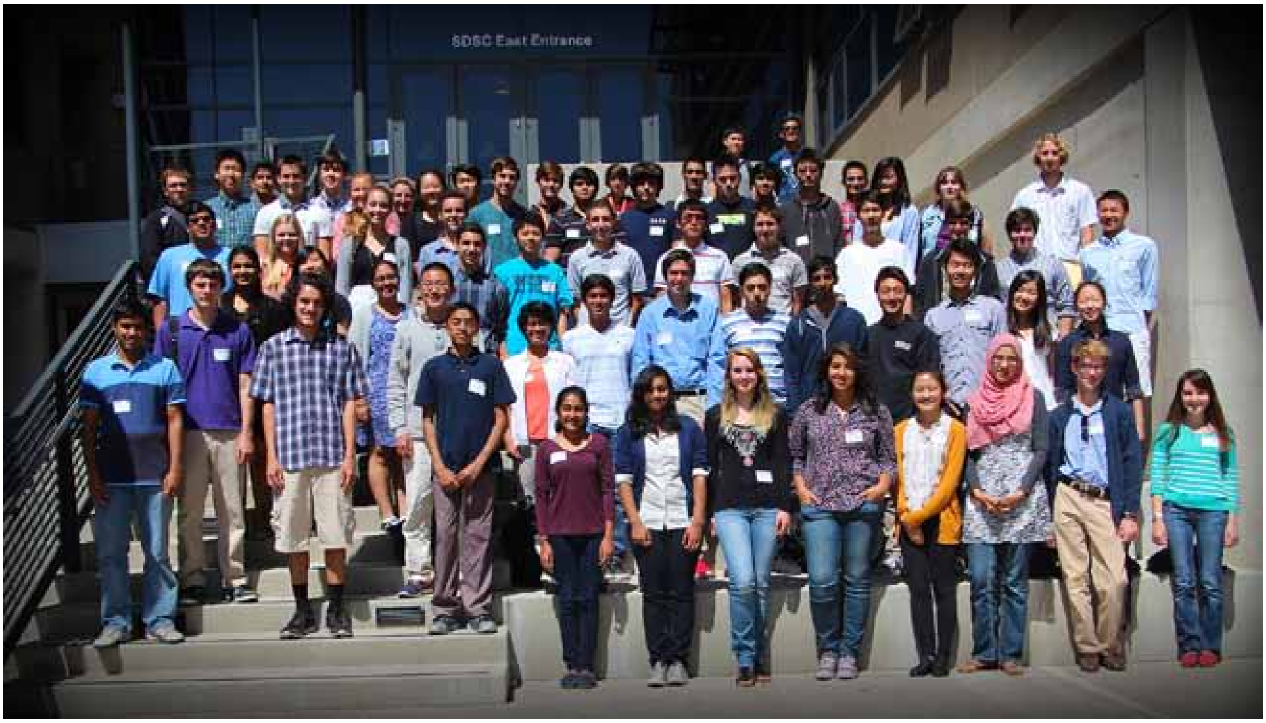
In 2014, SDSC completed it most successful and well-attended Research Experience for High School Students (REHS) summer program, where students join multidisciplinary research teams and staffers at the Center to gain experience across a wide array of computational research.

"Thanks to an expanded selection of internships and broad participation from the SDSC staff, we had a total of 60 high school students enrolled—almost twice the number of students from the previous year's program," said Ange Mason, SDSC's education program manager. "It also has become quite challenging for us to select students from the more than 200 applications we received, because collectively, these students have an amazing diversity of backgrounds and interests in all areas of computational science."

REHS is intended to serve as a stepping stone for those students who are considering a computational science curriculum as a major or minor when they enter college. Subject areas included working with SDSC researchers in the areas of molecular dynamics software development, advancing drug designs for Parkinson's Disease, understanding and managing large data sets and scientific workflows, working on projects related to predictive analytics, developing reliable network and information technology infrastructures, and learning how to effectively communicate and publicize research projects and their results.

"I learned the basics of how 3D printing works, and was already creating my own 3D designs," said participant Sarah Hempton, who attends the Halstrom Academy. "I really appreciated the mentoring as we explored this technology and others."

"I viewed this internship as a kind of sandbox for me," said Olivia Palid, who attends the Academy of Our Lady of Peace, and chose science communications. "There's a kind of super coolness to computing if one gets into the coding and programming end. For now I want to hone my writing skills, especially when it comes to effectively communicating conclusions."



...TO TAKING HOME THE GOLD

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

Canyon Crest Academy student Ezra Kosviner recently won first place in not one, but two science fair competitions for his biochemistry project that involved creating an accurate 3D model of Naeglariapore, a potentially powerful antimicrobial.

Kosviner, a high school senior, won first place in the biochemistry section of both the 2014 San Diego Science and Engineering Fair and the California State Science Fair. His research project was part of a scientific research elective course at school called Quest.

"Three-dimensional structures of proteins are important because the 3D structure largely determines the protein's mechanisms of action," said Peter Rose, an SDSC researcher who mentored Kosviner. "These models are also important because when one actually looks at a protein, one can find drug-able pockets, which are areas that a molecule can be placed to alter the function of the protein."

Rose also is the Scientific Lead for the RCSB Protein Data Bank project (www.rcsb.org), which provides access to the PDB, the single worldwide repository for the three-dimensional structures of proteins and nucleic acids.

Earlier, another senior at Canyon Crest Academy won a trifecta of student-level science competitions after being mentored by two UC San Diego professors in a project that combined supercomputer modeling with experimental research to speed up the discovery of influenza virus inhibitors. In all, Eric Chen was awarded $250,000 in prize money after winning the 2014 Intel Science Talent Search; the 2013 national Siemens Competition in Math, Science & Technology; and the grand prize in the international Google Science Fair.

Chen won the awards for a project that combined chemistry, biology, and computer modeling to find compounds capable of blocking the activity of an enzyme called endonuclease that all flu viruses need to reproduce and spread.

Chen received his training in computational biology at UC San Diego, where he used resources at SDSC, including the *Gordon* supercomputer, to run molecular dynamics computations as part of the BioChemCoRe outreach program organized by Rommie Amaro, director of UC San Diego's National Biomedical Computational Resource (NBCR). Amaro and Gen-Sheng Feng, a professor of pathology in the UC San Diego School of Medicine, mentored Chen.

(left) Erik Kosviner

(below) Eric Chen and UC San Diego Professor Rommie Amaro.
Image credit: Eric Jepsen UC San Diego Publications

STATE IMPACT AND INFLUENCE

Aligning with Principles and Partnerships

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

Although founded by the National Science Foundation (NSF) almost three decades ago as one of the nation's first academic supercomputer centers, SDSC's mission has since evolved and expanded to include the creation and fostering of collaborations across the University of California system. As part of that effort, SDSC has worked to align itself with three UC principles when it comes to UC-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 all essential ingredients to stimulating collaboration across the UC system, whether it is finding ways to better predict the impact of earthquakes and wildfires or developing new drugs to combat often debilitating diseases.

Confronting such societal challenges requires collaboration among researchers who have the scientific vision, technological skill, and innovative approaches to advance discovery. To that end, 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. While this initiative will be significantly expanded in 2015 and beyond, several elements were launched during 2014:

UC Collaborative Research Opportunity

This program lets SDSC experts interested in collaborating with UC researchers apply for Collaborative Research Opportunity (CRO) mini-grants to support collaborative work, leading to extramural grant proposals. In addition to the strategic 'top down' CRO 'LHC@UC' initiative (see feature story on page 36), SDSC already has several PI-initiated CRO awards in the areas of seismic simulations with UC Riverside, a GPU (graphics processing unit) cluster involving UC Irvine and UC Riverside, a Pacific Research Platform project (see right column) with UC Berkeley, and a brain research collaboration with UC Los Angeles and UC San Francisco, among other universities.

'Made in UC'

This project focuses on the cataloguing of 'Made in UC' software and technologies that specifically relate to data science and computational science. SDSC is partnering with UC researchers on benchmarking various technologies and extending software so that they can be run on different types of clusters and scale to big data problems. Where possible, SDSC is working to create 'Made in UC' bundles, e.g. running Berkeley's AMPLab-produced BDAS stack incorporating the UCR Suite, and UCSC's Probabilistic Soft Logic (PSL) on UCI's AsterixDB or UCM's GLADE platforms. This project has already spawned several collaborative brainstorming sessions between SDSC PIs and researchers at UCI, UCM, UCR, and UCSC.

Pacific Research Platform

SDSC has made significant contributions to the Pacific Research Platform (PRP), which engages all UC campuses including LBNL as well as other universities across California and the western region. The immediate objective of the PRP, recently funded by the National Science Foundation (NSF), is to develop a "regional Science DMZ" across the regional partners that opens up the capabilities of high-performance networking to advance data transfers and scientific collaborations for researchers across all scientific domains. This initiative is receiving high visibility across all campuses at the CIO and VCR levels. Working with Larry Smarr, Director of Calit2, SDSC staff are active contributors, and participated in a recent high-profile demonstration of this capability at the CENIC 2015 conference. SDSC Deputy Director Richard Moore wrote a proposal for an NSF-sponsored summer workshop on this initiative, with Larry Smarr as PI. In addition, Moore joined the UC IT Leadership Council, representing the IT requirements and solutions for UC system-wide researchers. Phil Papadopoulos, SDSC's Division Director of Cloud and Cluster Software Development, is a co-PI on the project and will coordinate the efforts of the large group of network engineers, network providers, and measurement programmers at the PRP institutions. Frank Würthwein, Distributed High-Throuput Computing Lead at SDSC and a UC San Diego physicist, is also a co-PI. Würthwein will lead technical development of the application groups and monitor progress from the scientists' perspective.

SDSC Summer Institute

The SDSC Summer Institute is an ideal program for UC researchers who would like to become more familiar with advanced computation as it relates to data management, running jobs on SDSC resources, reproducibility, database systems, and other techniques for turning data into knowledge. Participants are scheduled to receive hands-on training using SDSC's *Gordon* and *Comet* clusters. SDSC HPC Specialist Andrea Zonca has already been assisting colleagues at UC Irvine, UCLA, and Lawrence Berkeley Labs in becoming familiar with a range of computational tools, and will be assisting in the planning and implementation of the 2015 SDSC Summer Institute.

LHC@UC: UC SAN DIEGO PHYSICIST FRANK WÜRTHWEIN JOINS SDSC

Frank Würthwein, a noted expert in high-energy particle physics and advanced computation, has joined SDSC as head of the Center's Distributed High-Throughput Computing Group to develop and deploy a high-capacity data cyberinfrastructure across all UC campuses based on his involvement in processing massive data sets associated with the Large Hadron Collider (LHC).

"Frank's expertise paves the way for him to pioneer a shared data and compute platform, anchored at SDSC, across the entire UC system," said SDSC Director Michael Norman. "His appointment is just one of many ways we are committed to strengthening our UC engagement efforts."

A UC San Diego physics professor since 2003, Würthwein was recently named executive director of the Open Science Grid (OSG) project, a multi-disciplinary research partnership funded by the U.S. Department of Energy and National Science Foundation. He was one of OSG's founding executives during 2005.

Würthwein is no stranger to processing extremely large data sets. In 2013, he and his team used SDSC's *Gordon* supercomputer to provide auxiliary computing capacity to the OSG by processing massive data sets generated by the Compact Muon Solenoid (CMS), one of two particle detectors at the Large Hadron Collider near CERN, Switzerland. Nor is he a stranger to collaborations across the UC system. His group has been supporting data-intensive computing for CMS colleagues at UC Santa Barbara, UC Riverside, and UC San Diego for the last several years, with SDSC providing Wide Area Network connectivity. During the last year, SDSC helped Würthwein's lab establish 80 Gb/s (gigabits per second) network connectivity to LHCONE, the international wide area network in support of LHC science. This is an important first step toward realizing the LHC@UC goals.

Frank Würthwein is the executive director of the Open Science Grid, a national cyberinfrastructure to advance the sharing of resources, software, and knowledge; and a physics professor at UC San Diego.

"The goal of what we now call 'LHC@UC' is to provide tangible benefits to as many UC campuses as possible and across as wide a range of scientific domains," said Würthwein. "It makes sense to start with the LHC community because eight of the 10 UC campuses are already involved with it. Any LHC@UC member should be able to analyze his or her data transparently across the cloud, loosely networked high-throughput computing, and tightly coupled high-performance computing."

One of the key benefits of LHC@UC is that individual PIs (principal investigators) across all UC campuses will have direct access to SDSC's expertise and resources from their home institutions. "We view this network as a key solution and enabler for data-enabled research, and we can see the day when other university systems and research enterprises follow suit with similar systems," added Norman.

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

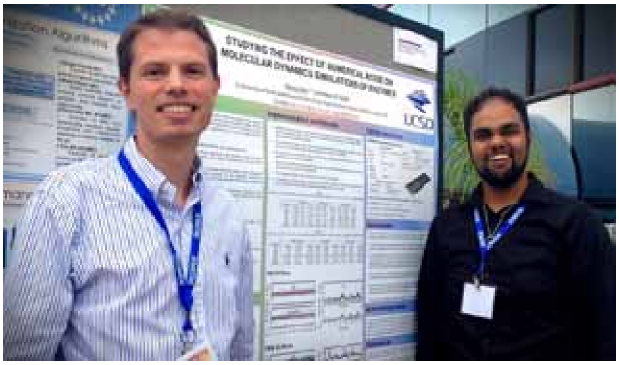
UC GRADUATE SUMMER FELLOWSHIP PROGRAM

Providing Opportunities in Computational Science for UC Graduates

In 2014, SDSC put the finishing touches on another program directly related to UC@SDSC—a UC Graduate Student Summer Fellowship program that provides opportunities for graduate students throughout the UC system to learn about SDSC's expertise and use the Center's resources to advance their own research.

"This eight-week program is designed to increase awareness of the value of computational science among the other UC campuses," said Diane Baxter, SDSC's associate director for education. "These graduate student fellows will also gain exposure to a more diverse range of career options, gain hands-on computational experience, and add computational research scientists as essential mentors who will help them succeed in their careers."

Fellows are expected to work with an established research team at SDSC to learn new skills that complement their own interests, or they may focus on their own research project while learning from an SDSC mentor whose area of expertise augments the fellow's advisory team.



SDSC Researcher Andreas Goetz (left) with Rahul Nori, of the University of North Dakota, at a project poster session at the National Science Foundation's eXtreme Science and Engineering Discovery Environment (XSEDE) 2013 annual conference.

To read the full press release, use a QR code reader or visit http://goo.gl/CzR5XX

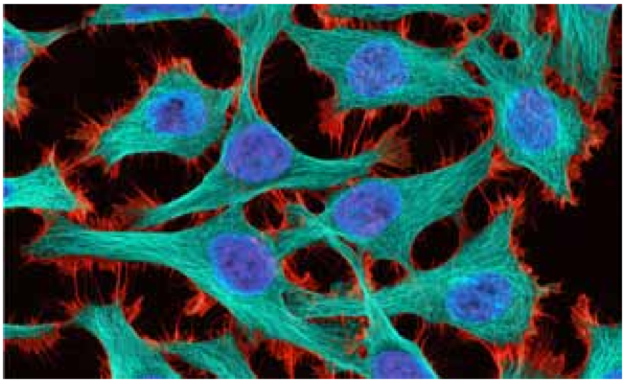
BREAKING THE BOTTLENECK IN INTERPRETING CANCER GENOMES

Researchers from the UC San Diego and UC San Francisco – with support from a diverse team of collaborators including SDSC – are undertaking an ambitious new project to determine how all of the components of a cancer cell interact.

"We're going to draw the complete wiring diagram of a cancer cell," said Nevan Krogan, director of the UC San Francisco division of QB3, a quantitative biosciences research institute, in announcing the Cancer Cell Map Initiative, or CCMI. Krogan is an investigator at Gladstone Institutes and co-director of CCMI with Trey Ideker, chief of medical genetics in the UC San Diego Department of Medicine and founder of the UC San Diego Center for Computational Biology & Bioinformatics.

The CCMI will provide key infrastructure for the recently announced alliance between UC San Diego Health Sciences and San Diego-based Human Longevity Inc., which plans to generate thousands of tumor genomes from UC San Diego cancer patients. It also will leverage resources and information from the National Cancer Institute (NCI), including large databases of cancer genomes and pathways that are being developed in collaboration with SDSC and UC Santa Cruz.

SDSC currently hosts the UC Santa Cruz Cancer Genomics Hub (CGHub), a secure repository for storing, cataloging, and accessing cancer genome sequences, alignments, and mutation information from the Cancer Genome Atlas (TCGA) consortium and related projects.



Cultured HeLa cancer cells. Image credit: Thomas Deerinck, National Center for Microscopy and Imaging Research, UC San Diego.

To read the full press release, use a QR code reader or visit http://goo.gl/I8VELE

NATIONAL IMPACT AND INFLUENCE

SDSC'S ADVANCED COMPUTING RESOURCES AND EXPERTISE AT THE NATIONAL LEVEL

As one of the first four supercomputer centers opened by the National Science Foundation (NSF) cooperative agreement, SDSC has a long history of "turning data into discovery" at the national level.

Collaborations have yielded hundreds of published papers and presentations at prestigious scientific meetings, leading to further scientific discovery across a broad range of fields, from biochemistry and cosmology to studying ways to sustain and improve Internet topology.

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

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

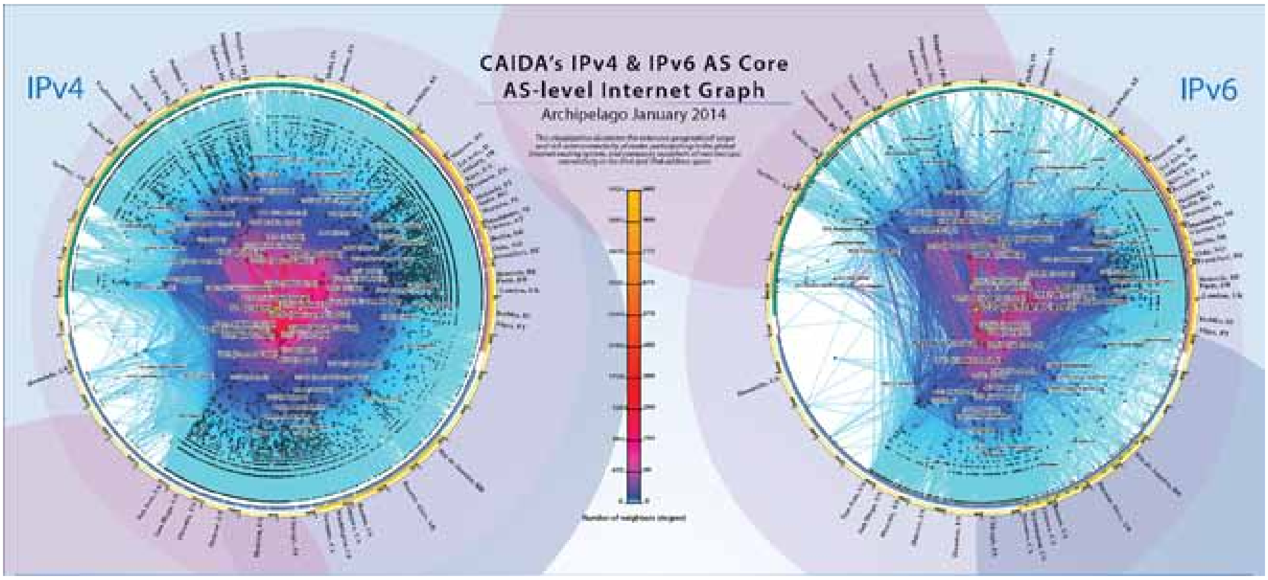
XSEDE, a partnership of 19 institutions, represents the most advanced collection of integrated advanced digital resources and services in the world. SDSC, the only supercomputer center participant on the West Coast, provides advanced user support and expertise for XSEDE researchers across a variety of HPC applications, in addition to support for the organization's central accounting database. SDSC also offers grid monitoring services through Inca, used by leading grid projects worldwide to identify, analyze, and troubleshoot user-level grid problems and failures. In 2014, SDSC was the first XSEDE partner to implement a new feature of the Globus software that allows researchers using the Center's computational and storage resources to easily and securely access and share large data sets with colleagues.

The Open Science Grid (OSG) Consortium and Project

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, and ideas. To further facilitate this sharing, the project operates services that allow for transparent computation across more than 150 clusters worldwide, including National Grid Initiatives in Europe, Asia, and the Americas. In 2014, SDSC Director Michael Norman became a member of the OSG Council, the governing body that defines policy and strategic direction. In early 2015, UC San Diego Physics Professor Frank Würthwein joined SDSC, and became the OSG Executive Director, thus leading the OSG project. SDSC is currently one of the lead institutions in the OSG.

Making Data-intensive HPC Resources Available

SDSC's new *Comet* cluster and its data-intensive *Gordon* system currently are accessible via the XSEDE allocation process to U.S. researchers as well as those affiliated with U.S.-based research institutions. *Comet* replaces SDSC's *Trestles* system, which came online in 2011. The result of an NSF award currently valued at $21.6 million including hardware and operating funds, *Comet* entered service in early April 2015 with the mission of expanding access and capacity among traditional as well as non-traditional research domains. (See page 7 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* is ideal for data mining and exploration, where researchers have to churn though tremendous amounts of data just to find a small amount of valuable information.

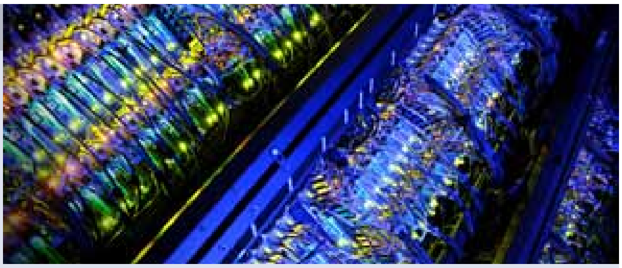


(Above) This visualization represents macroscopic snapshots of IPv4 and IPv6 Internet topology samples captured by CAIDA researchers in January 2014. For the IPv4 map, CAIDA collected data from 74 monitors located in 33 countries on six continents. For the IPv6 map, CAIDA collected data from 33 Ark monitors located in 21 countries on six continents.

Internet Research for Cybersecurity and Sustainability

The Center for Applied Internet Data Analysis (CAIDA), based at SDSC, is a collaboration among organizations in the commercial, government, and research sectors aimed at promoting greater cooperation in the engineering and maintenance of a robust, scalable global Internet infrastructure. In late 2014, CAIDA was recently awarded a $1.2 million NSF grant to measure and quantify the changing nature of the Internet's topology and what it means for the Internet's future in terms of design, operations, scientific study, and public policy. The goals of this three-year project are aligned with the NSF's Networking Technology and Systems (NeTS) program and include: advancing our fundamental understanding of how content distribution dynamics affect ISP network management capabilities; developing metrics to quantify the impact of emerging interconnection patterns on the resiliency, efficiency, and market power of modern networks; and revisiting longstanding but now questionable topology modeling assumptions and offering new models that are better empirically grounded.

For the 2008-2012 fiscal year period, the Center received more than $66.29 million in sub-awards from 47 non-UC San Diego research institutions, accounting for 69 total awards. Similarly, SDSC allotted nearly $7 million in sub-awards to 33 non-UC San Diego research partners during the same time period, underscoring the Center's research impact beyond UC San Diego.



Comet has Landed!

Comet, SDSC's new petascale supercomputer, began early operations in April 2015. *Comet* is designed to transform advanced scientific computing by expanding access and capacity among traditional as well as non-traditional research domains. The result of an NSF award currently valued at $21.6 million including hardware and operating funds, *Comet* is capable of an overall peak performance of two petaflops, or two quadrillion operations per second. "*Comet* is all about providing high-performance computing to a much larger research community—what we call 'HPC for the 99 percent'—and serving as an innovative gateway to discovery," said SDSC Director Michael Norman, the project's principal investigator. Specifically, *Comet* was designed to provide a solution for emerging research requirements often referred to as the 'long tail' of science, which describes the idea that the large number of modest-sized computationally-based research projects still represents, in aggregate, a tremendous amount of research and resulting scientific impact and advance.

See page 8 for more details about Comet.
To read the full press release, use a QR code
reader or visit http://goo.gl/fM9nHF

XSEDE SCIENCE GATEWAYS

Nancy Wilkins-Diehr, 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, is a leader in the construction of Science Gateways, which foster collaborations and the exchange of ideas among research communities.

A Science Gateway is a community-developed set of tools, applications, and data that is integrated through a web-based portal or a suite of applications. 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 both time and money.

A single gateway can give thousands of users access to current, optimized versions of analysis codes at any time. Codes with a large user base can be used by thousands through a single installation rather than hundreds of local installations. Researchers can focus on their scientific goals without having to know how supercomputers and other data cyberinfrastructures work.

Gateways also help foster collaborations and the exchange of ideas among researchers and have shown tremendous growth in terms of the number of users, the number of processing hours used on HPC resources by the broader user community, and in the number of published research papers enabled. They also can be readily used for teaching classes, workshops, and tutorials without having to set up codes on HPC resources, or create new accounts for students/attendees.

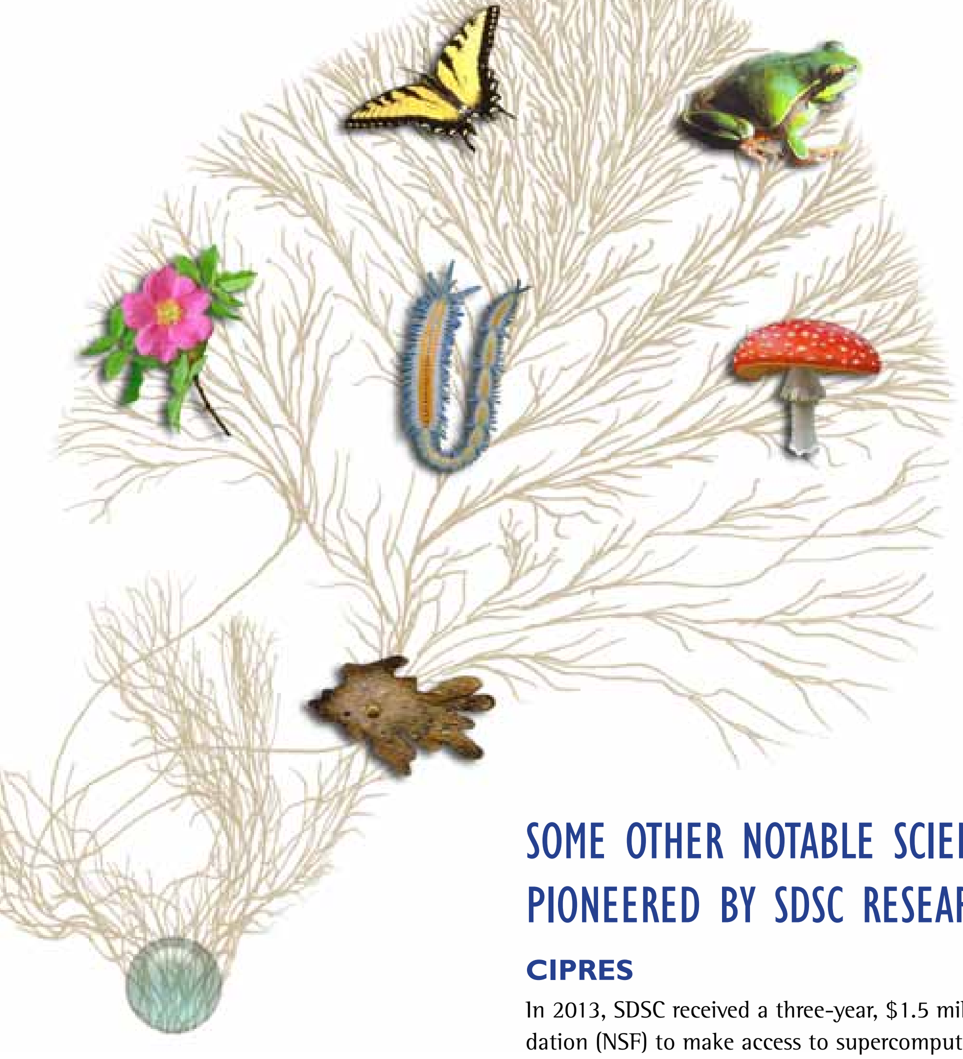
In early 2015, the National Science Foundation (NSF) named Science Gateways as one of two focus areas for the implementation phase of its Software Institute program. Award decisions (up to $3 million a year for five years for Science Gateways) may be made by the end of the year.

SDSC has been a leader in developing Science Gateways for a wide range of research communities. In 2012, SDSC was named the lead institution on an NSF planning grant for a conceptualization phase of a Science Gateway Institute that would offer a complete range of services aimed at connecting numerous individual groups developing domain-specific, user-friendly, Web-based portals and tools that enable scientific research. Partner institutions include Elizabeth City State University; Indiana University; Purdue University; the Texas Advanced Computing Center (TACC) at The University of Texas, Austin; and the University of Michigan, Ann Arbor.

Through a one-year, $500,000 conceptualization award, the team developed a strategic plan for a much larger Science Gateway Institute as part of the NSF's Software Institutes program. If funded, the Institute would provide a full support system for those developing gateways—from technical expertise to licensing advice to long-term planning and project management.

> *"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."*
>
> --Nancy Wilkins-Diehr, Associate Director, SDSC and Principal Investigator, Science Gateway Institute

The team envisions the Science Gateways Institute as offering a startup, or incubation service, which would include a complete development environment and hosting service, as well as consulting, documentation, and software recommendations to ensure the gateway is properly planned for maximum participation and success. An extended support team could build gateways for research teams that request support, transferring knowledge by teaching those teams what it takes to build, enhance, and operate gateways in the process. A 29,000-person survey of NSF principal investigators, university administrators and others showed overwhelming interest in such services.



Tree of Life image courtesy of Nick Kurzenko, Greg Rouse, and the U. S. Fish and Wildlife Service.

SOME OTHER NOTABLE SCIENCE GATEWAYS PIONEERED BY SDSC RESEARCHERS

CIPRES

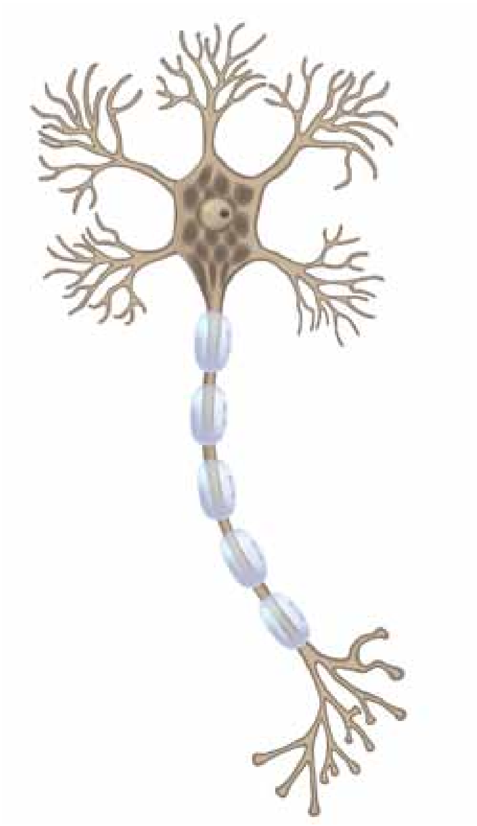
In 2013, SDSC received a three-year, $1.5 million award from the National Science Foundation (NSF) to make access to supercomputing resources simpler and more flexible for phylogenetics researchers. The award, which follows an earlier NSF grant that ran from 2003 to 2008, is for the CIPRES Science Gateway, a web site that allows researchers to explore evolutionary relationships between species using supercomputers provided by the NSF XSEDE project. CIPRES stands for CyberInfrastructure for Phylogenetic REsearch, and is among the most popular gateways in the XSEDE community. The CIPRES Gateway allows scientists to conduct their research in significantly shorter times without having to understand how to operate supercomputers, according to Mark Miller, principal investigator of the CIPRES gateway and an SDSC biologist. To date, the CIPRES Science Gateway has supported more than 12,000 users and has led to more than 1,300 publications of phylogenetic studies involving species in every branch of the Tree of Life.

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

For more information on SciGaP, use a QR code reader or visit www.scigap.org



Neuroscience Gateway

UC San Diego and Yale University are currently working under a collaborative NSF grant called "Advanced Biological Informatics Development: Building A Community Resource for Neuroscientists" to develop a Neuroscience Gateway (NSG) that gives neuroscientists broadened access to essential high-performance computing resources. The Neuroscience Gateway is a science gateway software infrastructure that makes neuroscience-specific computational tools conveniently available to researchers and students. The NSG offers high-performance compute time to neuroscience users through a streamlined process using a simple web portal-based environment for uploading neuronal models, running neuronal simulations on XSEDE's HPC resources, querying the status of jobs, and retrieving and storing output results. SDSC's Amit Majumdar, PI (principal investigator) of the NSF grant along with Yale PI Ted Carnevale, have seen tremendous adoption of NSG by the computational neuroscientists whose computing needs for simulation of large and complex brain models exceeds the resources available within their labs or institutions. NSG has been in production since early 2013 and within the first two years has provided over 3 million core hours to computational neuroscientists on SDSC's *Trestles* cluster. SDSC's training and outreach programs have supported the gateway's early success.

For more information on the Neuroscince Gateway, use a QR code reader or visit www.nsgportal.org

Amit Majumdar is division director of SDSC's Data-Enabled Scientific Computing division.

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

Going Globus

In 2014, SDSC implemented a new feature of the Globus software that will allow researchers to easily and securely access and share large data sets with colleagues. Described as a "dropbox for science", Globus is already widely used by resource providers and users who need a secure and reliable way to transfer files. SDSC was the first supercomputer center within the NSF's XSEDE program to offer Globus sharing. While SDSC has been offering file transfer capability via Globus to users for several years, the Center has been providing several Globus Plus accounts to selected users free of charge so that they can allow their collaborators, including those who don't have an account on SDSC clusters, to access (read and write to their shared file space) data on SDSC resources. "Integrating the Globus sharing capability into SDSC's widely used data-intensive computing and storage systems lets researchers and resource providers hand off the challenges of data sharing and movement to a hosted service that manages the entire process, while also monitoring performance and providing status reports," said Amit Majumdar, head of SDSC's Data-Enabled Scientific Computing division.

HEALTHCARE INFORMATION TECHNOLOGIES

Sandeep Chandra is division director, Health Cyberinfrastructure, SDSC.

Sherlock is SDSC's Center of Excellence focused on managed information technology and data services in healthcare for academia and government. Services, offered to federal, state, and local governments as well as the University of California system and universities nationwide, include compliant cloud hosting, cyber security, data management, application development, and visualization. "Data management, technology, and policy challenges, especially in the health sector, can be overwhelmingly complex and confusing," said Sandeep Chandra, Sherlock's director. "We have developed and deployed specific services designed to provide a solid and secure foundation for a wide range of initiatives, including how Sherlock is taking on healthcare fraud."

See page 52 to learn more about Sherlock.

To read the full press release, use a QR code reader or visit http://sherlock.sdsc.edu

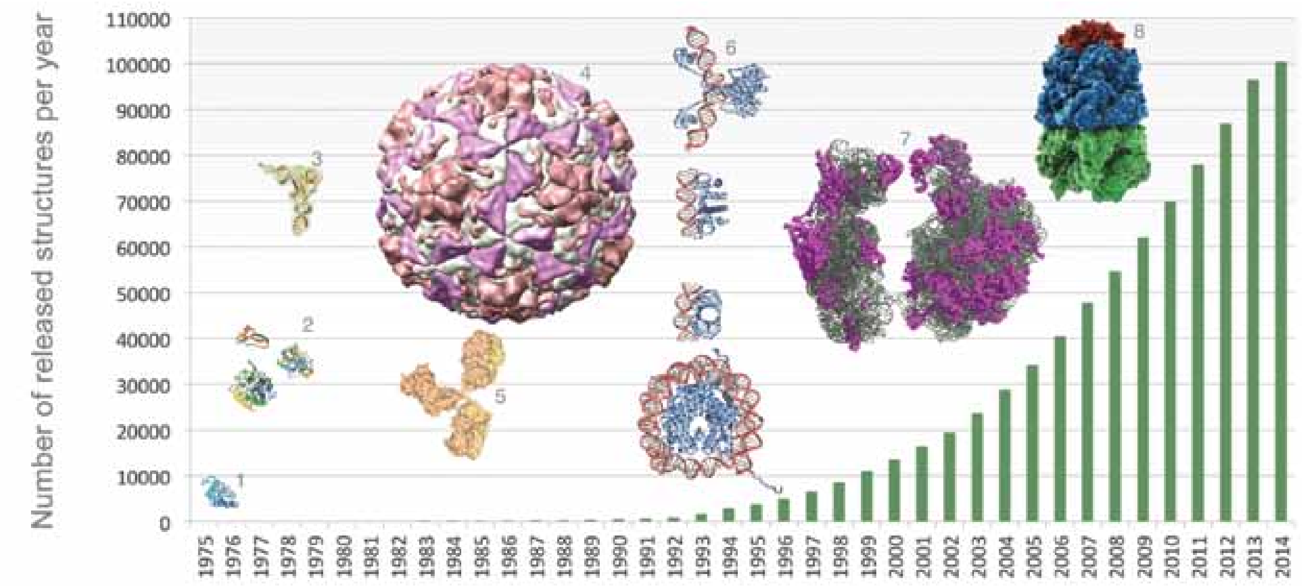


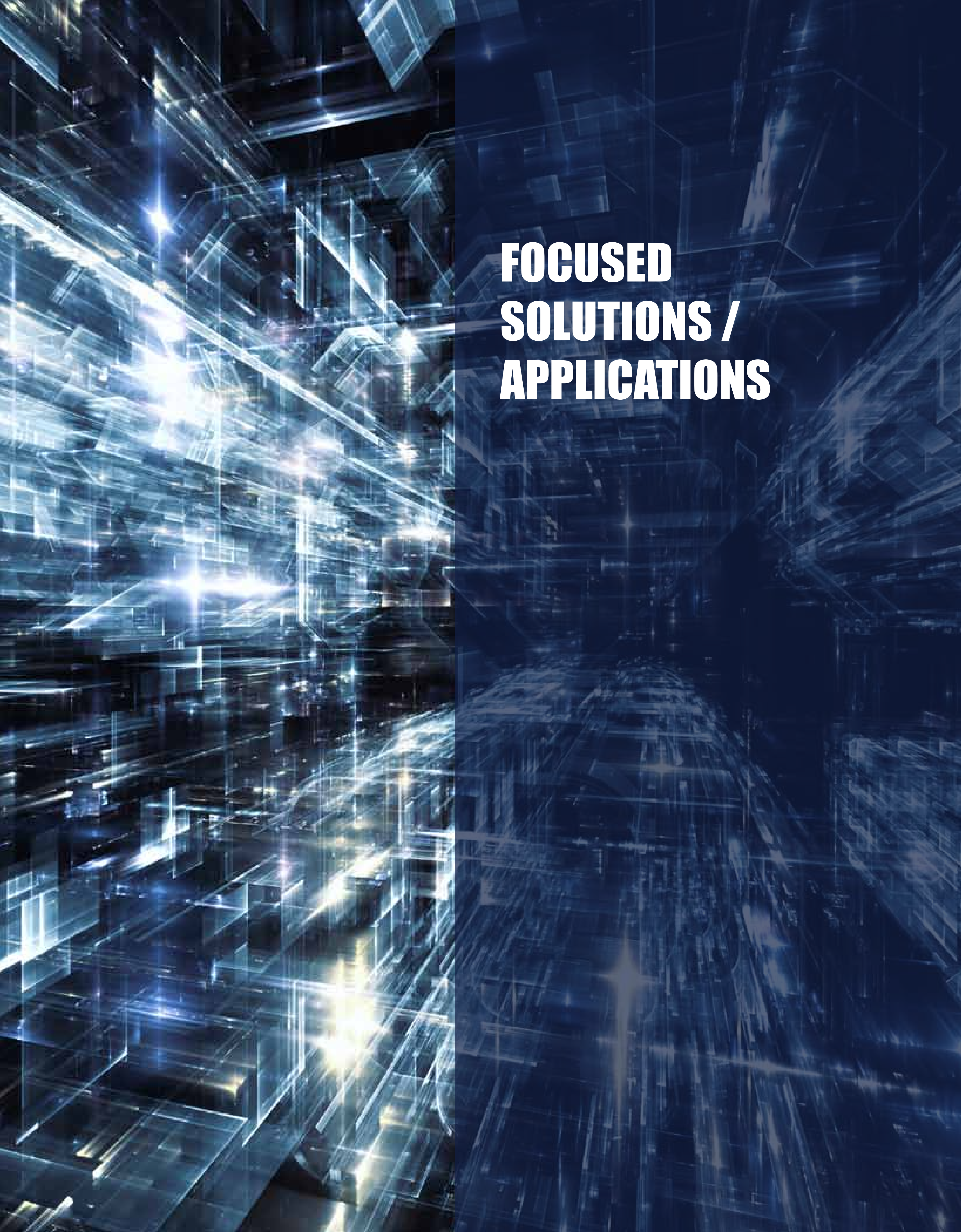
Image credit: Protein Data Bank

Protein Data Bank Structures Surpass 100,000

In 2014, the Protein Data Bank (PDB), the single worldwide repository for the three-dimensional structures of large molecules and nucleic acids, archived its 100,000th structure, doubling its size in just six years. Co-located at Rutgers, The State University of New Jersey; and SDSC in conjunction with UC San Diego's Skaggs School of Pharmacy and Pharmaceutical Sciences, 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. "SDSC has provided a safe haven for the 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." Also in 2014, PDB introduced a free mobile application device that enables users from the general public and expert researchers to quickly search and visualize the 3D shapes of proteins, nucleic acids, and molecular machines.

To read the full press release, use a QR code reader or visit http://goo.gl/6JHilF

FOCUSED SOLUTIONS / APPLICATIONS



FOSTERING PARTNERSHIPS across Industry, Academia, and Government

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.

INDUSTRY PARTNERS PROGRAM

SDSC's engagements with industrial partners in 2014 continued to be driven by three key themes: Big Data, Predictive Analytics, and Advanced Computing Technology.

For Big Data, computing and storage support for human genomics-related research and commerce was a primary motivator. The rapidly decreasing cost and time associated with sequencing human genetic material is resulting in massive amounts of data that must be stored and analyzed. Biogenetic and pharmaceutical organizations recognize that determining the genetic factors in sickness and disease for retargeting approved drugs for treatment, and for conducting "personalized" and "precision" medicine, represent the next wave in health care.

SDSC, located in the heart of San Diego's biotechnology cluster, has formed relationships and collaborations with local biotech companies and research institutes such as Janssen R&D and the J. Craig Venter Institute. The Center has provided storage and computational analyses of large human genome datasets on its *Gordon* and *TSCC* supercomputers, while its researchers and technical staff have provided assistance in bioinformatics programming, optimizing resource configurations, and setting up computational "pipelines" for genome analysis.

Predictive Analytics and Data Mining has been a strong theme at SDSC for several years, and this continued during 2014. Many enterprises are recognizing the value of analyzing and mining their troves of data for competitive advantage. In turn, this has triggered a strong impact on workforce perspectives as companies look to hire and handsomely compensate "data scientists" to guide their predictive analytics initiatives. As a result, SDSC saw strong turnouts in 2014 for its 'Data Mining Boot Camps' program. This Boot Camp is built around an intensive, two-day format designed to introduce concepts and impart practical skills to working professionals. Data mining boot camps and additional training topics in this format are continuing in 2015. In addition to training, companies such as the professional services firm The Marlin Alliance, Inc., have established partnerships with SDSC to explore use cases in predictive analytics and data science to benefit both their internal staff and clients.

For Advanced Computing Technology, SDSC's computational scientists and technical staff continued to be recognized by industry as an expert resource for exploring and evaluating high-performance computing (HPC) architectures and emerging computing technologies. During 2014, SDSC continued to perform work for Intel and other companies in the performance modeling and characterization of HPC architectures, including new processors and memory subsystems. For a company in the energy sector, Center researchers conducted evaluations of new architectures for computing and visualization.

SDSC experts also continued their partnership with visual computing technology company NVIDIA to develop and optimize scientific codes for NVIDIA's advanced general purpose graphics computing units in the areas of earthquake simulation and computational chemistry. As testimony to its expertise in one critical area, the Center was designated as an Intel Parallel Computing Center focusing on molecular dynamics code optimization and training for Intel's Xeon Phi® many-core computing architecture.

TSCC COMPUTING “CONDO” CONTINUES TO GROW

While SDSC’s national supercomputing systems are available to bona fide UC San Diego researchers, access to the national systems is highly competitive and user queues can be lengthy. As a result, in 2013 SDSC launched the *Triton Shared Computing Cluster (TSCC)* to provide UC San Diego investigators with a high-performance computing (HPC) system dedicated to their needs with quick access and reasonably short wait times. Following an extensive study of successful research computing programs across the country, SDSC selected “condo computing” as the main business model for *TSCC*. As its name implies, condo computing is a form of shared ownership in which researchers use equipment purchase funds from grants or other sources to buy and contribute compute “nodes” (servers) to the system. The result is a researcher-owned computing resource of medium to large proportions.

In 2014, some 14 UC San Diego labs/groups and 230 users participated in the program, for a total of 170 nodes (approximately 3,000 processors) and more than 80 teraflops of computing power. Participating researchers/labs are working in a variety of disciplines including engineering, computational chemistry, genomics, oceanography, high-energy physics, and others.

Also during 2014, TSCC became part of the Integrated Digital Infrastructure (IDI) program, the UC San Diego Chancellor’s initiative to advance and streamline the delivery of cutting-edge IT services to campus faculty, researchers, and students in the lab.

Working with all the IT providers on campus, UC San Diego’s IDI directs researchers to high-end and big data-friendly services to support research and instruction, including high-speed network connections, high-performance computing, colocation facilities, storage, tools, and training.

Features and Benefits

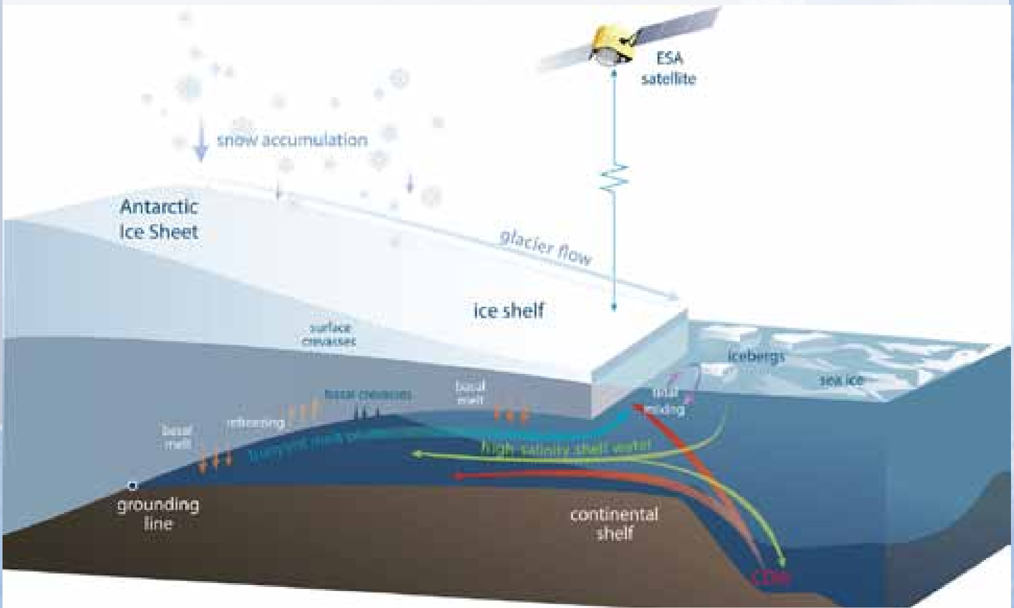
The condo computing model implemented for *TSCC* provides many features:

- Researchers use equipment purchase funds to buy “nodes” (compute servers) for the cluster. An additional “infrastructure fee” covers the purchase of shared components (racks, network switches, cables, etc.).
- Participants may then have dedicated use of their purchased nodes, or they may run larger computing jobs by sharing other researchers’ idle nodes. As a result, the researcher effectively has access to a much larger cluster than would typically be available to his/her lab.
- Researchers also pay an annual “operations fee” for each of their purchased nodes. This fee covers labor for maintaining the system, utilities, software licenses, etc. Currently, the IDI program substantially subsidizes this fee for UC San Diego researchers.
- Researchers may run jobs in a “glean queue” which does not count against their annual allocation of computing time.
- Participation in the program runs nominally for three years, which is the duration of the equipment warranty. Researchers may leave their nodes in the system for a fourth year, though equipment failing during this period may not be repaired.
- Researchers may remove and take possession of their nodes at any time; once equipment is removed from the cluster it may not be returned.

Benefits of the program include:

- Access to a much larger cluster than most labs could typically afford, at a fraction of the total cost.
- Professionally administered/managed postdocs and graduate students can focus on research instead of maintaining a computing cluster.
- Equipment is housed in a secure, climate-controlled, energy-efficient data center.
- Access to high-performance hardware with latest generation Intel server processors and a high-bandwidth, low latency network for maximum parallel computing performance.
- Many installed software packages, or researchers may install their own.
- Researchers have access to a community of participants and users that can share tips and information.

To support researchers who do not have funds to purchase equipment or have short-term or sporadic computing needs, *TSCC* also operates a “hotel” section where users may “recharge” (purchase) arbitrary blocks of computing time and run as a guest user on the system.



Schematic diagram of an Antarctic ice shelf showing the processes causing the volume changes measured by satellites. Ice is added to the ice shelf by glaciers flowing off the continent and by snowfall that compresses to form ice. Ice is lost when icebergs break off the ice front, and by melting in some regions as warm water flows into the ocean cavity under the ice shelf. Under some ice shelves, cold and fresh meltwater rises to a point where it refreezes onto the ice shelf. Image credit: Helen Amanda Fricker, SIO, UC San Diego.

SDSC SUPERCOMPUTERS ASSIST IN SCRIPPS' STUDY OF THINNING ANTARCTIC ICE

A recently published study led by researchers at Scripps Institution of Oceanography at UC San Diego (SIO) that shows a significant decline in the thickness of Antarctica's floating ice shelves was made possible with the computational prowess of SDSC's *TSCC (Triton Shared Computing Cluster)*.

The widely reported study, published in March 2015 in the journal *Science*, also used *TSCC's* predecessor, the *Triton Compute Cluster*, to analyze 18 years of satellite data.

"We used parallel processing to handle a relatively large amount of observations," said Scripps graduate student Fernando Paolo, lead author of the study. "Those 18 years of raw satellite data consist of 20 observations per second continuously, so we could perform within a few hours processes what would take weeks or even months on a typical desktop computer."

The study was conducted over a period of three years, using SDSC's shared compute resources. "We needed high-performance computing to speed up locating and processing crossing points between satellite orbits, from which we estimated changes in the ice-shelf thickness," said Paolo.

Before further processing the data, the researchers had to first calculate the points where samples from two satellite orbits crossed each other. This is a very time-consuming task because there are thousands of orbits and millions of potential crossing points, explained Paolo.

Paolo assisted Scripps glaciologist Helen Amanda Fricker and oceanographer Laurie Padman of Earth and Space Research, a non-profit institute in Corvallis, Oregon, in constructing a new high-resolution record of ice shelf thickness based on satellite radar altimetry missions of the European Space Agency from 1994 to 2012. The study reports that the ice shelves decreased in volume by as much as 18 percent in certain areas over almost two decades, providing new insights on how the Antarctic ice sheet is responding to climate change.

"Eighteen percent over the course of 18 years is a substantial change," said Paolo in an SIO press release. "Not only is the total ice shelf volume decreasing, but we see an acceleration in the last decade."

For a Scripps video showing the ice melt due to rising temperatures, use a QR code reader or visit https://goo.gl/K8SNio

SDSC CENTERS OF EXCELLENCE

In recent years, SDSC has created several ‘Centers of Excellence’ as part of a larger strategic focus to help researchers across all domains – including those who are relatively new to computational science – better manage ever-increasing volumes of digitally-based information. These centers formally represent key elements of SDSC’s wide range of expertise, from ‘big data’ management to the analysis and advancement of the Internet.

Ilkay Altintas is director of the WorDS Center and principal investigator for the WIFIRE project, a university-wide collaboration funded by the National Science Foundation to create a cyberinfrastructure to effectively monitor, predict, and mitigate wildfires.

Workflows for Data Science (WorDS) Center

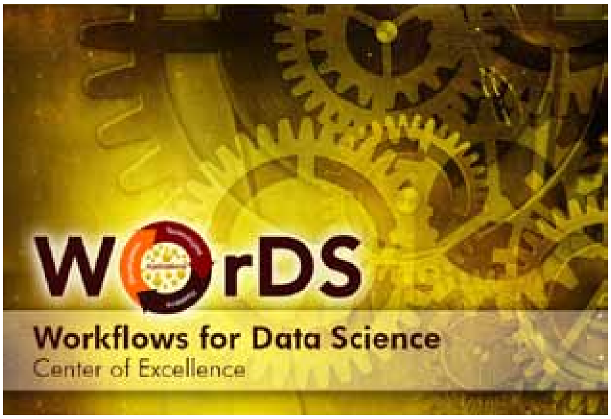
Called the WorDS Center for ‘Workflows for Data Science’, this new center, formed in 2014, leverages more than a decade of experience within SDSC’s Scientific Workflow Automation Technologies Laboratory that develops and validates scientific workflows for researchers involved in computational science, data science, and engineering.

“WorDS is designed to serve those researchers at the intersection of distributed and parallel computing, big data analysis, and reproducible science, while fostering a collaborative working culture,” said Ilkay Altintas, director of WorDS. “Our aim is to assist researchers in creating workflows to better manage the tremendous amount of data being generated across a wide range of scientific disciplines, from natural sciences to marketing research, while letting them focus on their specific areas of research instead of having to solve workflow issues or the computational challenges that arise as data analysis progresses from task to task.”

Expertise and services offered by the WorDS Center 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, Cascading;
- Technology briefings and classes on end-to-end support for data science.

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



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

Sherlock

Sherlock is SDSC's center of excellence focused on managed information technology and data services in healthcare 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 its 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-complaint 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. "We've developed and deployed specific services designed to provide a solid, secure foundation for a wide range of initiatives, including taking on healthcare fraud."

Sherlock's resources are physically located within the SDSC Data Center, and as needed for redundancy, in a secure data center in Northern California. Sherlock Cloud systems interconnect with a 10Gb/s (gigabits per second) network fabric within the SDSC Data Center, and wide-area networking utilizes more than 100Gb/s of high-bandwidth connections to the Internet and research networks such as Internet2, National Lambda Rail (NLR), and the Corporation for Education Network Initiatives in California (CENIC).

(left) For Sherlock, use QR code reader or visit http://sherlock.sdsc.edu

(right) For CLDS, use QR code reader or visit http://clds.sdsc.edu

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

Center for Large-scale Data Systems Research (CLDS)

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 big data era. CLDS specializes in developing applicable concepts, frameworks, analytical approaches, case analyses and systems solutions to big data management, with a related goal of developing 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.

Predictive Analytics Center of Excellence (PACE)

PACE, also announced in 2012, was started to foster collaboration and education among industry, government, and academia to provide a multi-level curriculum that gives business and science enterprises the critical skills to design, build, verify, and test predictive data models. Natasha Balac leads and manages the PACE initiative at SDSC. Balac has been with SDSC since 2003, leading multiple large projects and collaborations across a wide range of organizations in industry, government, and academia including the Centers for Medicare and Medicaid Services (CMS), National Science Foundation (NSF), National Institutes of Health (NIH), and the California Energy Commission (CEC). Her research includes several large-scale predictive analytics projects including collaborations with the UC San Diego School of Medicine and the campus-wide Smart Energy Grid, as well as public and private sector partnerships such as Sustainable San Diego.

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

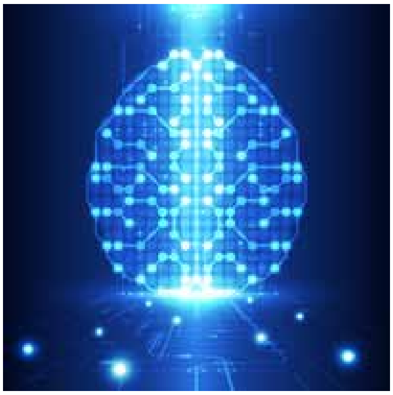
Center for Applied Internet Data Analysis (CAIDA)

CAIDA, formed in 1997, is a collaborative undertaking among organizations in the commercial, government, and research sectors aimed at promoting greater cooperation in the engineering and maintenance of a robust, scalable global Internet infrastructure. In 2014, Kimberly Claffy, CAIDA's principal investigator and co-founder, was awarded the Institute of Electrical and Electronics Engineers' Internet Award. In 2014, CAIDA was awarded a three-year, $1.2 million NSF grant to measure and quantify the changing nature of the Internet's topology and what it means for the World Wide Web's future in terms of design, operations, scientific study, and public policy. David Clark, a senior research scientist at the MIT Computer Science and Artificial Intelligence Laboratory (MIT/CSAIL), the largest research laboratory at MIT and one of the world's leading centers of information technology research, is collaborating on this project.

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

PROVIDING SOLUTIONS for Data Scientists

SDSC's wide range of expertise in advanced computation has resulted in more 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 new solutions to both age-old disasters and emerging threats around the world. Below are a few projects made possible by SDSC's staff and resources in 2014:



SUPPORTING THE NATIONAL BRAIN INITIATIVE

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.

UC San Diego's Center for Brain Activity Mapping (CBAM) also was established in 2013 as the nation's first academic center dedicated to the brain mapping effort. CBAM aims to develop a new generation of tools for recording neuronal activity, conducting brain-mapping experiments, and analyzing collected data. Computational modeling of cells and networks is an essential part of neuroscience research, and investigators are using such simulations to address problems of ever-increasing complexity.

To facilitate access to high-performance computing resources, the National Science Foundation (NSF) is funding a collaborative effort between SDSC and the Yale School of Medicine to develop a Neuroscience Gateway (NSG). The web-based portal, www.nsgportal.org, will provide users with computer time for research and instruction, a streamlined process for uploading models, and a community forum to collaborate and share data.

SDSC/UC SAN DIEGO RECEIVES MULTIPLE 2014 HPCWIRE AWARDS FOR WIFIRE PROJECT

Calit2 Director Larry Smarr (left) and Ilkay Altintas (center), Director of SDSC's Workflows for Data Science Center of Excellence (WorDS) and PI of WIFIRE accept three awards for the WIFIRE project from Tom Tabor, CEO of Tabor Communications and publisher of HPCwire.

SDSC achieved a proverbial hat trick by garnering three awards for its university-wide WIFIRE project as part of the annual HPCwire Readers' and Editors' Choice Awards, presented in November at the 2014 International Conference for High Performance Computing, Networking, Storage and Analysis (SC14) in New Orleans.

Ilkay Altintas, SDSC's deputy coordinator for research and PI (principal investigator) of the project, was on hand with Calit2 Director Larry Smarr to accept the three awards for the WIFIRE, a multi-year, $2.65 million NSF-funded project to create a cyberinfrastructure to more effectively monitor, predict, and mitigate wildfires.

SDSC, along with its partners in the WIFIRE project, was recognized with the following honors:

- Readers' Choice: Best Application of Big Data in HPC;
- Editors' Choice: Best Application of Big Data in HPC;
- Editors' Choice: Best Data-Intensive System, End-user Focused.

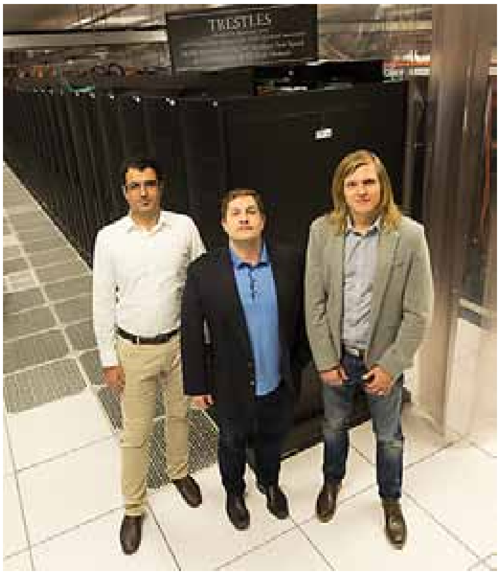
Aside from SDSC and Calit2, participants in the WIFIRE project include researchers from UC San Diego's Mechanical and Aerospace Engineering (MAE) department in the Jacobs School of Engineering and the University of Maryland's Department of Fire Protection Engineering.

To read the full press release, use a QR code reader or visit http://goo.gl/5inyLT

EXAMINING EXTREME EVENTS THROUGH SUPERCOMPUTER SIMULATIONS

A new center at the UC San Diego Jacobs School of Engineering focused on developing new ways to protect buildings and infrastructure as well as the human body from extreme events such as blasts from terrorist attacks or high-impact collisions, is joining forces with SDSC to provide data-intensive supercomputer simulations. The Center for Extreme Events Research (CEER) brings together a unique combination of experts in experimental and computational research. Yuri Bazilevs, an expert in finite element analysis and isogeometric analysis and associate director of CEER, is heading the center's initiatives in the field of computational simulation. His research group focuses on the development of computational methods for large-scale coupled fluid-structure interaction problems and their implementation in high-performance computing environments.

"Systems such as SDSC's *Trestles* supercomputer enable simulation of full-scale engineering systems in a much more efficient manner," said Bazilevs, who used *Trestles* previously to develop code and run cardiovascular applications. "High-performance computing plays a key role in all the methods and software development, which is essential to CEER's projects. Having access to thousands of compute cores means we can perform simulations with more sophisticated mechanics modeling and higher temporal and spatial resolution, resulting in greater physical realism and numerical accuracy."



Yuri Bazilevs (center) is associate director of CEER and an expert in finite element analysis and isogeometric analysis. Also pictured is postdoctoral researcher Kazem Kamran (left) and Ph.D. student Artem Korobenko.

To read the full press release, use a QR code reader or visit http://goo.gl/OPBk01

SDSC AND LEIDOS DEVELOPING CYBERSECURITY PROTOCOLS FOR ELECTRICAL MICROGRIDS

SDSC and Leidos (formerly SAIC) announced plans in early 2014 to jointly develop protocols aimed at increasing security levels of systems used to manage electrical microgrids worldwide. Microgrids are small-scale versions of traditional larger power grids that draw energy from clean sources such as the wind and sun, as well as from conventional technology. Microgrids can more efficiently manage real-time demand, supply, and storage of energy at a lower cost and with less pollution than a conventional grid, and their use has grown significantly during the past decade.

The SDSC project will focus on analyzing the cybersecurity aspect of one of the world's most advanced microgrids, located on the UC San Diego campus. The university saves more than $8 million a year in power costs due to its microgrid operation. The campus' microgrid project has also spurred investment: the nearly $4 million that the Energy Commission has invested in the microgrid since 2008 has been leveraged to garner more than $4 million from other funding sources, both public and private.

Solar panels such as the ones atop the Hopkins Parking Structure are part of the UC San Diego microgrid, which generates 92 percent of the electricity used on campus. Image: UC San Diego

To read the full press release, use a QR code reader or visit http://goo.gl/ZNR2WZ

"Improving security for our sophisticated microgrid is extremely important to us," said Byron Washom, director of Strategic Energy Initiatives at UC San Diego. "This project will allow us to establish effective baseline security controls that could be applied to microgrids all over the world."

"SDSC will contribute its security experience in dealing with complex supercomputer systems as well as securing sensitive data such as FISMA- and HIPAA-compliant databases, to help raise awareness of security issues facing microgrids," added Winston Armstrong, SDSC's chief information security officer. "SDSC will also recommend improvements to harden microgrid control systems in general."

Hans-Werner Braun is an SDSC research scientist who helped manage the Area Situational Awareness for Public Safety Network (ASAPnet). Braun, along with UC San Diego Scripps Institution of Oceanography Seismologist Frank Vernon, co-founded HPWREN in 2000.

HPWREN/ASAPNET: A WIRELESS EDUCATION AND SAFETY NETWORK FOR SCIENCE AND SOCIETY

HPWREN, the High-Performance Wireless Research and Education Network, provides high-speed Internet access to field researchers from disciplines including geophysics, astronomy, and ecology; educational opportunities through connections to learning centers in several communities; and advanced warning and monitoring systems to firefighters in distant sections of San Diego County.

In addition, HPWREN has been working with industry and government officials to expand its Area Situational Awareness for Public Safety Network (ASAPnet), a grid of high-speed, wireless communications connecting dozens of backcountry fire stations in the region.

Started in 2000 under a National Science Foundation (NSF) grant, HPWREN today consists of a collaboration of researchers at SDSC, the Scripps Institution of Oceanography's Institute (SIO) of Geophysics and Planetary Physics, UC San Diego's California Institute for Telecommunications and Information Technology (Calit2) Qualcomm Institute, Caltech's Palomar Observatory, San Diego Gas & Electric (SDG&E), and various San Diego firefighting agencies.

During 2014, HPWREN/ASAPnet underwent a significant build-out of environment-sensing cameras, connecting more than 60 firefighter locations including fire stations, air bases, and camps under an expansion supported by SDG&E. In addition, Calit2 provided an 80-terabyte storage array and staff support to provide for long-term storage of camera images.

HPWREN/ASAPnet is a participant in the WIFIRE project that has been cataloging and integrating data related to dynamic wildfire models from a variety of resources including sensors, satellites, and scientific models, and creating visual programming interfaces for using that data in scalable wildfire models. WIFIRE encompasses HPWREN's remote sensor network.

"San Diego County is already well positioned to monitor and analyze these dynamics through sensors within and outside of our research networks," said HPWREN co-founder Hans-Werner Braun, who is also a co-PI of WIFIRE. "We have been collecting environmental data for more than 10 years through HPWREN, merging large volumes of data and computational models into sophisticated visualizations, and have forged new networks through government and industry partners including CAL FIRE, the U.S. Forest Service, SDG&E, and the San Diego County Emergency Operations Center to direct and share our research."

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

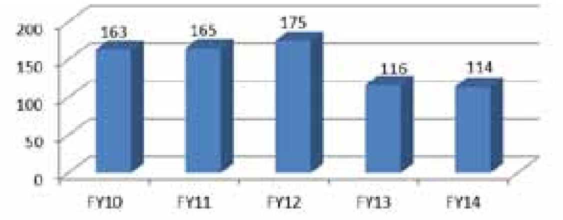
FACTS & FIGURES

Proposal Success Rate

	FY10	FY11	FY12	FY13	FY14
Proposals Submitted	119	92	116	84	76
Proposals Funded	44	44	51	40	30
Success Rate	37%	48%	44%	48%	39%

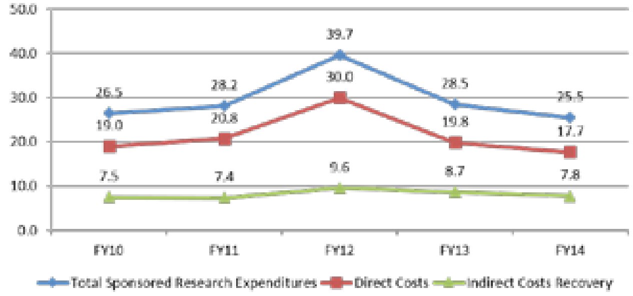
In perhaps the most competitive landscape for federal funding in the last two decades, SDSC's overall success rate on federal proposals remains at about 43%, compared to a national average of roughly 15% for computer science and engineering proposals at the National Science Foundation.

Number of Sponsored Research Awards



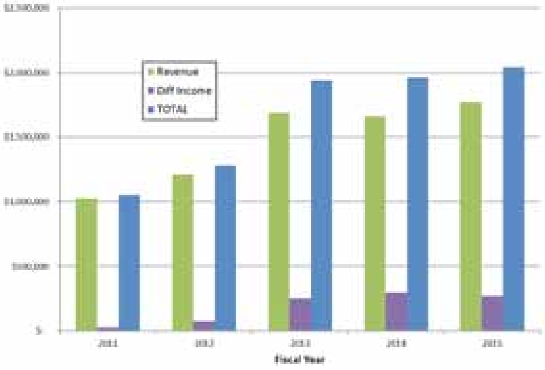
FY13-FY14 decline represents a shift to fewer awards with larger budgets.

Sponsored Project Expenditures ($M)

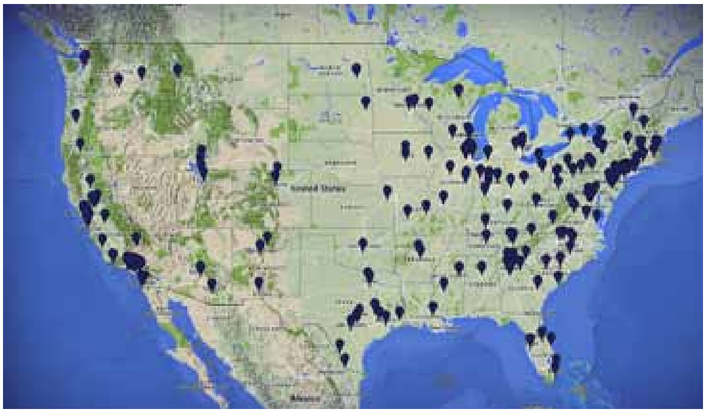


Apart from the extraordinary research impact of SDSC collaborations and partnerships, a quick look at the fiscal impact of these collaborations is impressive. During its 28-year history, SDSC revenues have exceeded $1 billion, a level of sustained funding matched by few academic research units in the country. In the above graph, FY12 expenses are higher due to a $10M hardware purchase for NSF's *Gordon* Supercomputer.

Industry Revenue Data 2010-2015



Geographical Distribution of National Users of SDSC HPC Resources

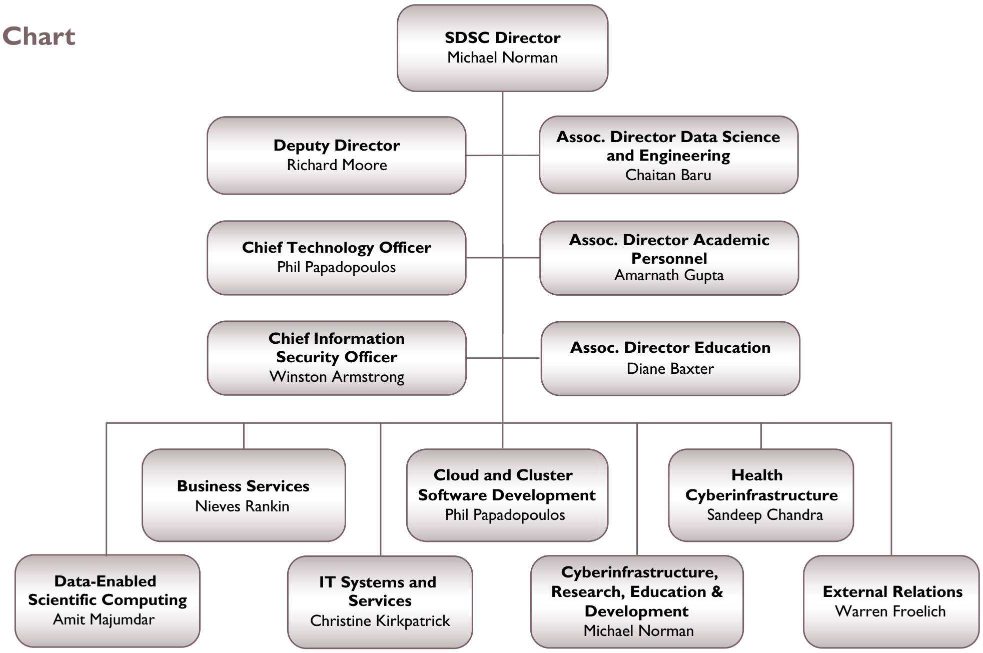


A total of 1,265 unique users from around the world accessed SDSC's HPC resources (*Gordon* and *Trestles*) during FY2014. Of these users, 1,245 were based in the United States. The adjacent map displays a geographic disbursements of users from different cities across the U.S.

On *Gordon*, a total of 91,754,679 service units (SUs) were used, 94% of which were charged against XSEDE accounts. On *Trestles*, a total of 58,476,499 SUs were used, 97% of which were charged against XSEDE accounts.

ORGANIZATION & LEADERSHIP

SDSC Org Chart



SDSC Director
Michael Norman

Deputy Director
Richard Moore

Assoc. Director Data Science and Engineering
Chaitan Baru

Chief Technology Officer
Phil Papadopoulos

Assoc. Director Academic Personnel
Amarnath Gupta

Chief Information Security Officer
Winston Armstrong

Assoc. Director Education
Diane Baxter

Business Services
Nieves Rankin

Cloud and Cluster Software Development
Phil Papadopoulos

Health Cyberinfrastructure
Sandeep Chandra

Data-Enabled Scientific Computing
Amit Majumdar

IT Systems and Services
Christine Kirkpatrick

Cyberinfrastructure, Research, Education & Development
Michael Norman

External Relations
Warren Froelich

Executive Team

Chaitanya Baru
Assoc. Director Data Science and Engineering

Warren Froelich
Division Director External Relations

Ronald Bruce Hawkins
Director Industry Relations

Christine Kirkpatrick
Division Director IT Systems and Services

Richard Moore
Deputy Director

Michael Norman
SDSC Director

Philip M. Papadopoulos
Division Director Cloud & Cluster Software Development

Nieves Rankin
Division Director Business Services

Frank Würthwein
Lead, Distributed High-Throughput Computing

Executive Committee

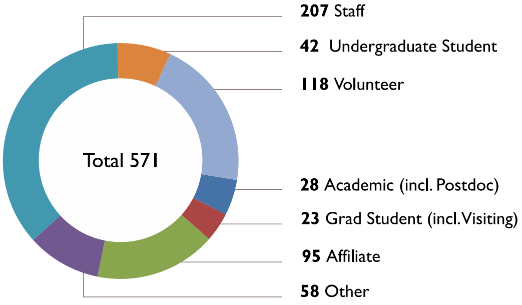
UC SAN DIEGO

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

Chaitanya Baru
Sandeep Chandra (interim)
Warren Froelich
Christine Kirkpatrick
Richard Moore
Michael Norman
Philip M Papadopoulos
Nieves Rankin

SDSC Census FY2014



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

Yifeng Cui, Ph.D.
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

Bob Sinkovits, Ph.D.
Scientific Applications Lead, SDSC
Gordon applications
Data-intensive high-performance computing
Computational physics and fluid dynamics
Bioinformatics
Relationship databases
Compute clusters systems administration

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.
Co-Principal Director, XSEDE at SDSC
Co-Director for Extended Collaborative Support, XSEDE
Science gateways
User services
Aerospace engineering

Frank Würthwein, M.S.
Distributed High-Throughput Computing Lead, SDSC
Professor of Physics, UCSD
High-capacity Data Cyberinfrastructure
High-energy Particle Physics

SDSC Data Scientists

Ilkay Altintas, Ph.D
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

Natasha Balac, Ph.D.
Director, Predictive Analytics Center of Excellence, SDSC
Director of Data Application and Services, SDSC
Lecturer, Computer Science & Engineering, UCSD
Data mining and analysis
Machine learning
Predictive analytics
Data-intensive computing
Big Data analytics

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, SDSC
Adjunct Professor, College of Sciences, SDSU
Director/PI, High Performance Wireless Research and Education Network (HPWREN)
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.
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

Jianwu Wang, Ph.D.
Assistant Project Scientist, SDSC
Lecturer, Computer Science & Engineering, UCSD
Scientific workflow automation
Data-intensive computing

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

SDSC

San Diego Supercomputer Center
University of California, San Diego
9500 Gilman Drive MC 0505
La Jolla, CA 92093-0505

www.sdsc.edu
twitter/SDSC_UCSD
facebook/SanDiegoSupercomputerCenter