SDSC ANNUAL REPORT FY2018/19

COMPUTING WITHOUT BOUNDARIES

SDSC

SAN DIEGO SUPERCOMPUTER CENTER

UNIVERSITY OF CALIFORNIA SAN DIEGO

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 petascale *Comet* supercomputer continues to be a key resource within the National Science Foundation's (NSF) eXtreme Science and Engineering Discovery Environment (XSEDE) program.

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
Division Director, External Relations
jzverina@sdsc.edu
858-534-5111

COMPUTING WITHOUT BOUNDARIES
SDSC Annual Report FY2018/19
(PDF version available online at the SDSC website)

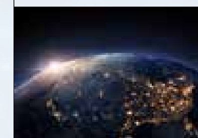
EDITOR:
Jan Zverina

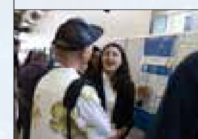
CO-EDITOR:
Kim Bruch

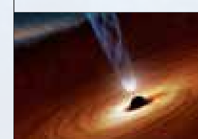
CONTRIBUTORS:
Kim Bruch, Julie Gallardo, Ron Hawkins, Fritz Leader, Susan Rathbun, Bob Sinkovits, Shawn Strande, Ben Tolo, Jan Zverina

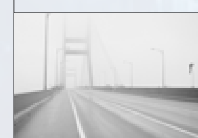
CREATIVE DIRECTOR:
Ben Tolo

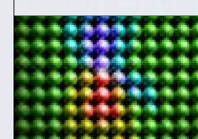
PHOTOGRAPHY:
Jon Chi Lou

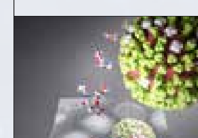


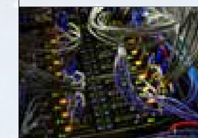


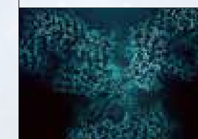


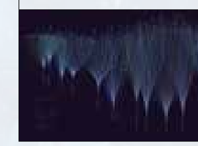


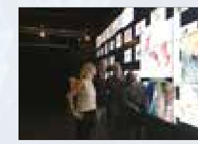












DIRECTOR'S LETTER

Computing without Boundaries

Expanse will be more than just Comet on steroids… [it] will enable researchers to push the boundaries of computing and answer questions previously not possible.

I am delighted to say that we closed the fiscal year on June 30 with a very strong note after having several other successes during the period. In June we received word from the National Science Foundation that SDSC was awarded a $10 million+ grant to deploy an all-new supercomputer we are calling *Expanse*, which is a very fitting name for this new research ecosystem. *Expanse* will eventually succeed our highly successful, NSF-funded *Comet* supercomputer, which will continue through March 2021 as a vital computational resource for researchers around the globe and in just about every area of science.

When it enters production during the second half of 2020, *Expanse* will be more than just *Comet* on steroids, with features such as direct scheduler-integration with major cloud providers, leveraging high-speed networks to ease data movement to and from the cloud, and opening up new modes of computing made possible by the combination of powerful capabilities and ease of connecting to software outside the system. As Principal Investigator, I'm confident that *Expanse* will enable researchers to push the boundaries of computing and answer questions previously not possible. You can read more about *Expanse* on page 34 of this annual report.

SDSC was also successful during the latest fiscal period in procuring several other NSF awards, including three cloud-related awards as we implement our hybrid cloud strategy:

- A five-year grant worth $5 million to SDSC, the University of Washington's eScience Institute, and UC Berkeley's Division of Data Science to develop and operate CloudBank, a set of managed services designed to simplify access to public clouds among the computer science community.
- A two-year grant worth about $400,000 to SDSC to deploy a new system called Triton Stratus as part of our *Triton Shared Computing Cluster* (*TSCC*), which primarily serves UC San Diego researchers and our industry partners. Triton Stratus will provide researchers with improved facilities for accessing emerging computing tools and scaling them to commercial cloud computing resources.
- Our popular CIPRES (CyberInfrastructure for Phylogenetic Research) science gateway was awarded an Internet2 grant funded by the NSF that lets researchers access more powerful processors available from commercial cloud provider AWS to accelerate discovery. Access is free for researchers using the gateway during the one-year period of the award.

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

I'm pleased to report that as of July 1, 2019 SDSC now reports to the Executive Vice Chancellor's Office, with Senior Associate Vice Chancellor (SAVC) Robert Continetti providing direct oversight. SAVC Continetti also provides oversight to other major interdisciplinary units on campus, such as the Design Lab and the Halicioglu Data Science Institute. I look forward to evolving SDSC under this new arrangement. At the same time I'd like to thank Vice Chancellor for Research Sandra Brown for her guidance throughout the time SDSC reported to her office.

I'm also proud to note that several initiatives launched by or involving SDSC have grown in impact from the local or state levels to a national and even global reach. One such example is the Pacific Research Platform, which is serving as a prototype for what is being called the Global Research Platform. You can read more about this high-speed data network on page 7 of this report.

Last but not least, please join me in congratulating Andreas Goetz as SDSC's 2019 π (Pi) Person of the Year. Now in its sixth year, this award recognizes SDSC researchers who have one leg, so to speak, in a science domain and the other in cyberinfrastructure technology. As director of SDSC's Computational Chemistry Laboratory and senior investigator for the Center for Aerosol Impacts on Chemistry of the Environment (CAICE) at UC San Diego, Andy's expertise is in quantum chemistry, molecular dynamics, GPU-accelerated computing, biophysics, drug design, and using machine learning models to predict the behavior of molecular liquids. Read more about Andy's work on page 4.

I invite you to browse through our latest Annual Report to get a sense of the depth and breadth of research activity at SDSC. The year ahead will be a very busy but rewarding one for all of us!

Michael L. Norman
SDSC Director

Welcome to SDSC!

MELISSA CRAGIN JOINS SDSC'S RESEARCH DATA SERVICES GROUP

In April 2019 SDSC announced the appointment of Melissa Cragin as Chief Strategist for SDSC's Research Data Services (RDS) group. Cragin comes from the University of Illinois at Urbana-Champaign (UIUC), where she was executive director of the Midwest Big Data Hub (MBDH) at the National Center for Supercomputing Applications (NCSA) since 2016. "We see potential in more research data management and stewardship as part of research computing services, making it easier for researchers to make their data FAIR (findable, accessible, interoperable, reusable)," said RDS Division Director Christine Kirkpatrick in welcoming Cragin to SDSC.

MICHAEL ZENTNER TO LEAD SDSC'S SUSTAINABLE SCIENTIFIC SOFTWARE GROUP

In June 2019 SDSC announced the appointment of Michael Zentner as director of Sustainable Scientific Software. Zentner joins SDSC following nine years with Purdue University, where he was an Entrepreneur in Residence at the Purdue Foundry and a senior research scientist. He continues as a senior research scientist and director of the HUBzero Platform for Science and Engineering, and remains co-PI of the Network for Computational Nanotechnology project, which operates the nanoHUB.org science gateway. Zentner also was named Principal Investigator of the Science Gateways Community Institute (SGCI), founded in 2016 by SDSC Associate Director Nancy Wilkins-Diehr, who retired June 30 after 26 years at SDSC.

SDSC's 2019 'PI' PERSON OF THE YEAR

MEET ANDREAS GOETZ

Please congratulate Andreas Goetz as SDSC's 2019 π (Pi) Person of the Year! Deputy Director Shawn Strande announced the award at the June 12 General Staff Meeting following a presentation by Andy, who was unaware that he was selected by an internal committee for this year's award. Now in its sixth year, the award recognizes SDSC researchers who have one "leg" in a science domain and the other in cyberinfrastructure technology.

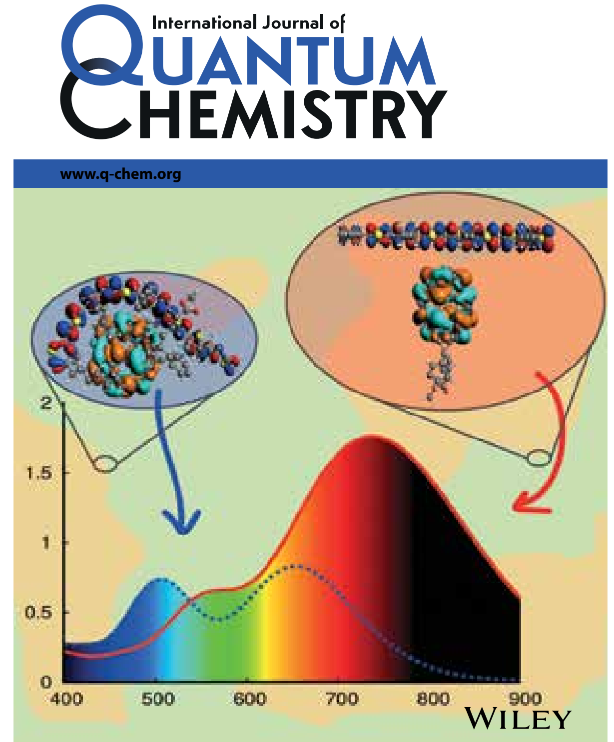
As director of the Computational Chemistry Laboratory at SDSC and senior investigator for the Center for Aerosol Impacts on Chemistry of the Environment at UC San Diego, Andy's expertise is in the areas of quantum chemistry, molecular dynamics, and GPU-accelerated computing. He is also a contributing author of the AMBER and ADF software packages for molecular simulations, both of which are widely used in academic and industrial research settings.

Andy collaborates with UC San Diego researchers, such as Chemist Francesco Paesani in Physical Sciences, but also beyond the university on a variety of projects in molecular simulation, computational catalysis, biophysics, and drug design, as well as using machine learning models and techniques to predict the behavior of molecular liquids.

"Some of our latest research is using smart algorithms as part of the computational process, or what we call an active learning approach," he told SDSC staffers during his presentation. "Predicting quantum energies can be extremely expensive and time consuming in terms of using compute time, so we are taking this new approach to do it in a smarter way."

In June, an image for one of Andy's recent articles was featured on the cover of the *International Journal of Quantum Chemistry*. Calculations were performed on *Comet*, as well as the *Stampede2* system at the Texas Advanced Computing Center (TACC) as part of a collaboration with researchers at the Instituto Politécnico Nacional de México in Mexico City.

In April 2018, a study by Andy and his colleagues at UC San Diego's Department of Chemistry and Biochemistry was published that used machine learning techniques to develop models for simulations of water with what they



called "unprecedented accuracy." Their pioneering work, which appeared in *The Journal of Chemical Physics*, showed how popular machine learning techniques can be used to construct predictive molecular models – in this case water, but applicable also to other "generic" molecules, based on quantum mechanical reference data.

This study was based on the highly accurate and successful "MB-pol" (many-body potential) for water developed in Paesani's lab. MB-pol has recently emerged as an accurate molecular model for water simulations from the gas-to-liquid-to-solid phases. "This is a new methodology that could revolutionize computational chemistry," said SDSC Director Michael Norman.

The development of complex neural networks with associated optimization processes was performed on SDSC's *Comet* supercomputer GPU resources and TACC's *Maverick* system, with allocations provided by the Extreme Science and Engineering Discovery Environment (XSEDE). Also participating in the study were researchers at the École Polytechnique Fédérale de Lausanne in Switzerland, Cambridge University in England, and the University of Göttingen in Germany.

Andy also enjoys teaching the next generation of scientists in computational chemistry, software engineering, and molecular simulation methods. He has been an avid supporter of SDSC's Research Experience for High School Students program (REHS) for 10 years. "It's fun to welcome these motivated young students as interns to the research group."

Before joining SDSC in 2009, Goetz worked as postdoctoral researcher at the VU University in Amsterdam, the Netherlands. He obtained his undergraduate and Ph.D. degrees in chemistry with specialization in theoretical chemistry from the Friedrich-Alexander University in Erlangen, Germany.

In June, SDSC Researcher Hans-Werner Braun received SDSC's Distinguished Achievement Award, which is the highest honor at the Center and given at the sole discretion of the Director. Hans-Werner was recognized for his outstanding vision and leadership of the High Performance Wireless Research and Education Network (HPWREN) which he co-founded almost 20 years ago with Frank Vernon, a research geophysicist with Scripps Institution of Oceanography.

Initially funded by the National Science Foundation, HPWREN has successfully transitioned to a user-funded system while undergoing a significant expansion throughout southern California's most rugged terrain. While many researchers know HPWREN as a network of internet-based high-tech cameras and weather stations that alert local authorities about various conditions including storms, wildfires, and earthquakes, HPWREN has for many years served science and society in a multitude of other ways.

Hans-Werner built quite an eclectic community of users across a wide swath of research domains, from astronomy to wildlife conservation, as well as comprehensive education and training programs for the National Park Service and California State Parks.

"I commend Hans-Werner for his tireless work and dedication to building and maintaining HPWREN as a vital resource for numerous scientific, public safety, and conservation communities across greater San Diego, Southern California, and the nation," said SDSC Director Michael Norman.

IMPACT & INFLUENCE

SDSC'S **NATIONAL MISSION** IN **ADVANCED CYBERINFRASTRUCTURE**

As one of the country's first four supercomputer centers opened in 1985 by the National Science Foundation (NSF), SDSC has an impressive history of programs and partnerships that have benefited science and society across an increasing variety of research domains. SDSC's mission has expanded in recent years to encompass more than just advanced computation, which has served as a foundation to include new and innovative applications and expertise related to the ever-increasing amount of digitally-based science data generated by researchers.

"With regard to our national cyberinfrastucture (CI) mission, we currently have 112 active awards totaling about $140 million," said SDSC Director Michael Norman. "We also generated almost $5.5 million in revenue and support from industry during the latest fiscal year, up 28 percent from the previous year."

SDSC Receives HPCwire's Editors' Choice Award for Top HPC-Enabled Scientific Achievement

In late 2018, SDSC received one of HPCwire's 'Top Achievement' awards, for *Comet's* role in assisting an international team of scientists at the NSF-funded IceCube Neutrino Observatory in finding the first evidence of a source of high-energy cosmic neutrinos. SDSC shared this award with several other organizations, including the IceCube Neutrino Observatory and the NSF's eXtreme Science and Engineering Discovery Environment (XSEDE) program. The awards were presented by HPCwire at the 2018 International Conference for High Performance Computing, Networking, Storage and Analysis (SC18), in Dallas, TX.

QRGO.PAGE.LINK/TYIRA

PACIFIC RESEARCH PLATFORM GOES NATIONAL AND GLOBAL

To meet the needs of researchers in California and beyond, in 2015 the NSF awarded a five-year grant to fund the Pacific Research Platform (PRP), a high-speed data transfer network led by researchers at UC San Diego and UC Berkeley that now connects more than 50 institutions including the Department of Energy and multiple research universities around the world.

The PRP's data-sharing architecture, with end-to-end 10/100 gigabits per second (Gbps) connections, enables region-wide virtual co-location of data with computing resources and enhanced security options. In 2018 and 2019, significant additions were made to the PRP as it met its goal of becoming a global data-intensive research cyber-infrastructure. One was the adoption of Kubernetes, an extensible open-source platform for managing containerized workloads and services that was originally developed by Google.

"These were a real game-changer for us," said Larry Smarr, a UC San Diego computer science and engineering professor, director of the California Institute for Telecommunications and Information Technology (Calit2) at UC San Diego, and principal investigator of the PRP initiative. "In fact the PRP has been a prototype for a National Research Platform or NRP, and now the GRP, for Global Research Platform."

During the period PRP also adopted the Rook Cloud-Native storage Orchestrator, which runs inside the Kubernetes software and allows researchers to store, deliver, and protect their data. "Together, this completes the web revolution to data and software," said Smarr.

The first GRP Workshop was held September 7-18, 2019 at UC San Diego's Qualcomm Institute/Calit2.

SDSC staff involved in the PRP project include Frank Würthwein, a PRP co-PI and lead for distributed high-throughput computing; and Thomas Hutton, a networking architect. Phil Papadopoulos, formerly director of cloud and cluster software development at SDSC, is also a PRP co-PI.

QRGO.PAGE.LINK/LJ4RY

NATIONAL IMPACT & INFLUENCE
Highlights of Key National Partnerships

EXTREME SCIENCE AND ENGINEERING DISCOVERY ENVIRONMENT (XSEDE)

The NSF's XSEDE program allows scientists to interactively share computing resources, data, and expertise. As the only supercomputer center participant on the West Coast, SDSC provides advanced user support and expertise for XSEDE researchers across a variety of applications. SDSC's *Comet* supercomputer is accessible via the XSEDE allocation process to U.S. researchers as well as those affiliated with U.S.-based research institutions. *Comet* is among the most widely used systems in XSEDE's resource portfolio.

WWW.XSEDE.ORG

OPEN SCIENCE GRID CONSORTIUM

OSG is a multi-disciplinary research partnership specializing in high-throughput computational services funded by the U.S. Department of Energy and NSF. Through a partnership with XSEDE, OSG scientists have access to resources such as *Comet* to further their research. The integration of *Comet* into the OSG provisioning system was led by a team including Frank Würthwein, an expert in experimental particle physics and advanced computation, and SDSC's lead for distributed high-throughput computing. Würthwein also serves as OSG's Executive Director. OSG operates services that allow for transparent computation across more than 150 computing clusters worldwide, including National Grid Initiatives in Europe, Asia, and the Americas.

OPENSCIENCEGRID.ORG

SUPPORTING THE NATIONAL BRAIN INITIATIVE THROUGH THE NEUROSCIENCE GATEWAY

Charting brain functions in unprecedented detail could lead to new prevention strategies and therapies for disorders such as Alzheimer's disease, schizophrenia, autism, epilepsy, traumatic brain injury, and more. The BRAIN Initiative (Brain Research through Advancing Innovative Neurotechnologies), launched by President Barack Obama in 2013, is intended to advance the tools and technologies needed to map and decipher brain activity, including advanced computational resources and expertise.

WWW.NSGPORTAL.ORG

NATIONAL DATA SERVICE (NDS)

Founded by a consortium of U.S.-based research computing centers, governmental agencies, libraries, publishers, and universities, NDS builds on data archiving and sharing efforts already underway within scientific communities and links them together with a common set of services. NDS is a vision for how scientists and researchers across all disciplines can find, reuse, and publish data, while providing a workbench for research projects and a training platform for future curators, data stewards, and scientists. Its value is in making it easier to use resources together through dependency management and increased interoperability between cyberinfrastructure services. SDSC has taken a leadership role in NDS through SDSC's Christine Kirkpatrick, the organization's first executive director.

NATIONALDATASERVICE.ORG

SDSC SHERLOCK DIVISION

In spring 2019 SDSC's Sherlock Division announced a partnership with Boise State University to provide a Health Insurance Portability and Accountability Act (HIPAA)-compliant environment to protect the sensitive data needs of Idaho's metropolitan research university of distinction. The partnership evolved due to Boise State's need to protect sensitive data stemming from its collaboration with St. Luke's Health System on a variety of research initiatives involving HIPAA data. SDSC's Sherlock Division, formerly called the Health Cyberinfrastructure (Health CI) Division, specializes in protecting various types of sensitive data. Read more about Sherlock on page 45.

ORGO.PAGE.LINK/HQMX

OPEN STORAGE NETWORK (OSN)

SDSC was one of four partners awarded a $1.8 million grant from the NSF in mid-2018 for a distributed, data storage demonstration project. The team is actively developing the Open Storage Network (OSN), which will allow academic researchers across the nation to work with and share their data more efficiently. The overall project is led by Alex Szalay of Johns Hopkins University, and SDSC's Christine Kirkpatrick. OSN will benefit from integrating with previous NSF investments, such as the CC* project that brought 10/100 Gbps connectivity to several U.S. universities. OSN leverages key data storage partners throughout the U.S. including the National Data Service and the four NSF-funded Big Data Regional Innovation Hubs.

OPENSTORAGENETWORK.ORG

NSF WEST BIG DATA INNOVATION HUB (WBDIH)

The NSF supports four regional Big Data Innovation Hubs throughout the U.S. The Western region, comprised of 13 states with Montana, Colorado, and New Mexico marking the eastern boundary, is led from SDSC, UC Berkeley, and the University of Washington's eScience Institute. In mid-2019 the NSF awarded a second round of funding for hubs, with the goal of convening events and building collaborations where academics, non-profits, industry, and local to federal government agencies collaborate to help solve grand challenges of regional importance through the use of big data and associated methods and technologies. While SDSC Director Michael Norman is the Principal Investigator (PI) of the WBDIH, Christine Kirkpatrick, who leads SDSC's Research Data Services division, is co-PI. "This next phase allows for fine-tuning of offerings that give rise to new collaborations, ideas, and combinations of data resources between academics and government agencies," said Kirkpatrick.

WESTBIGDATAHUB.ORG

GO FAIR

SDSC's Research Data Services (RDS) division is hosting the first GO FAIR coordination office (GFISCO) in the U.S. as part of the division's role in the NDS initiative. GO (Global Open) FAIR is a 'bottom up' initiative aimed at implementing the FAIR principles to ensure that data is findable, accessible, interoperable, and reusable. GO FAIR's main goal is to kick-start the development of a global data commons for research and innovation. With the new U.S. office, the nation joins the Netherlands, Germany, and Brazil in establishing country-level coordinating offices. The US GFISCO hosts FAIR data stewardship training and will host a workshop for training trainers to build FAIR data stewardship, and will serve as a contact point for establishing US-based Implementation Networks (IN).

ORGO.PAGE.LINK/PBYDJ

UNIVERSAL SCALE STORAGE (USS)

In 2019, SDSC's Research Data Services (RDS) division launched the Universal Scale Storage (USS) service for UC San Diego, the UC system, and the local research community. The service provides a single scalable namespace, provisioned per PI, directly mounted on campus research lab systems, SDSC supercomputers, and compute clusters hosted in the SDSC data center. Data is protected with an erasure code that stripes across disks within clustered storage nodes, rendering the storage highly tolerant to hardware failure. Recognizing extensive research demand by a global audience, RDS architected the USS Scale Storage Cloud Connector, which extends local access methods outside the UC San Diego campus, enabling cloud-native applications to interface with USS via the internet. USS cluster network connection speeds were doubled from 40 Gigabit (Gb) per node to 80 Gb, further realizing the advantages of the high-speed SDSC intranet and connections to the wider internet.

ORGO.PAGE.LINK/SC7ML

Providing **Science Gateways** for Researchers

Nancy Wilkins-Diehr was an associate director of SDSC and served as co-PI of XSEDE (eXtreme Science and Engineering Discovery Environment) as well as director of XSEDE's Extended Collaborative Support Services. Nancy retired in mid-2019 after 26 years with SDSC.

In mid-2016, a collaborative team led by SDSC Associate Director Nancy Wilkins-Diehr was awarded a five-year, $15 million National Science Foundation (NSF) grant to establish a Science Gateways Community Institute (SGCI) to accelerate the development and application of highly functional, sustainable science gateways that address the needs of researchers across the full spectrum of NSF directorates and other federal agencies.

Science gateways make it possible to run the available applications on supercomputers such as *Comet* so results come quickly, even with large data sets. Moreover, browser access offered by gateways allows researchers to focus on their scientific problem without having to learn the details of how supercomputers work.

In April 2018, SGCI was approved by an NSF review panel to move into the execution phase. SGCI provided consulting or in-depth expertise to more than 65 projects and written over 100 letters of commitment for interested researchers in the first two-and-a-half years of the program. Proposals with SGCI support have enjoyed a 41% success rate.

SGCI external-facing activities during late 2018 included the Gateways 2018 conference at the Texas Advanced Computing Center in September, as well as implementation of an intense, in-person boot camp program, held twice a year for five-day periods to help attendees building gateways form lasting bonds among peers who share common research challenges. Gateways 2019 was held in San Diego, where SGCI was invited to co-locate with the eScience conference.

In the area of workforce development, SGCI continues with its successful student intern program, in its third year in 2019. Student programs also include a four-week coding institute to develop skills for the internships and several hackathons. After two successful events, SGCI's workforce development lead, Linda Hayden from Elizabeth City State was asked to lead a hackathon event at PEARC19.

"Our core focus is connecting people and resources to accelerate discovery by empowering the entire science gateway community," said Wilkins-Diehr. "Our target market is US-based academic and non-profit students, researchers, and educators who are eager to support their communities. While we make and measure progress one calendar quarter at a time by focusing on prioritized deliverables, SGCI's ultimate goal is to be an autonomous world-class leader and think tank for science gateways."

Wilkins-Diehr retired in mid-2019 after 26 years with SDSC. Succeeding her as PI of SGCI is Michael Zentner, who joined SDSC in June 2019 after nine years with Purdue University. Read more about Michael Zentner on Page 3.

SCIENCEGATEWAYS.ORG

The Gateways conference series is hosted each fall at a location in the US. It offers gateway developers and users an opportunity to connect with and learn from colleagues with diverse experiences, approaches, and academic interests.

Science Gateways Pioneered by SDSC Researchers

SDSC Biology Researcher Mark Miller is PI for the popular CIPRES Science Gateway.

CIPRES

One of the most popular science gateways across the entire XSEDE resource portfolio is the CIPRES science gateway, created as a portal under the NSF-funded CyberInfrastructure for Phylogenetic RESearch (CIPRES) project in late 2009. Based at SDSC, CIPRES allows scientists to explore evolutionary relationships by comparing DNA sequence information between species using supercomputers allocated by the NSF's XSEDE project.

In early 2019, CIPRES was awarded a one-year Internet2 grant funded by the NSF that gives users free access to more powerful compute processors available from commercial cloud provider Amazon Web Services to accelerate their scientific discoveries. Internet2 leads a project called Exploring Clouds for Acceleration of Science (E-CAS) in partnership with representative commercial cloud providers. The project is funded under a collaboration with the NSF to support academic-based researchers who can benefit from access to commercial clouds. Access to AWS will be free for researchers using CIPRES during the period of the award. Please read more about CIPRES on page 39.

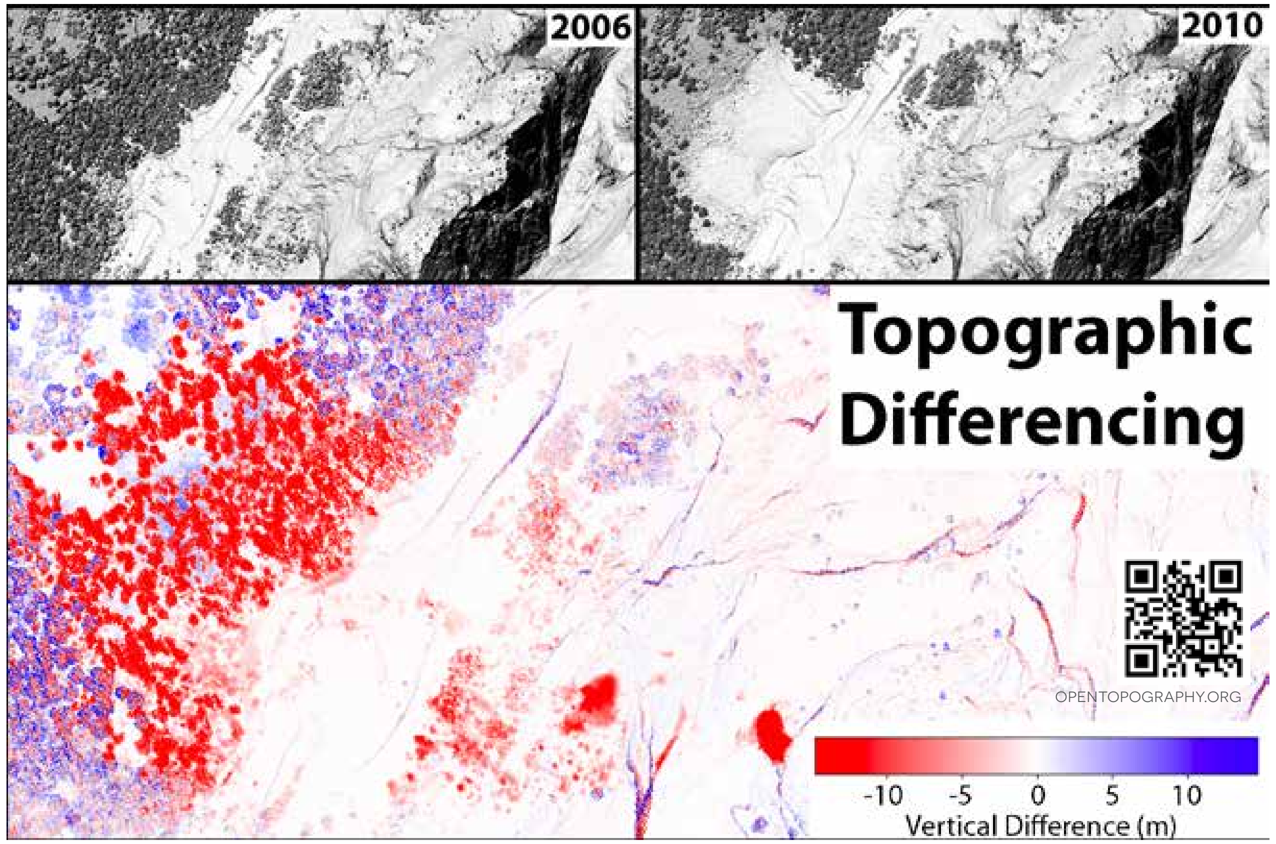
WWW.PHYLO.ORG

Shining Light on Landfills to Uncover 'Dark Life'

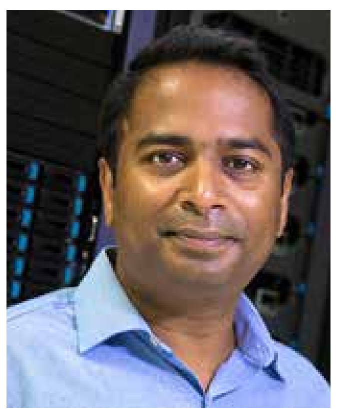
Researchers increasingly turn to computational genomics to assist in bioremediation, a process that uses naturally occurring organisms to help turn hazardous wastes into less toxic or non-toxic materials.

In the YouTube video QR/link above, Assistant Professor Laura Hug at the University of Waterloo in Canada explains how municipal landfills are mined for new types of bacteria–called 'dark life'–because many such organisms were not previously known to exist. Scientists extract DNA from such microbes and sequence it using supercomputers such as SDSC's petascale *Comet* system.

"Now we're getting thousands and thousands of DNA sequences from creatures we've never seen before," says Mark Miller, an SDSC biologist who leads the NSF-funded CIPRES phylogenetics gateway.



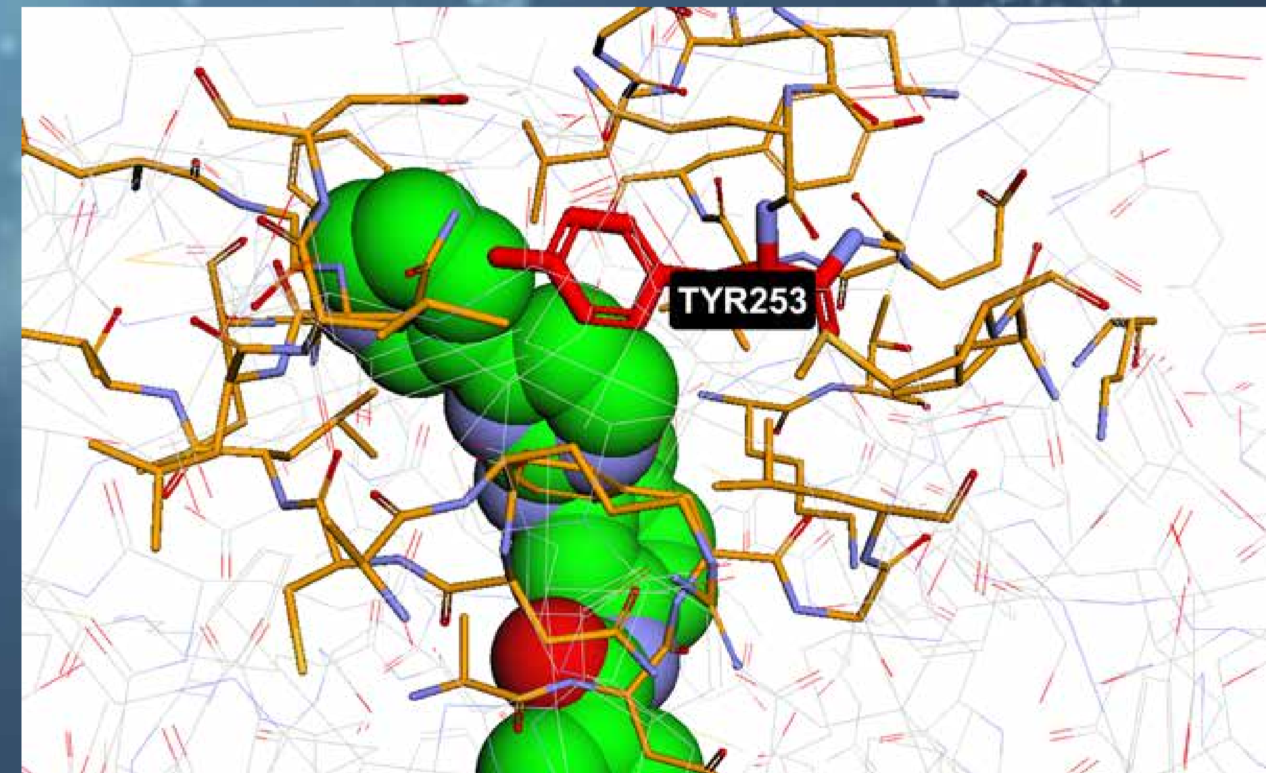
The sheer rock faces and towering mountains of Yosemite National Park may be beautiful, but they're also dangerous. The area has experienced more than 1,000 rockfalls in the past 150 years, posing a threat to people and buildings. Arizona State University researcher Chelsea Scott used lidar imaging data from OpenTopography to map where rocks and trees fell at Yosemite between 2006 and 2010. The bottom image shows the movement of rock from areas that lost height (red) to those that gained height (blue). Getting a clearer picture of the most unstable areas can help park managers keep visitors out of harm's way. *Credit: Chelsea Scott, Arizona State University.*



Viswanath Nandigam is PI and chief software architect for the Open-Topography Facility. He is also associate director of SDSC's Advanced Cyberinfrastructure Development group.

OPENTOPOGRAPHY

Initiated in 2009 with funding from the NSF, OpenTopography provides easy access to earth science-oriented, high-resolution topographical data and processing tools for a broad spectrum of research communities. As a collaboration between UC San Diego, Arizona State University and UNAVCO, OpenTopography employs sophisticated cyberinfrastructure that includes large-scale data management, distributed high-performance computing (HPC), and service-oriented architectures, providing researchers with efficient web-based access to large, high-resolution topographic datasets. OpenTopography's current data holdings comprise 292 lidar point cloud datasets with over 1.27 trillion lidar returns and 164 high-resolution raster datasets covering 242,646 km2, and six global datasets including the highly popular Satellite Radar Topography Mission (SRTM) global 30m dataset. The recently released Community DataSpace that lets researchers publish and distribute their "long tail" topography data (of modest size but great value) via the OpenTopography platform has over 30 national and international datasets. OpenTopography has established itself as resource that supports a broad interdisciplinary user base in academia and beyond. Since its initiation, over 97,000 unique users have run more than 372,000 jobs via the gateway portal. An additional 455,750 jobs were run via available API. More than 400 peer-reviewed articles and other publications have been produced using OpenTopography resources. These include academic works in Earth science, ecology, hydrology, geospatial and computer science, and engineering.



Analysis of the effects of mutations on drug binding run in Jupyter Notebook. The drug molecule Imatinib (Gleevec) for the treatment of chronic myeloid leukemia is shown in green, bound to the protein c-Abl kinase (PDB ID: 1IEP). The drug binding site is shown in orange. An amino acid (red) will affect drug binding as a result of a mutation reported in the ClinVar database. *Image courtesy of Peter Rose, SDSC.*

Peter Rose is the director of the Structural Bioinformatics Laboratory at SDSC, and leads Bioinformatics and Biomedical Applications for the SDSC Data Science Hub.

Reproducible and Scalable Structural Bioinformatics

Scientists face time-consuming barriers when applying structural bioinformatics analysis, including complex software setups, non-interoperable data formats, and lack of documentation – all which make it difficult to reproduce results and reuse software pipelines. A further challenge is the ever-growing size of datasets that need to be analyzed. To address these challenges, SDSC's Structural Bioinformatics Laboratory, directed by Peter Rose, is developing a suite of reusable, scalable software components called MMTF-PySpark, using three key technologies: its parallel distributed processing framework provides scalable computing; the MacroMolecular Transmission Format (MMTF), a new binary and compressed representation of Macromolecular structures, enables high-performance processing of Protein Data Bank structures; and Jupyter Notebooks is used to bundle code, 2D and 3D visualization, machine learning tools, and more into reproducible and reusable workflows. "The use of MMTF-PySpark could easily shave off a year of a graduate student's or postdoc's work in Structural Bioinformatics," said Rose. "We bank on contributions from the community to develop and share an eco-system of interoperable tools."

MMTF.RCSB.ORG

IMPACT & INFLUENCE

STATE AND UC ENGAGEMENT
ALIGNING WITH PRINCIPLES & PARTNERSHIPS

HPC@UC

This novel program provides UC researchers access to SDSC's supercomputing resources and expertise. Since its 2016 inception it has assisted some 50 individual research projects spanning all 10 UC campuses, with about 18 million core-hours of compute time allocated on SDSC's petascale *Comet* supercomputer funded by the National Science Foundation (NSF). In the last year, SDSC added 13 new projects in the areas of pharmacology, neurological surgery, physics, nanoengineering, astronomy, energy research, climate science, biology etc. HPC@UC is offered in partnership with the UC Vice Chancellors of Research and campus CIOs. The program has helped UC researchers accelerate their time-to-discovery across a wide range of disciplines, from astrophysics and bioengineering to earth sciences and machine learning. The program is specifically intended to:

- Broaden the base of UC researchers who require advanced and versatile computing,
- Seed promising computational research,
- Facilitate collaborations between SDSC and UC researchers,
- Give UC researchers access to cyberinfrastructure that complements what is available at their campus, and
- Help UC researchers pursue larger allocation requests through the NSF's eXtreme Science and Engineering Discovery Environment (XSEDE) program and other national computing initiatives.

WWW.SDSC.EDU/COLLABORATE/HPC_AT_UC.HTML

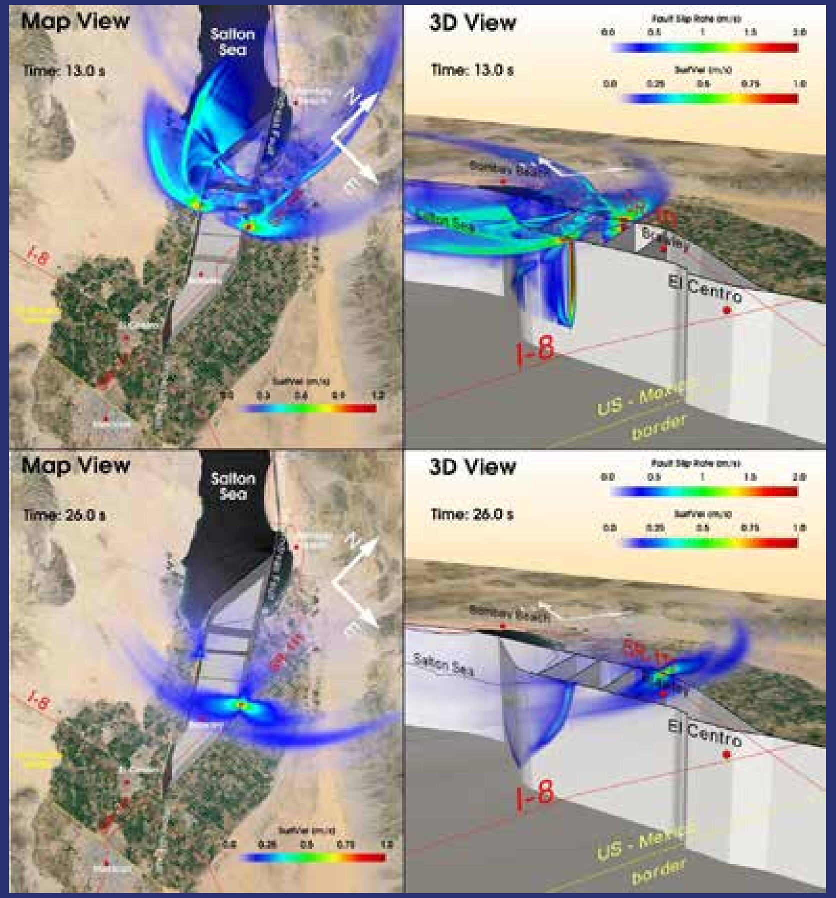
HIGH-PERFORMANCE COMPUTING AND DATA SCIENCE WORKSHOPS ACROSS THE UC SYSTEM

SDSC staff conducts numerous workshops during the fiscal year to raise awareness among UC researchers about the advantages of using high-performance computing resources such as *Comet* as well as understanding of data science. These sessions, usually just one day, are led by research staff from SDSC's Data-Enabled Scientific Computing (DESC) division and Cyberinfrastructure Research, Education and Development (CI-RED) division. SDSC speakers who teach at these workshops have Ph.Ds. in physics, astrophysics, aerospace engineering, computer science, cognitive science, and more. They have attracted several hundred attendees including graduate students, faculty, and post docs at the UCs. Topics covered include machine learning using *Comet*, scaling and optimization of scientific applications on HPC, software tools for life sciences applications, and working with Python and Jupyter notebooks. The workshops, which promote interaction among UC researchers and serve as an easy "on-ramp" for allocations on *Comet*, have been conducted since the start of the UC@SDSC program in 2014.

Scientists are using supercomputers to better predict the behavior of the world's most powerful, multiple-fault earthquakes. A science team used simulations to find dynamic interactions of a postulated network of faults in the Brawley seismic zone in southern California. Map (left panels) and 3D (right panels) views of supercomputer earthquake simulations show how different stress conditions affect rupture propagation across the complex network of faults. The top panels show a high-stress case scenario (leading to very fast rupture propagation, higher than the S wave speed) while the bottom panels show a medium stress case simulation.

Credit: Christodoulos Kyriakopoulos, UC Riverside

YOUTU.BE/P5F3JSCKPZM



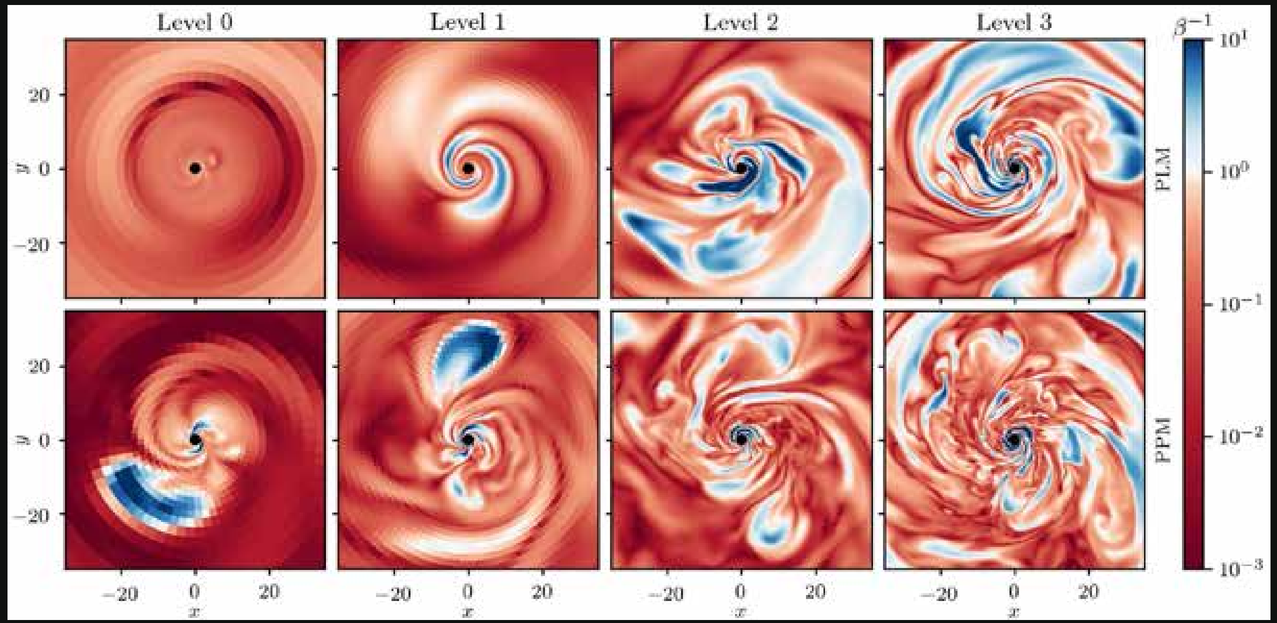
SDSC's Comet Supports UC Riverside Study of San Andreas Fault System

Aerial photo of San Andreas Fault looking northwest onto the Carrizo Plain.

Credit: John L. Wiley, flickr.com/jw4pix

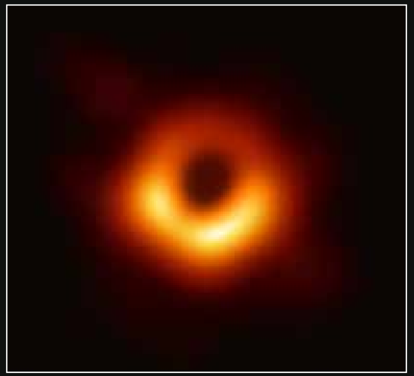
Multi-fault earthquakes can span fault systems of tens to hundreds of kilometers, with ruptures propagating from one segment to another. During the last decade, seismologists have observed several cases of this complicated type of earthquake rupture, and are relying on supercomputers to provide detailed models so they can better understand the physical processes that occur. Some of the findings point to the possibility of a multi-fault earthquake in Southern California, which could have dire consequences and impact the lives of millions of people in the U.S. and Mexico. The simulations focused on the dynamic interactions of a postulated network of faults in the Brawley seismic zone in Southern California. "We found that a rupture on the Southern San Andreas Fault could propagate south of Bombay Beach, considered to be the southern end of the San Andreas Fault," said Christodoulos Kyriakopoulos, a research geophysicist at the University of California Riverside and lead author of a study published in the *Journal of Geophysical Research: Solid Earth*. Supercomputers are able to simulate the interactions between different earthquake faults. These models showed researchers how seismic waves travel from one fault to another and influence the stability of other faults. "Such models are useful to investigate large-scale earthquakes of the past, and perhaps more importantly, possible earthquake scenarios of the future," said Kyriakopoulos.

Read more about how SDSC is advancing earthquake research on page 51.



These *Comet*-enabled images depict the hot, thick disk of plasma surrounding a black hole. As resolution and reconstruction order increases, more and thinner sheets of plasma appear while highly magnetized bubbles become more numerous.

Credit: Christopher White, UC Santa Barbara.



Using the Event Horizon Telescope, scientists obtained an image of the black hole at the center of galaxy M87.

Credit: Event Horizon Telescope collaboration et al.

UC Scientists Use Comet to Simulate Black Holes and their Magnetic Bubbles

In April 2019, when the Event Horizon Telescope team released the first picture ever taken of a black hole – one that resides in the Messier 87 (M87) galaxy some 55 million light years from Earth, that image spawned questions from curious cosmologists around the world. One of them was Christopher White, a computational astrophysicist at the Kavli Institute for Theoretical Physics at UC Santa Barbara. White and his colleagues – Eliot Quataert, a professor of astronomy and physics at UC Berkeley, and James Stone, chair of the Department of Astrophysical Sciences at Princeton University – published a study in the April 2019 issue of *The Astrophysical Journal* about their visualizations after logging some 760,000 core-hours on SDSC's *Comet* supercomputer. The study focused on the processes by which matter falls into these supermassive black holes in the centers of galaxies. This specific scenario, similar to what scientists think occurred with the M87 black hole, involves a hot, thick disk of plasma and a strong magnetic field. Scientists have long believed that this magnetic field is responsible for launching the extremely powerful jets of particles and energy associated with some black holes.

QRGO.PAGE.LINK/U72FN

IMPACT & INFLUENCE

Image credit: Erik Jepsen, UC San Diego Publications

CAMPUS AND EDUCATION
STRENGTHENING TIES ACROSS CAMPUS AND OUR LOCAL COMMUNITIES

SDSC's strategic plan is in close alignment with the top priorities of UC San Diego. Toward that end, SDSC has been focused on identifying and expanding its education, outreach, and training (EOT) initiatives at the undergraduate/graduate/post-doc levels in the areas of research-based computing. "SDSC's development and operation of high-performance computing (HPC) resources at the national level provides substantial and tangible benefits to UC San Diego researchers, as well as San Diego's burgeoning research infrastructure," said SDSC Director Michael Norman.

EMPOWERING THE NEXT GENERATION OF RESEARCHERS

SDSC's EOT programs range from grade school to high school as a catalyst for making students aware of opportunities within computational science at an early age, and then at the university levels with numerous data-centric courses including those in collaboration with the Halicioğlu Data Science Institute (HDSI), which is based in SDSC's East Building, as well as online courses that attract participants from all over the world. This initiative also extends into serving the growing computational science workforce with workshops such as SDSC's Summer Institute, the International Conference on Computational Science, IEEE Women in Data Science Workshop, and more.

EDUCATIONAL IMPACT & INFLUENCE
ACROSS CAMPUS AND OUR LOCAL COMMUNITIES

ON THE LOCAL FRONT

RESEARCH EXPERIENCE FOR HIGH SCHOOL STUDENTS

The Research Experience for High School Students (REHS) program, which celebrated its 10th year in 2019, was developed to help increase awareness of computational science and related fields of research among students in the greater San Diego area. During the eight-week summer program, students are paired with SDSC mentors to help them gain experience in an array of computational research areas, while also gaining exposure to career options and work readiness skills. They learn how to formulate and test hypotheses, conduct computational experiments and draw conclusions from those experiments, and effectively communicate the science and societal value of their projects to a wide range of audiences. To date, more than 400 students have participated in REHS.

MENTOR ASSISTANCE PROGRAM

San Diego-area high school students interested in pursuing a career in scientifically-based research are invited to apply to UC San Diego's Mentor Assistance Program (MAP), a campus-wide initiative designed to engage students in a mentoring relationship with an expert from a vast array of disciplines. Launched five years ago by SDSC and UC San Diego School of Medicine, MAP's mission is to provide a pathway for student researchers to gain access to UC San Diego faculty, post-doctoral fellows, Ph.D. candidates, and staff to mentor them in their own field of interest. Mentors are recruited from across campus from fields that include athletics, biology, chemistry, aerospace engineering, network architectures, pharmaceutical sciences, physics, social studies, and more.

UCSD SUPERCOMPUTING CLUB

In early 2019, SDSC hosted a meeting for students who are interested in HPC. More than 100 students attended, including undergraduates and graduates from over 20 fields/majors. In response to this interest the students formed a Supercomputing Club, and SDSC started the HPC Students Program to support student activities at SDSC. The Supercomputing Club is an independent UC San Diego club, organized and run by university students and sponsored by SDSC. Club activities include organizing club meetings and topics-based workshops; learning to build and administer clusters including updating Meteor, a Raspberry PI visualization cluster used by the visual arts department; and fostering interactions between students and SDSC staff. Near-term plans include hosting an NVIDIA machine learning workshop and a Big Data Hackathon event, as well as a Student Research Expo and 'Soft Skills' workshop focused on career development.

HPC STUDENTS PROGRAM

The HPC Students Program focuses on activities such as organizing, coordinating, and supporting Club activities; purchasing/loaning tool and cluster hardware to the Club; and sponsoring students to travel to the annual Supercomputing (SC) conference. This program also hosts the Student HPC Training classes in collaboration with the Club, where students are taught to design, build, and run small clusters, and learn to run scientific applications on those systems. The program also organizes and awards Co-Curricular Record (CCR) credits to SDSC interns while assisting PIs to create new CCRs.

ON THE NATIONAL FRONT

ONLINE DATA SCIENCE & BIG DATA

UC San Diego offers a four-part Data Science series via edX's MicroMasters® program with instructors from the campus' CSE Department and SDSC. In partnership with Coursera, SDSC created a series of MOOCs (massive open online courses) as part of a Big Data Specialization that has proven to be one of Coursera's most popular data course series. Consisting of five courses and a final Capstone project, this specialization provides valuable insight into the tools and systems used by big data scientists and engineers. In the final Capstone project, learners apply their acquired skills to a real-world big data problem. To date, the courses have reached more than a million students in every populated continent – from Uruguay to the Ivory Coast to Bangladesh. A subset of students pay for a certificate of completion.

HPC WEBINAR & WORKSHOP SERIES

SDSC's HPC training webinar and workshop series focuses on researchers who would like to learn more about using *Comet* and the *Triton Shared Computing Cluster* (*TSCC*). Since early 2018 there have been 11 training events with hundreds of participants from both UC San Diego and nationally via the XSEDE program. Topics included running parallel jobs on HPC systems, GPU computing, parallel computing with Python, Python for data scientists, machine learning, parallel visualization, using Singularity containers for HPC, and using Jupyter Notebooks for HPC and data science. As of September 2019, monthly training events began with the addition of topics such as more Jupyter Notebook technologies, interactive computing, and cloud computing.

HPC SUMMER INSTITUTE

SDSC's Summer Institute is an annual week-long training program offering introductory to intermediate topics on high-performance computing (HPC) and data science, with interactive classes and hands-on tutorials using SDSC's *Comet* supercomputer. The program has been expanded to cover new topics such as machine learning at scale, distributed programming in Python, cluster computing with Spark, data management and CUDA programming, all while retaining traditional HPC topics such as performance tuning, visualization and parallel programming with MPI and OpenMP. The Summer Institute is aimed at researchers in academia and industry, especially in domains not traditionally engaged in supercomputing, and who may have challenges that cannot typically be solved using local computing resources. The 2019 event had more than 100 applicants from nearly 70 institutions/companies. Generally, about half of the attendees are graduate students, with the remainder being post-docs, professors, and research staff at universities, national labs, and industry. SDSC has conducted the Summer Institute since the mid-1990s.

EDUCATION.SDSC.EDU

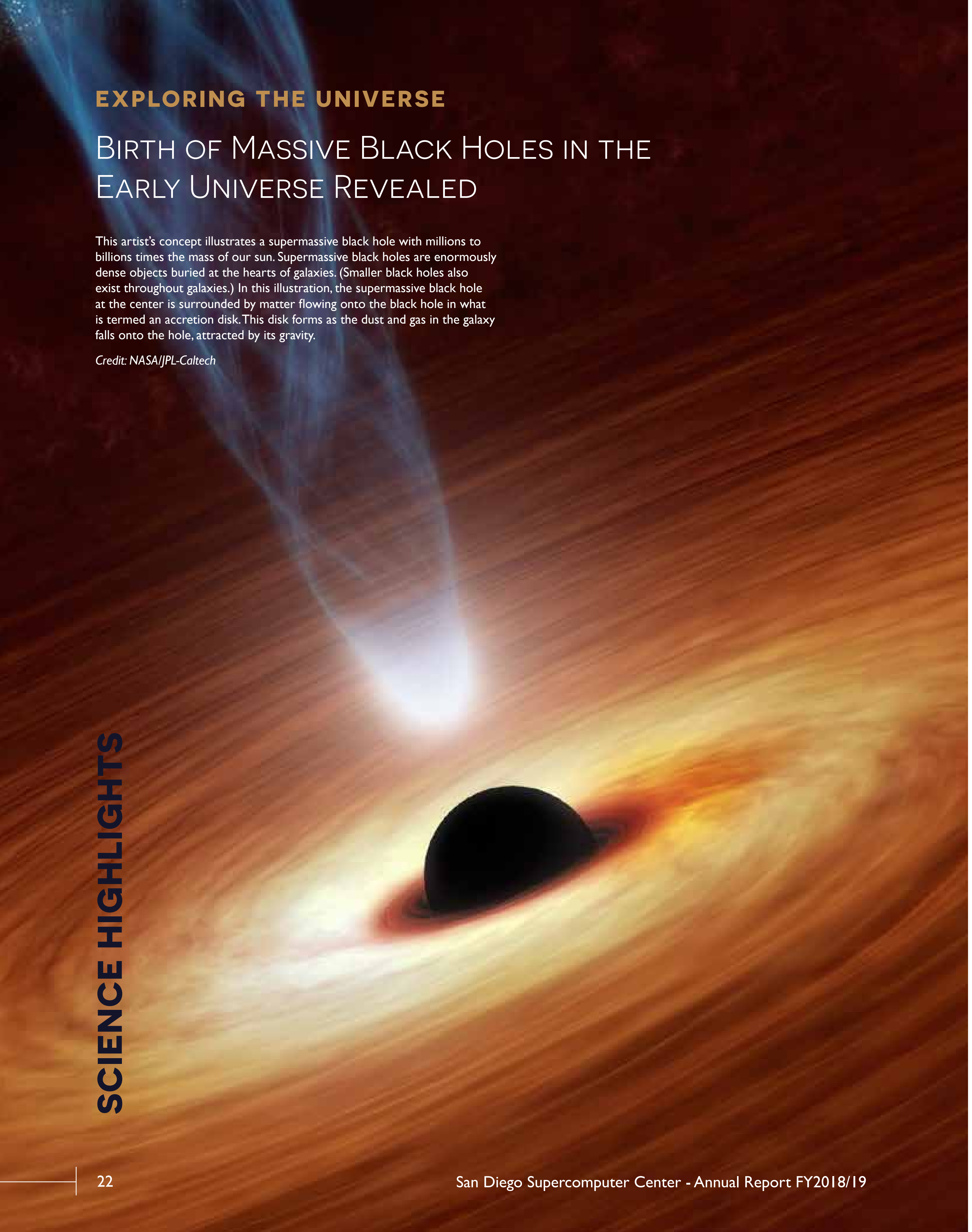
Validating a model's performance is accomplished by comparing its prediction of school locations with UNICEF's data. The orange squares represent tiles identified by the model as likely school locations, and the pink circles are actual schools from UNICEF's data.

Credit: Dan Crawl, SDSC.

SDSC Researchers Team with UNICEF for Liberian Schools Project

In the aftermath of Liberia's second civil war, which lasted 14 years and claimed some 250,000 lives before ending in 2003, the United Nations Children's Fund (UNICEF) has been working with SDSC and other parts of UC San Diego to determine the location of existing schools in the West African country so they can provide them with resources and work with policy makers to plan for additional ones to help rebuild their education infrastructure. In presenting their findings at the 11th IEEE/ACM International Conference on Utility and Cloud Computing in December 2018, SDSC researchers focused on their use of deep learning, a subfield of artificial intelligence that makes use of computational models that can automatically learn patterns from data instead of being explicitly programmed. "Using a particular type of deep learning model called a convolutional neural network (CNN), we extracted salient features from the satellite image tiles," said Mai Nguyen, a senior SDSC data scientist. "We then applied cluster analysis to organize the image tiles into clusters where tiles with similar features were grouped together. Clusters with the majority of tiles containing schools were then identified using ground truth data (on-site information provided by UNICEF) if available or by visual inspection of the cluster contents. These school clusters were then used to identify likely locations of schools in a region."

QRGO.PAGE.LINK/ENHT2



EXPLORING THE UNIVERSE

Birth of Massive Black Holes in the Early Universe Revealed

This artist's concept illustrates a supermassive black hole with millions to billions times the mass of our sun. Supermassive black holes are enormously dense objects buried at the hearts of galaxies. (Smaller black holes also exist throughout galaxies.) In this illustration, the supermassive black hole at the center is surrounded by matter flowing onto the black hole in what is termed an accretion disk. This disk forms as the dust and gas in the galaxy falls onto the hole, attracted by its gravity.

Credit: NASA/JPL-Caltech

SCIENCE HIGHLIGHTS

New findings by astrophysics researchers from the Georgia Institute of Technology, Dublin City University in Ireland, Michigan State University, UC San Diego, and SDSC provides an extremely promising avenue for solving the cosmic riddle of how the first black holes in the universe formed, according to conclusions of a simulation-based study published in early 2019 in the journal *Nature*.

The team showed that when galaxies assemble extremely rapidly – and sometimes violently – that can lead to the formation of very massive black holes. In these rare galaxies, normal star formation is disrupted and black hole formation takes over.

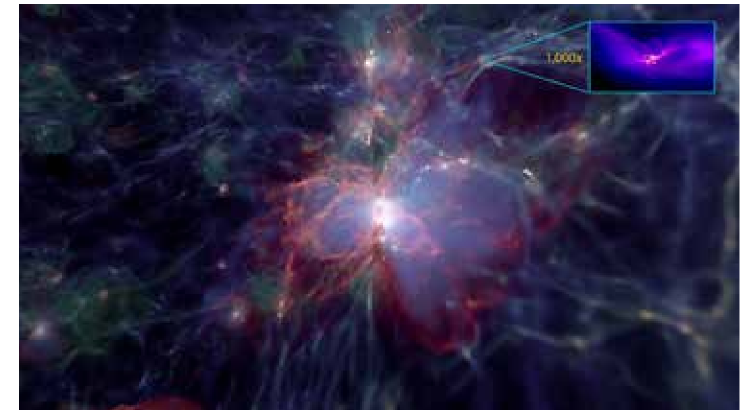
The new study found that massive black holes form in dense starless regions that are growing rapidly, challenging the long-accepted belief that massive black hole formation was limited to regions bombarded by the powerful radiation of nearby galaxies. The research also found that massive black holes are much more common in the universe than previously thought.

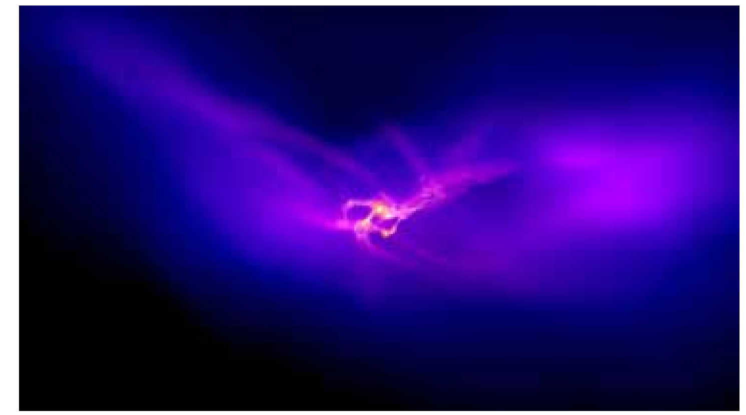
The earlier theory relied on intense ultraviolet radiation from a nearby galaxy to inhibit the formation of stars in the black hole-forming halo, according to SDSC Director Michael Norman, a co-author of the study and a Distinguished Professor of Physics at UC San Diego. "While UV radiation is still a factor, our work has shown that it is not the dominant factor, at least in our simulations."

When the research team found these black hole formation sites in the simulation they were at first stumped, said John Regan, research fellow in the Centre for Astrophysics and Relativity in Dublin City University. The previously accepted paradigm was that massive black holes could only form when exposed to high levels of nearby radiation.

"Previous theories suggested this should only happen when the sites were exposed to high levels of star-formation killing radiation," said Regan. "As we delved deeper, we saw that these sites were undergoing a period of extremely rapid growth. That was the key. The violent and turbulent nature of the rapid assembly, the violent crashing together of the galaxy's foundations during the galaxy's birth prevented normal star formation and led to perfect conditions for black hole formation instead. This research shifts the previous paradigm and opens up a whole new area of research."

The research was based on the Renaissance Simulations suite, the most comprehensive simulations of the earliest stages of the gravitational assembly of the pristine gas composed of hydrogen and helium and cold dark matter leading to the formation of the first stars and galaxies. They use a technique known as adaptive mesh refinement to zoom in on dense clumps forming stars or black holes.





(Top) A 30,000 light-year region from the Renaissance Simulation centered on a cluster of young galaxies that generate radiation (white) and metals (green) while heating the surrounding gas. A dark matter halo just outside this heated region forms three supermassive stars (inset) each over 1,000 times the mass of our sun that will quickly collapse into massive black holes and eventually supermassive black holes over billions of years. (Bottom) Zoom of the inner 30 light-years of the dark matter halo. The rotating gaseous disk breaks apart into three clumps that collapse under their own gravity to form supermassive stars.

Credit: Advanced Visualization Lab, National Center for Supercomputing Applications; John Wise, Georgia Institute of Technology

"The high resolution, rich physics, and large sample of collapsing halos were all needed to achieve this result," said SDSC's Norman. "These simulations are sufficiently rich so that other discoveries can be made using data already computed, and for this reason we created a public archive at SDSC called the Renaissance Simulations Laboratory, where others can pursue questions of their own."

QRGO.PAGE.LINK/6GFCB

EARTH SCIENCES

SDSC Supercomputer Simulations Reveal New Insight on Sea Fog Development

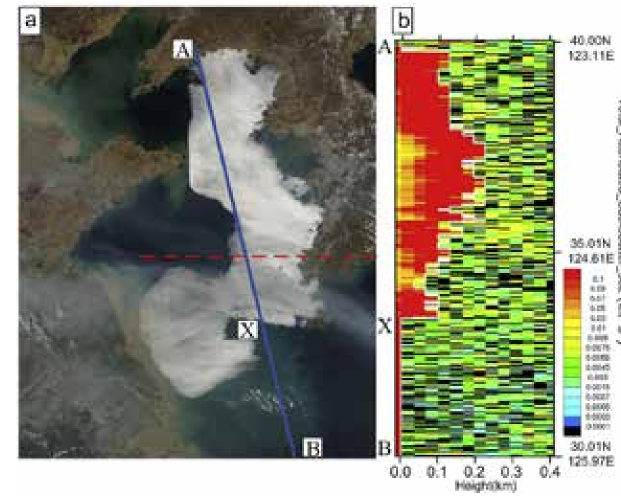
SCIENCE HIGHLIGHTS

A new study by an international team of researchers has shed new light on how and why a particular type of sea fog forms, using detailed supercomputer simulations to provide more accurate predictions of its occurrence and patterns to help reduce the number of maritime mishaps.

The study, published in the January 1, 2019 issue of *Atmospheric Research*, focused on a significant sea fog event that stretched approximately 400 miles across China's Yellow Sea and its surrounding land. Images captured by the Moderate Resolution Imaging Spectroradiometer (MODIS) on the Aqua satellite allowed the research team to better explain the sensitivity of sea fog simulations to vertical resolution. Researchers used SDSC's *Comet* supercomputer to develop detailed simulations.

NASA data affirms that spring fogs can occur for 50-80 days per year across the Yellow Sea. The fogs occur within the marine atmospheric boundary layer and are responsible for as many as half of the shipping accidents at the Qingdao port, according to the Chinese government. The primary type of fog covering the Yellow Sea is called advection fog, which usually forms when warm, moist air moves into areas with a cold surface.

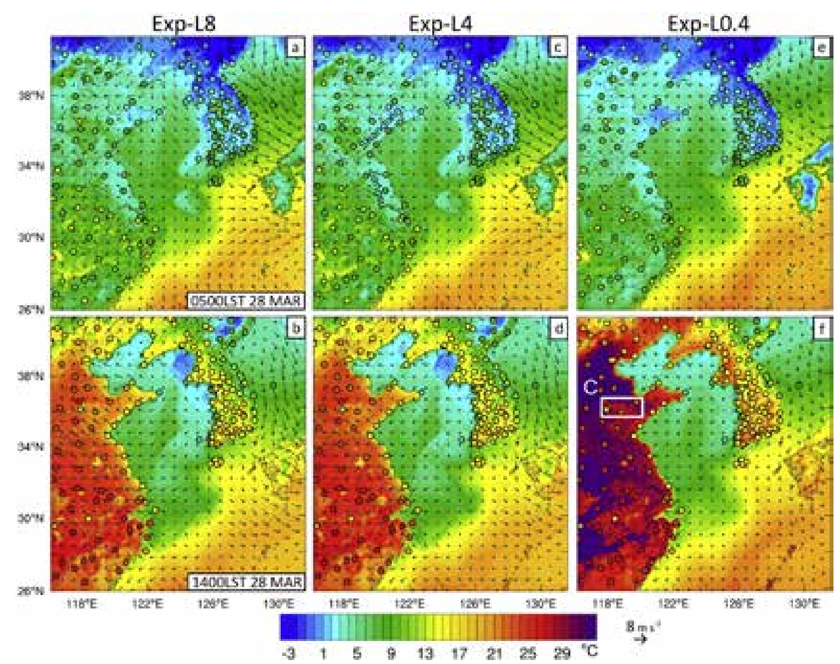
"Inaccurate predictions of sea fog over the Yellow Sea pose a severe risk or damage to many human activities such as aviation, shipping, and the general business economy," explained Xiao-Ming Hu, principal investigator for the study who is with the Center for Analysis and Prediction of Storms (CAPS) at University of Oklahoma. "Accurate predictions of sea fogs involve a few parameterizations for the atmospheric boundary layer and microphysics processes, all of which requires appropriate resolutions. Finding the optimal configurations among resolutions and parameterization schemes is not only beneficial for predicting sea fogs as demonstrated in this study, but also helps researchers to better understand and predict other meteorological events."



This image near Qingdao, from the CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations) satellite, shows the vertical cross-section of backscattering coefficients (indication of water droplets, i.e., sea fog) along the blue line AB.

Credit: Center for Analysis and Prediction of Storms (CAPS), University of Oklahoma

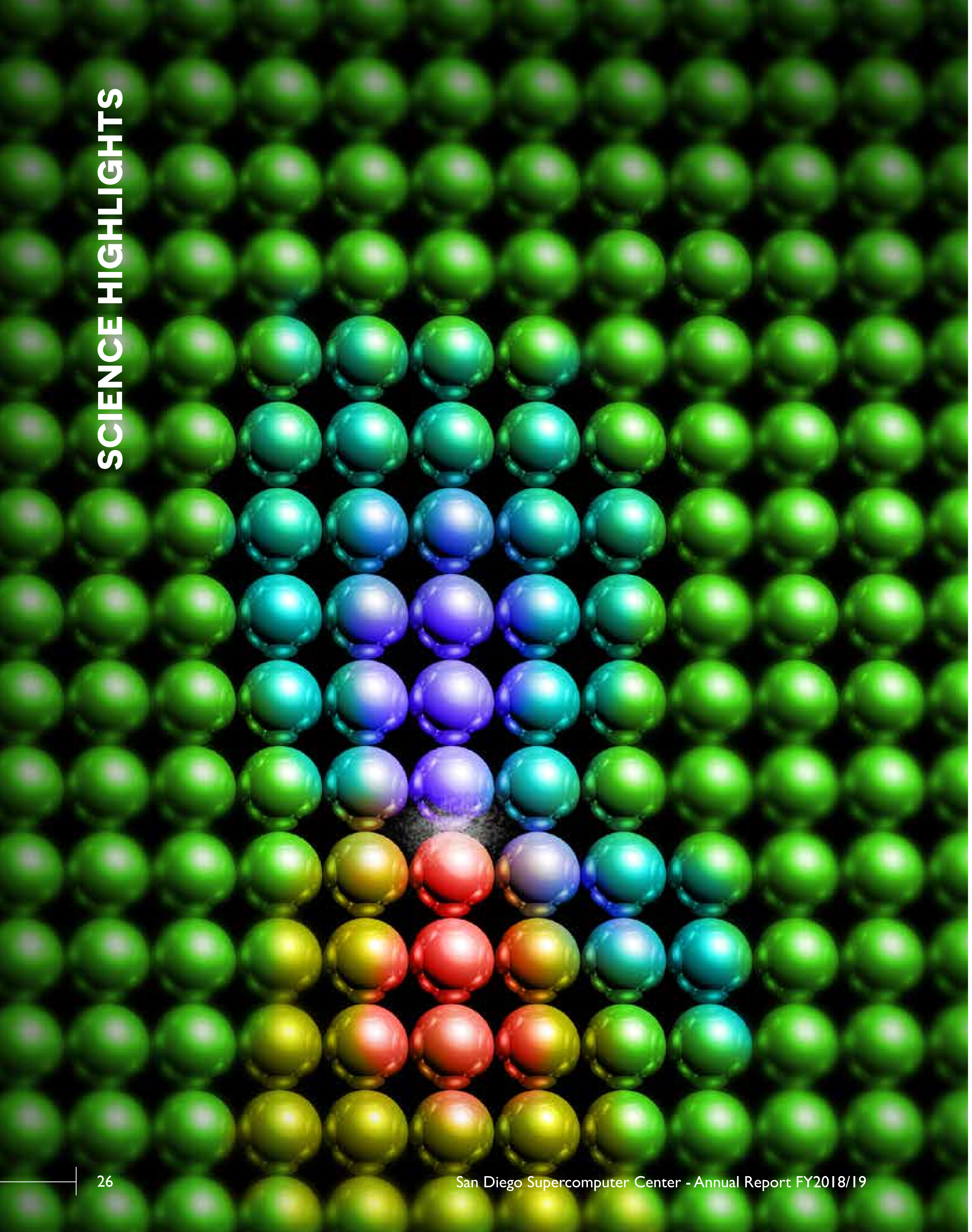
The researchers have been using SDSC's *Comet* supercomputer since 2016, with allocations obtained via the National Science Foundation's XSEDE (eXtreme Science and Engineering Discovery Environment) program. To date, some 2 million core hours have been used.



Evaluation of surface temperatures simulated with different vertical resolutions (from left to right) using station data over eastern Asia archived by NOAA. This helped researchers identify the optimal vertical resolution for accurately simulating sea fog.

Credit: Center for Analysis and Prediction of Storms (CAPS), University of Oklahoma

QRGO.PAGE.LINK/JM2VU



MATERIALS ENGINEERING

A New Way to See Stress

Contributing author, Jorge Salazar

Scientists take a lot for granted when they study the mechanics of stress, in this case the force per unit area on an object. They view stress mathematically by assuming it to have symmetry, meaning the components of stress are identical if one transforms the stressed object with something such as a turn or a flip.

Yet simulations done on supercomputers including *Comet* at SDSC, show that at the atomic level, material stress doesn't behave symmetrically. Those findings could help scientists design new materials, such as glass or metals that don't ice up.

"The commonly accepted symmetric property of a stress tensor in classical continuum mechanics is based on certain assumptions, and they will not be valid when a material is resolved at an atomistic resolution," according to Liming Xiong, an assistant professor in the Department of Aerospace Engineering at Iowa State University, who co-authored a study published in September 2018 in the *Proceedings of the Royal Society A*.

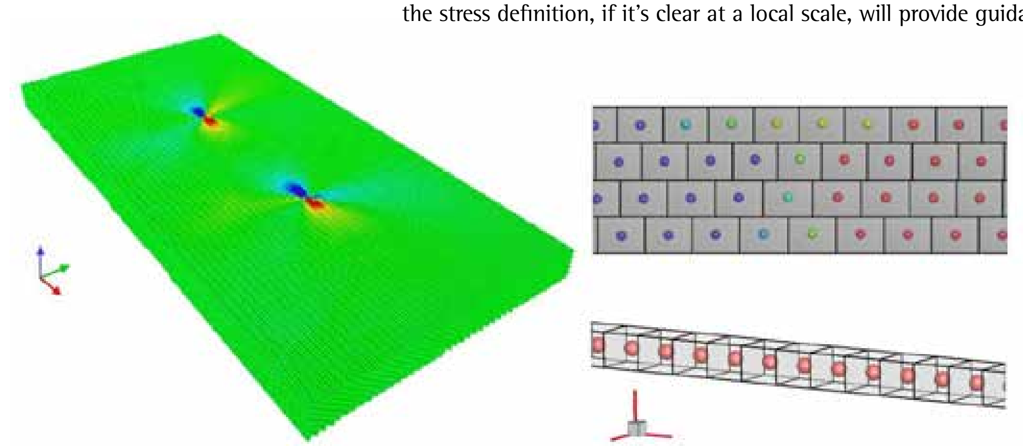
Xiong and colleagues treated stress in a different way than classical continuum mechanics. Instead, they used the definition by mathematician A.L. Cauchy of stress as the force per unit area acting on three rectangular planes. With that, they conducted molecular dynamics simulations to measure the atomic-scale stress tensor of materials with inhomogeneities caused by dislocations, phase boundaries, and holes.

The computational challenges near the limits of what's currently computable when dealing with atomic forces interacting inside a tiny fraction of the space of a raindrop. Add that to the lack of a well-established computer code that can be used for the local stress calculation at the atomic scale. So Xiong's team used the open-source LAMMPS Molecular Dynamics Simulator, incorporating the Lennard–Jones interatomic potential and modified through the parameters they worked out in the paper. Once the code was developed and benchmarked, it was ported to *Comet* to perform large-scale simulations using hundreds to thousands of processors.

"We were trying to meet two challenges – one was to redefine stress at an atomic level and the other was could we use supercomputer resources to calculate it?" said Xiong.

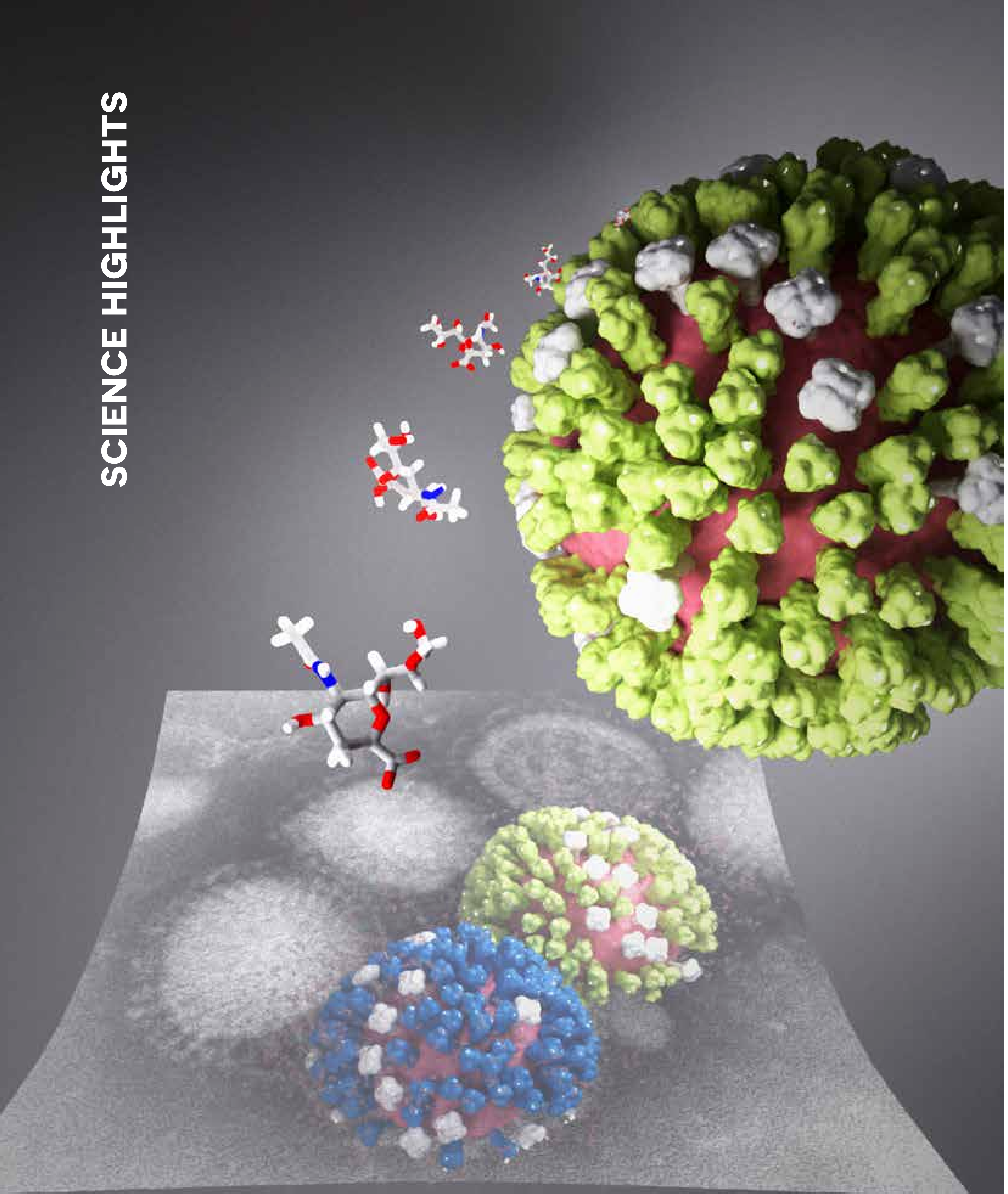
Xiong and his research group are working on several projects to apply their understanding of stress to design new materials with novel properties. "One of them is de-icing from the surfaces of materials, such as ice on a car window in cold weather." The force and energy required to remove that ice is related to the stress tensor definition and the interfaces between ice and the car window. "Basically, the stress definition, if it's clear at a local scale, will provide guidance to use in our daily lives."

QRGO.PAGE.LINK/JTTS8



(Left) Supercomputer simulations show that at the atomic level, material stress doesn't behave symmetrically. Molecular model of a crystal containing a dissociated dislocation; atoms are encoded with the atomic shear strain. (Right) Snapshots of simulation results showing the relative positions of atoms in the rectangular prism elements; each element has dimensions 2.556 Å by 2.087 Å by 2.213 Å and has one atom at the center.

Credit: Liming Xiong



SDSC SIMULATIONS HELP RESEARCHERS PROPOSE A NOVEL APPROACH TO BATTLING INFLUENZA

Rommie Amaro is a professor of chemistry and biochemistry at UC San Diego and director of the National Biomedical Computation Resource.

Every year, three to five million people around the world suffer from severe illness caused by influenza, primarily during the months of November through March. A new study by researchers from several universities including UC San Diego, published in *ACS Central Science*, suggests a novel approach for combating this sometimes deadly virus, thanks in part to detailed simulations of the virus created using SDSC's *Comet* supercomputer.

Although influenza has been studied for decades, the binding of the virus to host cells and its transmission remain a mystery. Led by Rommie Amaro, a professor of chemistry and biochemistry and director of UC San Diego's National Biomedical Computation Resource, and her colleagues at UC San Diego, UC Irvine, University of Pittsburgh, Point Loma Nazarene University, and the National Institutes of Health (NIH), said the new findings suggest that the proteins may use a 'secret' site to help in the binding and cutting process. Specifically, a closer look at two glycoproteins on the influenza viral surface, neuraminidase (NA) and hemagglutinin (HA), allow for a better understanding of virus.

"Another way to think about these proteins is that they have to work together. One protein (HA) acts like glue to help the virus bind to cells, whereas the other protein (NA) is like a pair of scissors – it helps the virus release itself from cells," explained Amaro. "If the virus sticks too much to the host cell, or if the scissors are too good, the virus itself cannot survive in the body. Understanding the balance between viral binding and release, or the functional balance of the two proteins, is an ongoing area of interest because it's believed to control how well the 'flu can transmit between people, and also how the virus may jump from pig to human, as it did in the 1918 and 2009 pandemic flu events."

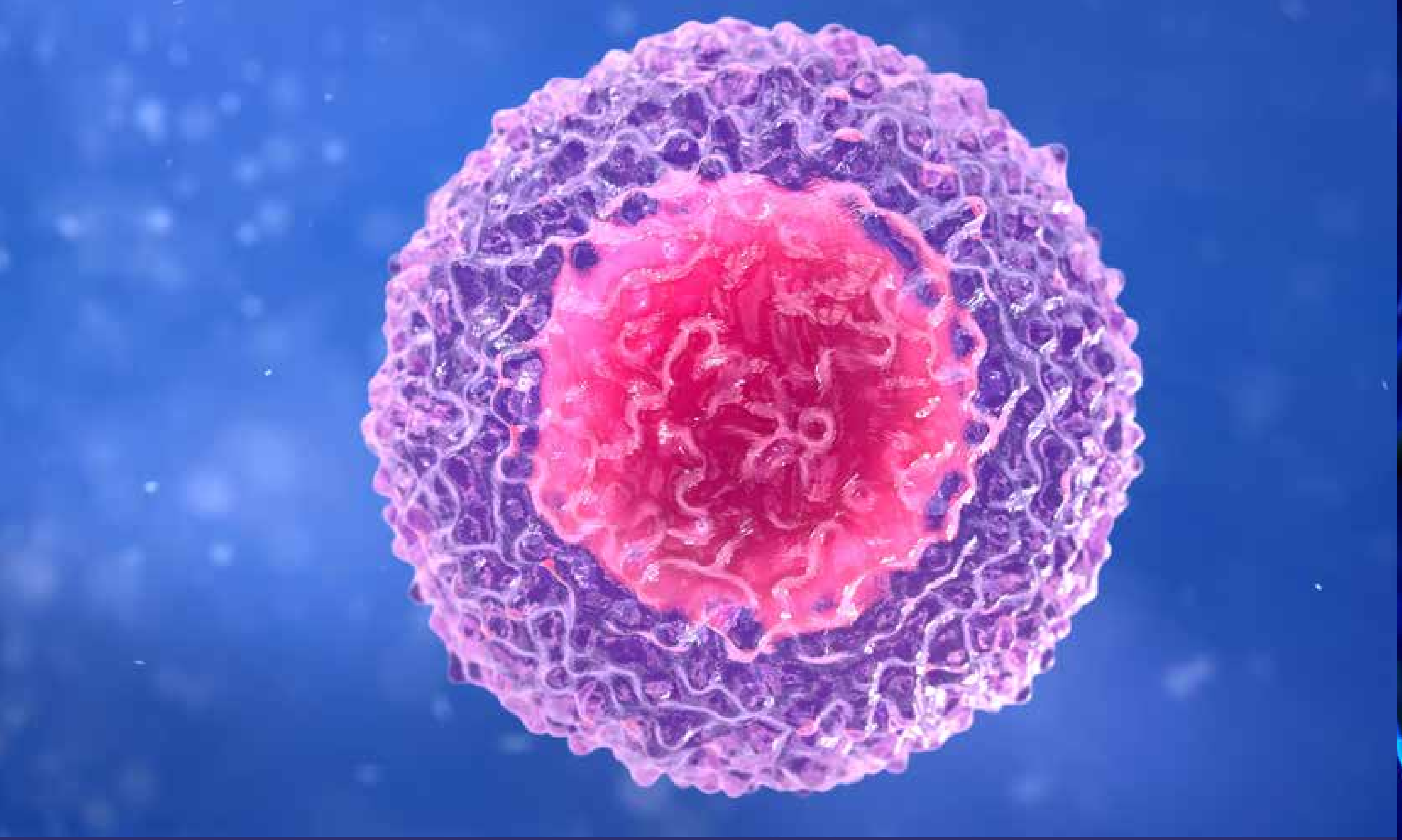
The study suggests a novel pharmaceutical strategy of disrupting the balance between HA and NA activity by targeting the secret site on NA (also called the 'secondary' site) with small-molecule ligands or drugs. While inhibitors of NA cleavage will continue to be important, therapeutics targeting the NA secondary site may provide an alternate, potentially less strain-specific approach for combating new influenza variants.

The simulations using *Comet* provided insight into different structural features of the flu virus and how it interacts with molecules. For example, in some of the most virulent strains of the flu virus, the research team found that the scissor protein (NA) is dramatically shortened, which allows for a better understanding of how the change in height affects the ability of the virus to bind to small molecules.

(Left) High-resolution models of influenza, developed by integrating X-ray crystallography with electron microscopy, enable the interrogation of the functional balance of the flu's two glycoproteins.

Credit: Jacob Durrant, University of Pittsburgh.

QRGO.PAGE.LINK/FEJBK



Robert Sinkovits is director of scientific computing applications at SDSC.

SDSC, Vanderbilt University Collaborate in Human Vaccines Project Study

For the first time ever, researchers are comprehensively sequencing the human immune system, which is billions of times larger than the human genome. According to a study published in the February 13 online issue of *Nature* by the Human Vaccines Project, researchers have sequenced a key part of this vast and mysterious system: the genes encoding the circulating B cell receptor repertoire.

Led by scientists at Vanderbilt University Medical Center and SDSC, this advancement was possible due to SDSC leveraging its considerable computing power to work with the multiple terabytes of data needed for the study. A central question was whether the shared sequences across individuals were the result of chance rather than some shared common biological or environmental factor.

To address this issue, researchers developed a synthetic B cell receptor repertoire and found that the overlap observed experimentally was significantly greater than what would be expected by chance, according to Robert Sinkovits, director of scientific computing applications at SDSC. “The Human Vaccines Project allows us to study problems at a larger scale than would be normally possible in a single lab and it also brings together groups that might not normally collaborate,” he said.

ORGO.PAGE.LINK/MK5OA

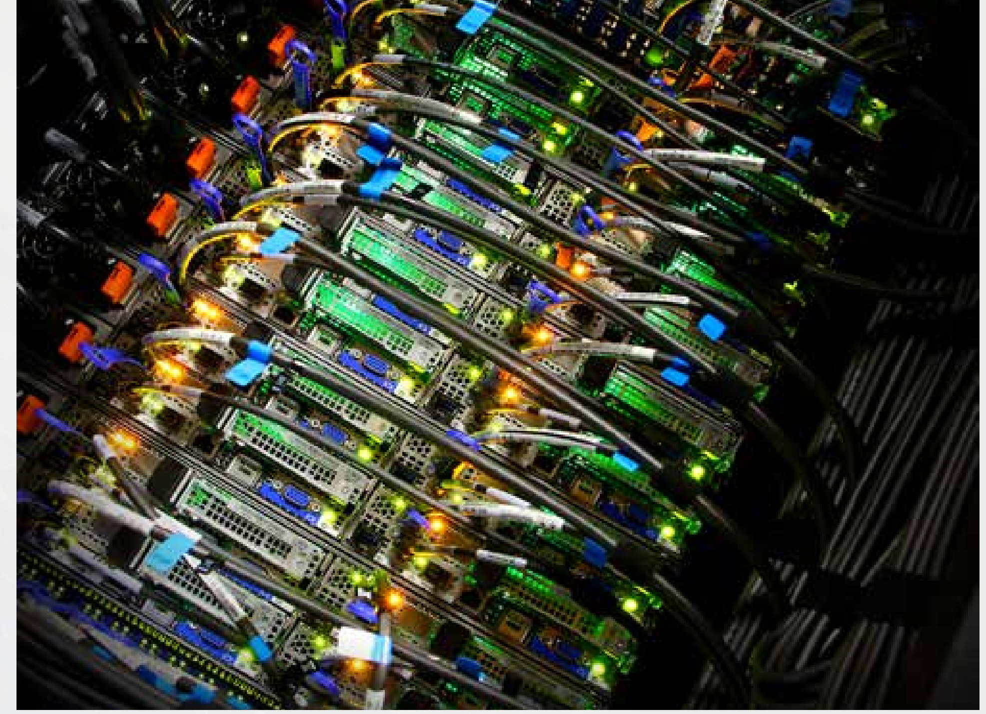
FOCUSED SOLUTIONS & APPLICATIONS

FOR ADVANCED COMPUTATION

SDSC's computational, storage, and networking resources – plus a high level of combined expertise required to configure, operate, and support them – create an advanced cyberinfrastructure that supports scientific discovery among numerous disciplines across academia, industry, and government.

Advanced but user-friendly resources such as SDSC's petascale-level *Comet* supercomputer underscore a vital need for systems that serve a broad range of research, with a focus on researchers who have modest to medium-scale computational needs, which is where the bulk of computational science needs exist.

*The **"long tail" of science** is the idea that the large number of modest-sized computationally based research projects represent, in aggregate, a tremendous amount of research that can yield scientific advances and discovery.*



COMET

A Computing Cyberinfrastructure for the 'Long Tail' of Science

While *Comet* is capable of an overall peak performance of about three petaflops – or three quadrillion calculations per second – its allocation and operational policies are geared toward rapid access, quick turnaround, and an overall focus on scientific productivity. *Comet* quickly established itself as one of the most widely used supercomputers in the National Science Foundation's XSEDE (eXtreme Science and Engineering Discovery Environment) program, which connects researchers to an advanced collection of integrated digital resources and services.

"SDSC's national mission to help pioneer an advanced research cyberinfrastructure has always been our core, and that has enabled us to support collaborations at the local and state levels," said SDSC Director Michael Norman, Principal Investigator (PI) for the *Comet* program, the result of an NSF grant now totaling more than $27 million.

33,000*
Individual users have run jobs on *Comet* through Science Gateways**

900,000,000+*
Core hours of computing provided by *Comet* since it began operations

4,500*
Individual users accessed *Comet* via traditional login from over 350 unique institutions

5,000,000+*
GPU hours of computing provided by *Comet* since it began operations

In mid-2018, the NSF extended *Comet's* service into a sixth year of operation, with the system now slated to run through March 2021. The research community depends on long-term availability and continuity in computing resources, and this $2.4M award extension from NSF ensures continued access to this highly productive and user-friendly system. SDSC also doubled the number of graphic processing units (GPUs) on *Comet* from 144 to 288 in direct response to growing demand for GPU computing across a wide range of research domains. The expansion makes *Comet* the largest provider of GPU resources available to the NSF's XSEDE program.

*As of mid-2019. ** A science gateway is a community-developed set of tools, applications, and data services and collections that are integrated through a web-based portal or suite of applications.

EXPANSE

COMPUTING WITHOUT BOUNDARIES

Coming mid-2020

In June 2019, SDSC was awarded a five-year NSF award valued at $10 million to deploy *Expanse*, a new supercomputer designed to advance research that is increasingly dependent upon heterogeneous and distributed resources. The system will go into production in mid-2020.

"Having access to supercomputers such as *Expanse* and *Comet* has become a necessity for researchers across an ever-growing number of science domains at institutions throughout the United States," said UC San Diego Chancellor Pradeep K. Khosla. "SDSC is nationally recognized for building high-performance computational systems specifically designed for scientific research. The NSF sought a national leader to help the organization deliver on its goal to make such resources available to a broader user base for the purpose of advancing scientific discovery."

"The name of our new system says it all," said SDSC Director Michael Norman, PI for *Expanse*, and a computational astrophysicist. "As a standalone system, *Expanse* builds on the exceptional performance and throughput of *Comet*. But with recent innovations in cloud integration and composable systems, as well as continued support for science gateways and distributed computing via the Open Science Grid, *Expanse* will enable researchers to push the boundaries of high-performance computing to answer questions previously not possible."

The NSF award, which runs from October 1, 2020 to September 30, 2025, is valued at $10 million for acquisition and deployment of *Expanse*. An additional award will be made in the coming months to support *Expanse* operations and user support.

Like SDSC's *Comet* supercomputer, *Expanse* will continue to serve what is referred to as the 'long tail' of science. Virtually every discipline, from multi-messenger astronomy, genomics, and the social sciences, as well as more traditional ones such as earth sciences and biology, depend upon these medium-scale, innovative systems for much of their productive computing.

Projected to have a peak speed of 5 petaflops, *Expanse* will approximately double the performance of *Comet* with next-generation processors and NVIDIA's GPUs. *Expanse* will increase throughput of real-world workloads by a factor of at least 1.3 for both CPU and GPU applications relative to *Comet*, while supporting an even larger and more diverse research community.

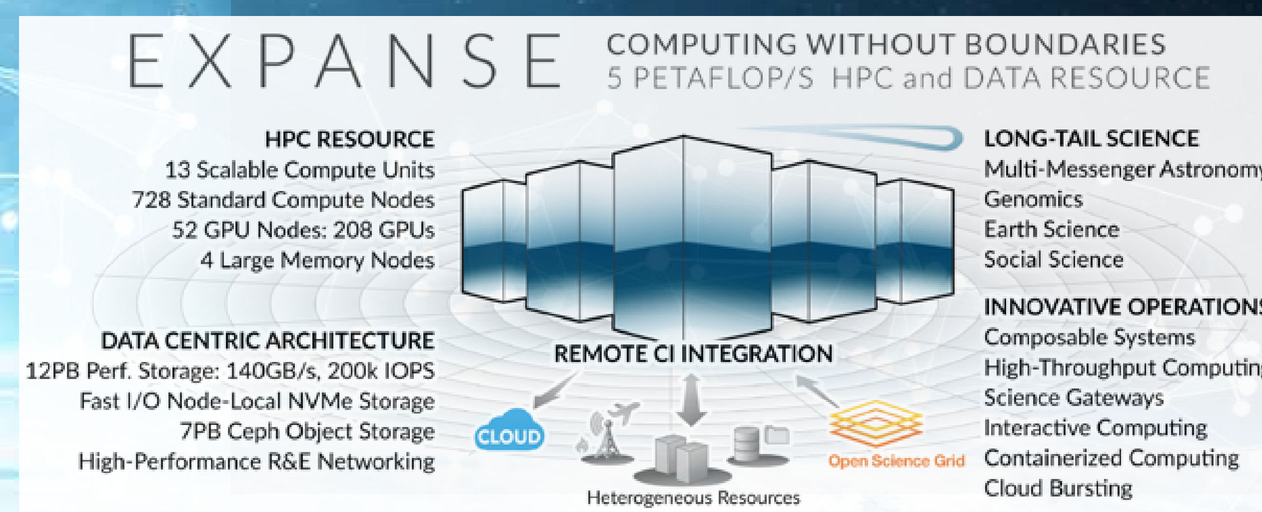
Expanse's accelerated compute nodes will provide a much-needed GPU capability to the user community, serving both well-established applications in areas such as molecular dynamics as well as rapidly growing demand for resources to support machine learning and artificial intelligence. A low-latency interconnect based on Mellanox High Data Rate (HDR) InfiniBand will support a fabric topology optimized for jobs of one to a few thousand cores that require medium-scale parallelism.

Expanse will support the growing diversity in computational and data-intensive workloads with a rich storage environment that includes 12PB of high-performance Lustre, 7PB of object storage, and more than 800TB of NVMe solid state storage.

While *Expanse* will support traditional batch-scheduled HPC applications, breakthrough research is increasingly dependent upon carrying out complex workflows that may include near real-time remote sensor data ingestion and big data analysis and interactive data exploration and visualization, as well as large-scale computation."

"One of the key innovations in *Expanse* is its ability to support so-called composable systems at the continuum of computing with dynamic capabilities," said SDSC Chief Data Science Officer Ilkay Altintas, an *Expanse* co-PI and the director of SDSC's Workflows for Data Science (WorDS) Center of Excellence. "*Expanse* will let researchers compose needed features across its system according to their own needs. Using tools such as Kubernetes, and workflow software we have developed for projects such as the NSF-funded WIFIRE and CHASE-CI programs, *Expanse* will extend the boundaries of what is possible by integration with the broader computational and data ecosystem."

This ecosystem is increasingly including public cloud resources. *Expanse* will feature direct scheduler-integration with the major cloud providers, leveraging high-speed networks to ease data movement to/from the cloud, and opening up new modes of computing made possible by the combination of *Expanse's* powerful HPC capabilities and ubiquity of cloud resources and software.



Triton Shared Computing Cluster (TSCC)

High-Performance Computing for Researchers and Innovators

Operated on behalf of UC San Diego by SDSC, the *Triton Shared Computing Cluster* (*TSCC*) supports research across a wide variety of domains, including biology and life sciences, chemistry, climate, engineering, political and social sciences, and more. Since its launch in 2013, *TSCC* has grown to 34 participating labs/groups with almost 300 researcher-owned compute nodes, plus an additional 75 common nodes available to anyone on campus through a pay-as-you-go recharge model. The latter is popular with individual researchers with occasional or temporary computing needs, students, and classes needing access to high-performance computing. In mid-2019 the National Science Foundation (NSF) awarded a two-year grant worth almost $400,000 to SDSC to deploy a new system called Triton Stratus as part of *TSCC*, which primarily serves UC San Diego researchers and SDSC's industry partners. Triton Stratus will provide researchers with improved facilities for accessing emerging computing tools such as Jupyter notebooks and RStudio to implement computational and data analysis functions and workflows and scale them to commercial cloud computing resources. SDSC Director of Industry Relations Ron Hawkins is the Principal Investigator for Triton Stratus, joined by SDSC's Robert Sinkovits, Subhashini Sivagnanam, and Mary Thomas as co-PIs.

34
Participating Labs/Groups

500+
Compute Nodes

7,000+
Compute Cores (300+ Nodes)

250+
Teraflops and Growing!

TSCC
GOO.GL/WWAQL5

Triton Stratus Award
ORGO.PAGE.LINK/SruXT

SDSC Co-location (Colo) Facility

SDSC offers colocation services to UC San Diego, the UC system, and the local research community. SDSC's 19,000 square-foot climate-controlled, secure data center is equipped with 13 megawatts of power, multi-10, 40, and 100 gigabit network connectivity, and a 24/7 operations staff. Within that center is a "co-location" facility that is free to UC San Diego researchers via a program aimed at saving campus funds by housing equipment in an energy-efficient, shared facility. The Colo facility houses computing and networking equipment for dozens of campus departments, every division and school, as well as local partners that include Rady Children's Hospital, the J. Craig Venter Institute, Simons Foundation, The Scripps Research Institute, and the Sanford-Burnham Medical Research Institute.

SDSC's Colo facility has resulted in more than $2.1 million annual energy savings and about $1.5 million in avoided construction costs in the latest fiscal year, while streamlining and improving the management of hundreds of campus systems. The facility is well-suited to installations that need to demonstrate regulatory compliance, as well those that require high-speed networking. SDSC welcomes inquiries from local companies interested in co-locating equipment to facilitate collaborations with UC San Diego and SDSC investigators.

Storage, Networking & Connectivity

SDSC's Research Data Services team administers a large-scale storage and compute cloud. UC San Diego campus users, members of the UC community, and UC affiliates are eligible to join the hundreds of users who already benefit from the multi-petabyte, OpenStack Swift object store. SDSC Cloud boasts a simplified recharge plan that eliminates fees such as bandwidth and egress charges. SDSC Cloud also includes an elastic compute facility, based on OpenStack Nova, using Ceph for storage. This comprehensive cloud environment provides researchers with a testbed and development environment for developing cloud-based services, and for many data science workflows. It is especially attuned to data sharing, data management, and data analytics services. SDSC Cloud is one of the platforms that underpins the National Data Service's suite of offerings (see page 8 for more on NDS). UC San Diego researchers who make use of SDSC Cloud using sponsored research funds will no longer incur overheard charges (IDC).

Data Oasis

SDSC's *Data Oasis* is a Lustre-based parallel file storage system that provides high performance I/O to *Comet* and *TSCC*. A critical component of SDSC's Big Data initiatives, *Data Oasis* features 12 petabytes (PB) of capacity and 200 gigabytes (GB) per second of bandwidth. *Data Oasis* ranks among the fastest parallel file systems in the academic community. Its sustained speeds mean researchers could retrieve or store 240 terabytes (TB) of data—the equivalent of *Comet's* entire DRAM memory—in about 20 minutes, significantly reducing time needed for retrieving, analyzing, storing, or sharing extremely large datasets. Recent enhancements to Lustre, such as Data on Metadata, improvements to the *Data Oasis* storage networking, and the presence of flash memory on HPC system flash memory on SDSC's HPC systems, combine to give users a flexible, high-performance storage environment capable of tackling the most demanding simulation and data analysis problems.

SDSC Universal Scale Storage

In March 2018, SDSC launched a scale-out storage service to UC San Diego, the UC system, and the local research community. To meet the needs of researchers, SDSC built the Universal Scale Storage service with Qumulo, a modern approach to scale-out storage that delivers fast and flexible storage with the real-time data usage and performance analytics necessary for visibility at the petabyte scale, all while containing costs. This service creates a single namespace that can be used in research labs or on SDSC's supercomputers and clusters. Universal Scale Storage already provides multiple petabytes of space to its current users and continues to rapidly expand to meet demand.

FOCUSED SOLUTIONS& APPLICATIONS

FOR LIFE SCIENCES COMPUTING

Life sciences computing is a key part of SDSC's strategic plan, with the goal of improving the performance of bioinformatics applications and related analyses on advanced computing systems. The initial work, co-sponsored and supported by Dell and Intel, involved bench-marking selected genomic and Cryo-electron Microscopy (Cryo-EM) analysis pipelines. SDSC's initiative focuses on developing and applying rigorous approaches to assessing and characterizing computational methods and pipelines.

CIPRES Gateway Wins Cloud Bursting Grant

One of the most popular science gateways across the entire XSEDE resource portfolio is the CIPRES science gateway, created as a portal under the NSF-funded CyberInfrastructure for Phylogenetic RESearch (CIPRES) project in late 2009. Based at SDSC, CIPRES lets scientists explore evolutionary relationships by comparing DNA sequence information between species using supercomputers allocated by the NSF's XSEDE project.

In early 2019 CIPRES was awarded a one-year Internet2 grant funded by the NSF that gives users free access to more powerful compute processors available from commercial cloud provider Amazon Web Services to accelerate their scientific discoveries. Internet2 leads a project called Exploring Clouds for Acceleration of Science (E-CAS) in partnership with representative commercial cloud providers. The project is funded under a collaboration with the NSF to support academic-based researchers who can benefit from access to commercial clouds. Access to AWS will be free for researchers using CIPRES during the period of the award.

"This award allows us to take the first step in cloud bursting: to let users access the latest graphics processing units (GPUs) at AWS," said SDSC Bioinformatics Researcher Mark Miller, principal investigator for CIPRES. "GPUs can speed up some CIPRES analyses by tens of times compared to traditional CPU cores. The availability of more GPUs will also let researchers ask much harder questions, while increasing research throughput and productivity."

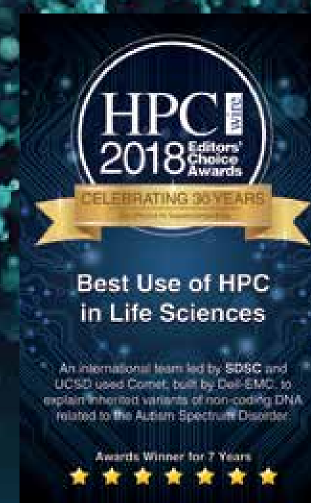
QRGO.PAGE.LINK/Z3TPR

Open EEGLAB Advances UC San Diego's Neuroscience Gateway Project

Even though electroencephalography (EEG) has been used for almost 100 years, this safe and painless test of brain activity remains an efficient method for recording aspects of rapid brain activity patterns supporting our thoughts and actions. Leveraging the power of SDSC's *Comet* supercomputer, UC San Diego researchers have demonstrated they can efficiently analyze more than 1,000 EEG 128-channel high-density data sets via the new Open EEGLAB Portal running on SDSC's Neuroscience Gateway (NSG).

Developed by Arnaud Delorme, Ramon Martinez, and Scott Makeig of UC San Diego's Swartz Center for Computational Neuroscience (SCCN) in collaboration with SDSC researchers Amitava Majumdar, Subhashini Sivagnanam, and Kenneth Yoshimoto, a first report on the Open EEGLAB portal was presented at the March 2019 International IEEE EMBS Conference on Neural Engineering in San Francisco. The Open EEGLAB Portal project is supported by the National Institutes of Health.

QRGO.PAGE.LINK/38AD



SDSC Awarded HPCwire Editors' Choice for Best Use of HPC in the Life Sciences

In late 2018 SDSC received this top award from HPCwire, recognizing *Comet's* role in new findings related to autism spectrum disorder (ASD) published in Science by researchers at UC San Diego's School of Medicine. "Life sciences is one of our three key strategic priorities at SDSC, so we are gratified for this continuing recognition by HPCwire," said SDSC Director Michael Norman in accepting the award at the 2018 International Conference for High Performance Computing, Networking, Storage and Analysis (SC18). "Awards such as these underscore the significant role that high-performance systems are playing in helping to advance discovery in many areas of life sciences."

QRGO.PAGE.LINK/TYIRA

FOCUSED SOLUTIONS& APPLICATIONS

FOR DATA-DRIVEN PLATFORMS AND APPLICATIONS

SDSC's mission has grown to include more than advanced computation as researchers require innovative applications and expertise related to the ever-increasing amount of digitally-based scientific data. While the Center's strategic plan focuses on three main areas – advanced and versatile computing, data science and engineering, and life sciences computing – SDSC has made significant inroads in the area of cloud-based computing, artificial intelligence and machine learning.

"SDSC has long been at the national forefront of large-scale scientific computing, which is foundational to the emergence of data-enabled scientific research," said SDSC Director Norman. "Our newest supercomputer, called *Expanse* and scheduled to go into service in mid-2020, includes cloud integration that also gives users access to new GPUs and processors as they become available from the public cloud providers. This is a great example of how our approach to computing without boundaries gives the research community capabilities that they just can't get with a static HPC system."

Data Science @ SDSC

A key goal of the Data Science @ SDSC (DS@SDSC) program is to foster partnerships across academia, industry, and government while serving as a hub of connectivity and collaboration for projects requiring high-level data analytics, 'big data' management, and distributed computing expertise, starting with cross-campus collaborations. "In order to effectively respond to today's data-rich research priorities, there must be a fertile ground of communication and collaboration between like-minded researchers to solve problems with high impact in science, society, and education," said SDSC Chief Data Science Officer and DS@SDSC Director Ilkay Altintas. "We are also connecting data science research initiatives with the local and very active data science community in San Diego." Over the last year, DS@SDSC focused on solving problems using artificial intelligence for science and created a reference architecture that can be customized for diverse applications involving scalable data-driven solutions.

With massive amounts of scientific data being generated every day, the need to extract meaningful value from all this information has become an essential task across the academic and industry domains.

"While SDSC is a nationally recognized center for advancing computational science and 'big data' expertise, DS@SDSC's role is to serve as a connectivity and collaboration hub that brings together the best and brightest minds across all of academia, industry, and government, starting right here on the UC San Diego campus," said SDSC Director Michael Norman.

Another focus area of DS@SDSC is education and outreach efforts, including the development of professional training initiatives to help establish a modern and fully capable data science workforce, as well as assistance in guiding career paths for data science researchers. DS@SDSC affiliates include principally Ph.D.-level researchers experienced in a wide range of areas needed to develop and deliver a robust curriculum for managing data-intensive scientific research. Examples of expertise areas include:

- Data modeling and integration
- Machine learning, including methods for artificial intelligence and graph analytics
- Workflow and collaboration management
- Performance modeling for 'big data' platforms and workloads
- Scalable, high-performance analytics
- Scientific visualization

DS@SDSC is working in concert with the Halicioğlu Data Science Institute (HDSI), which was launched in 2018. Research labs and offices for HDSI's senior staff and faculty are currently located in SDSC's East Building.

DS@SDSC Director, SDSC Chief Data Science Officer, and HDSI Fellow Ilkay Altintas gives a presentation at SDSC's Industry Partners Program Tech Forum.

DATASCIENCE.SDSC.EDU

'AWESOME' Social Media Data Platform

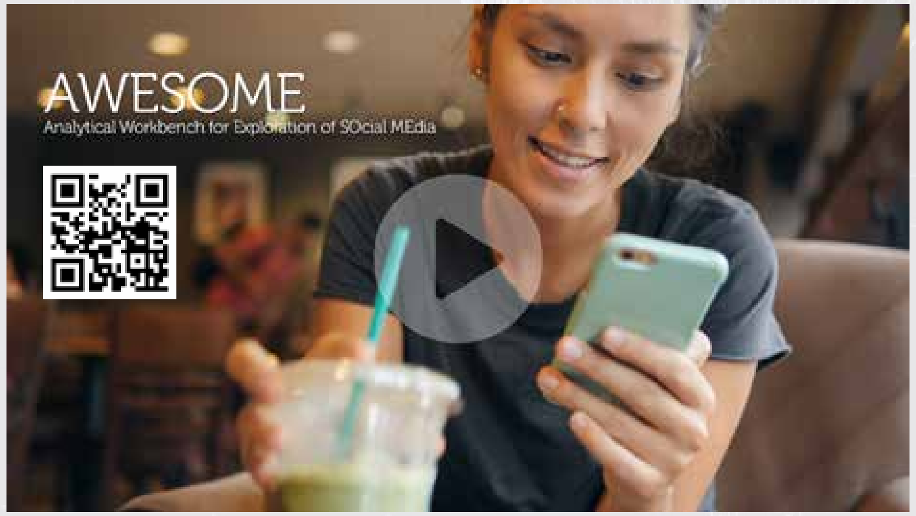
Amarnath Gupta, leader of the AWESOME project, is director of the Advanced Query Processing Lab and a research scientist at SDSC.

SDSC researchers have developed an integrative data analytics platform that harnesses the latest 'big data' technologies to collect, analyze, and understand social media activity, along with current events data and domain knowledge. Called AWESOME (Analytical Workbench for Exploring SOcial Media, the platform has the capability to continuously ingest multiple sources of real-time social media data and scalable analysis of such data for applications in social science, digital epidemiology, and internet behavior analysis. AWESOME is assisting social science researchers, global health professionals, and government analysts by using real-time, multi-lingual, citizen-level social media data and automatically crosslinking it to relevant knowledge to better understand the impact on and reaction to significant social issues.

Funded by the National Science Foundation (NSF) and National Institutes of Health (NIH), AWESOME's goal is to benefit society through areas as diverse as detecting free speech suppression, to shaping policy decisions, or even slowing the spread viruses. "Our tagline is 'heterogeneous analytics for heterogeneous data'," said SDSC researcher Amarnath Gupta, who leads the project.

After being awarded an NIH grant shared by SDSC, UCLA and UC Irvine, AWESOME was used to make timely predictions for the number of high HIV-risk patients at county levels. The award is the result of a collaboration between the groups under the banner of the UCOP-supported UC Institute for Prediction Technology (UCIPT), a multi-campus program to use new innovations in social technologies to predict human behaviors and outcomes.

Awesome uses a Polystore, a system made of several data management systems and multiple databases to achieve different data models. SDSC has been collaboratively extending AsterixDB to enable new political and social science analytics for the 21st Century China Center initiative of UC San Diego, a project that produces scholarly research and informs policy discussions on China and U.S.-China relations.



Learn more about the AWESOME social media data platform by visiting SDSC's YouTube channel.
Scan the QR code above or visit youtu.be/dQk4Hkr5rJY

Open Science Chain

Subhashini Sivagnanam, Principal Investigator for Open Science Chain and a principal scientific computing specialist with SDSC's Data-Enabled Scientific Computing division.

In August 2018, SDSC was awarded a three-year NSF grant to design and develop cyberinfrastructure called Open Science Chain (OSC), which allows researchers to contribute information about scientific data and independently verify the authenticity of scientific datasets. "Data sharing is an essential element of scientific research and associated publications, and facilitating the future reuse of that data in a secure and independently verifiable manner is critical to the advancement of research," said Subhashini Sivagnanam, Principal Investigator for the grant and a principal scientific computing specialist with SDSC's Data-Enabled Scientific Computing division. "Researchers have the ability to extend and build upon prior research when they are able to efficiently access, validate, and verify the data referenced."

The goal of Open Science Chain is to increase the confidence of scientific results and enhance data sharing, which in turn would result in greater research productivity and reproducibility. OSC uses open-source consortium blockchain framework for storing metadata and verification information of the scientific datasets. Information stored in the blockchain ledger is verifiable and immutable, which is essential for reproducibility and audits. As the datasets change or evolve over time, this new information is appended to the blockchain enabling researchers to view a detailed evolution history of that dataset over time.

"Open Science Chain is being designed and implemented using real-world scientific datasets from a diverse set of use cases ensuring broad applicability across scientific domains" said Viswanath Nandigam, co-PI for the project and associate director for SDSC's Advanced Cyberinfrastructure Development Lab. OSC can be used by broad community of users who are in need of a resource that manages an immutable record of their data that can be independently verified.

OPENSCIENCECHAIN.ORG

SeedMeLab

Amit Chourasia, Principal Investigator for SeedMeLab and leader of the SDSC Visualization Group.

SeedMeLab provides users a powerful way to search, manage, share, and visualize data with an innovative content and data management system. It eliminates content fragmentation (data, its context and its discussion) and enables quick reference of data/research context that is especially useful when writing research papers, while accelerating team productivity and knowledge transfer between lab members in a sustained way. A branded site enables the Principal Investigators, Projects, and Science Gateways to extend their brand and overcome data management gaps. SeedMeLab is available as an open-source software and as a managed service from SDSC.

OPENSCIENCECHAIN.ORG

SDSC Centers of Excellence

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

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

WORKFLOWS FOR DATA SCIENCE

Called WorDS for 'Workflows for Data Science', this center of excellence combines more than a decade of experience within SDSC's Scientific Workflow Automation Technologies laboratory, which developed and validated scientific workflows for researchers working in computational science, data science, and engineering. "Our goal with WorDS is to help researchers solve problems. We work on scalable and collaborative methods for teams of researchers to create their own workflows, and in turn to better manage the tremendous amount of data being generated in so many scientific disciplines, while letting them focus on their specific areas of research instead of having to solve workflow issues and other computational challenges," said Ilkay Altintas, director of WorDS and SDSC's Chief Data Science Officer. Funded by a combination of sponsored agreements and recharge services, WorDS' expertise and services include:

- World-class researchers and developers well-versed in data science, big data, and scientific computing technologies,
- Research on workflow management technologies that resulted in the collaborative development of the popular Kepler Scientific Workflow System,
- Development of data science workflow applications through a combination of tools, technologies, and best practices,
- Hands-on consulting on workflow technologies for big data and cloud systems, i.e., MapReduce, Hadoop, Yarn, Spark, and Flink,
- Technology briefings and classes on end-to-end support for data science.

WORDS.SDSC.EDU

CLDS Lead Scientist James Short is also director of the BlockLAB blockchain research laboratory at SDSC.

CENTER FOR LARGE SCALE DATA SYSTEMS

In September 2018, the Center for Large Scale Data Systems (CLDS) formally opened a new blockchain research laboratory with the objectives of exploring the principal technologies and business use cases in blockchains, distributed ledgers, digital transactions, and smart contracts. Founded in partnership with technology firms AEEC, Collibra, Decision Sciences, Dell Technologies, IBM, and Intel, the new laboratory, called BlockLAB, is conducting research in blockchain and distributed ledger technologies (DLT) and their potential business applications across a wide range of industrial and organizational settings. "One of our primary goals is to work closely with industry partners to provide foundational knowledge to help science-based and industrial companies evaluate the potential benefits and risks of applying these new technologies to critical, large-scale transaction and data-intensive business processes," said James Short, BlockLAB's director and SDSC's lead scientist at CLDS. BlockLAB's summary of research progress will be presented at SDSC's annual Data West technology conference in December 2019.

CLDS.INFO/
BLOCKLAB.HTML

Sandeep Chandra is director of SDSC's Sherlock Division.

SHERLOCK

SDSC's newly renamed Sherlock Division provides HIPAA- and FISMA-compliant cloud and data cyberinfrastructure to meet the secure computing and data management requirements throughout academia, government, and industry. In 2019 the division, formerly called Health CI, partnered with Microsoft Azure Cloud Services (Azure) to expand its portfolio of cloud services, continuing to build on its strategy of providing end-to-end compliance by coupling the infrastructure-level compliance offered by public Cloud platforms with the expansive compliance offered by Sherlock. The new service is available immediately, with plans to include additional major cloud platforms in the near future.

Sherlock Cloud and a new offering called Vyloc Cloud are managed services offered by this SDSC Center of Excellence. At the same time, the division announced the launch of Innovation Accelerator Platforms within Sherlock Cloud and Vyloc Cloud. These turn-key solutions provide quick access to on-demand, elastic, fit-for-use data platforms for secure processing, visualization, and storage of a wide range of data. "Not only are these platforms highly secure and innovative – they also provide value-added capability that caters to the still-emerging areas of research and development such as machine learning, artificial intelligence, and data analytics," said Sandeep Chandra, executive director of the Sherlock Division.

SHERLOCK.SDSC.EDU

CAIDA's founder and director, KC Claffy, is a resident research scientist at SDSC whose research interests encompass internet topology, routing, security, economics, future internet architectures, and policy.

CAIDA

Internet Research for Cybersecurity and Sustainability

The Center for Applied Internet Data Analysis (CAIDA) was the first center of excellence at SDSC. Formed in 1997, CAIDA is a commercial, government, and research collaboration aimed at promoting the engineering and maintenance of a robust and scalable global internet infrastructure. CAIDA's founder and director, KC Claffy, is a resident research scientist at SDSC whose research interests span internet topology, routing, security, economics, future internet architectures, and policy.

CAIDA was recently awarded a $4 million, five-year National Science Foundation grant to integrate several of its existing research infrastructure measurement and analysis components into a new Platform for Applied Network Data Analysis, or PANDA. The platform, to include a science gateway component, is in response to feedback from the research community that current modes of collaboration do not scale to the burgeoning interest in the scientific study of the internet.

CAIDA.ORG

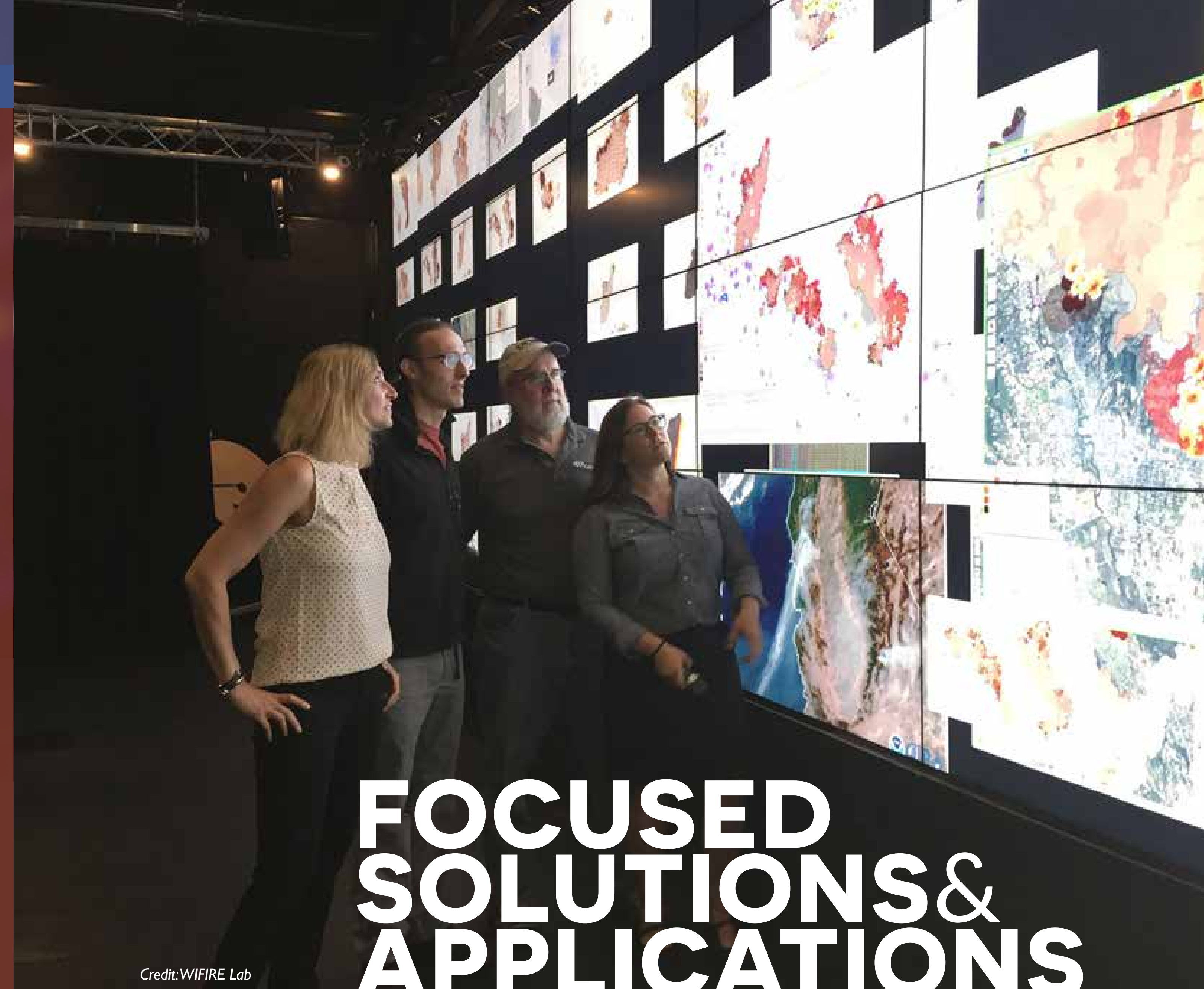
Background image credit: Garrett Britton, InternetSociety

KC Claffy
Inducted Into Internet Hall of Fame

KC Claffy, director of the Center for Applied Data Analysis (CAIDA) based at SDSC, was inducted into the Internet Hall of Fame this year for her pioneering work in the area of internet measurement and analysis. In her acceptance remarks, Claffy, also an adjunct professor in UC San Diego's Computer Science and Engineering Department, recognized the roles that SDSC, its Founding Director Sid Karin, SDSC Research Scientist Hans-Werner Braun, and then-UC San Diego Professor George Polyzos played early in her career. She also observed two historical inflection points: one at the beginning of her career – the commercialization and privatization of the internet infrastructure – and a second now emerging amidst recognition that "society is increasingly exposed to a range of harms serious enough to create a public interest in mitigating them." Claffy also noted that this inflection point toward regulation is "scarier..., because mistakes by governments are more dangerous and take longer to undo than mistakes by the private sector."

One of SDSC's latest YouTube playlists dives into the topic of network neutrality. This controversial topic has been referred to as a "political football" by CAIDA Director KC Claffy, who brings viewers up to speed on several key issues framing the debate, including the death of common carriage, the "dirt road problem", and more.

goo.gl/bSPhyg



Credit: WIFIRE Lab

FOCUSED SOLUTIONS& APPLICATIONS

FOR DATA-DRIVEN DISASTER RELIEF

Earthquakes, wildfires, and severe rainstorms are unfortunately all-too-common occurrences in California as well as other parts of the world. SDSC researchers have teamed up with scientists at UC San Diego and beyond to work with first-responder agencies and government officials to support and enhance a range of public safety programs using a key resource: collecting real-time data and using predictive analytics and other tools to help mitigate the often tragic effects of such events.

Gathering and analyzing such data is possible via internet-connected cyberinfrastructures to detect both ground and weather conditions that could potentially harm people and property in populated and more remote areas. In short, data equals knowledge, and SDSC is at the forefront with a range of advanced technologies to support public safety.

Fire officials use WIFIRE Lab's FireMap tool to track and predict the spread of a rapidly growing wildfire in Southern California. FireMap gathers real-time information about topography, flammable materials, and weather conditions among other variables from giant government data sets and on-the-ground sensors.

Credit: WIFIRE Lab

WIFIRE Lab Expands Partnerships with California Agencies

Started under a multi-year, $2.65 million National Science Foundation (NSF) grant, SDSC's WIFIRE cyberinfrastructure has been attracting high levels of interest among first-responders during the latest fiscal year as a resource to help them better monitor, predict, and mitigate wildfires by analyzing forecasts from the National Weather Service, vegetation readouts from the U.S. Department of Interior, and even satellite data from NASA. WIFIRE Lab now is an all-hazards data resource, enabling analysis and modeling to turn data into knowledge across environmental hazards, as most of the time hazards are connected, such as wildfires and debris flows.

The need for such data-rich resources has increased with the number and intensity of wildfires in recent years. California alone saw nearly 1.7 million acres burned across the state in 2018. "Human beings can probably do anything computers can do, but not at the speed and scale that computers can achieve," said SDSC Chief Data Science Officer and WIFIRE Principal Investigator Ilkay Altintas in a June 2019 *New York Times* interview. "The more data we can use, the more lives and property and businesses we can save."

During the last three years the Los Angeles Fire Department (LAFD) and many other California fire departments has been utilizing a program developed by the WIFIRE Lab called FireMap, which gathers real-time information about topography, flammable materials, and weather conditions among other variables from giant government data sets and on-the-ground sensors.

"In Los Angeles, with our population density, a fire is going to be burning houses down right away," LAFD Fire Chief Ralph Terrazas told the *New York Times*. "There's smoke, there's fire, there's sirens, and we have to make decisions in just minutes."

As an example for the system's potential public impact, through word-of-mouth and social media the Firemap tool was accessed by 800,000 public users over 8 million times to view information related to the devastating wildfires in California throughout the fall of 2017.

WIFIRE Lab works with firefighters throughout California formally, and with more than 130 organizations and individuals that reached out for informal help and collaboration. The Orange County Fire Authority, the Los Angeles County Fire Department, the Ventura County Fire Department, and the State of California Public Utilities Commission are also among those partnering with WIFIRE.

Participants in the WIFIRE Lab include researchers from SDSC, and other UC San Diego areas including the California Institute for Telecommunications and Information Technology's (Calit2) Qualcomm Institute, the Scripps Institution of Oceanography and the Mechanical and Aerospace Engineering (MAE) department at the university's Jacobs School of Engineering.

VOA/TEK gets an inside view of the high-tech world of firefighting as aerial and satellite imagery combine with supercomputing technology to help first-responders evacuate and battle raging wildfires in the Southwestern U.S.

To watch the video, scan the QR code or visit youtu.be/ptnLr-SMe28

Credit: VOA/TEK

FOCUSED SOLUTIONS & APPLICATIONS
For Data-Driven Disaster Relief

HPWREN's First Responder Efforts Continue Expansion beyond Greater San Diego

View from the HPWREN cameras atop Santiago Peak as the Holy Fire approaches in August 2018. *Credit: HPWREN/SDSC*

On August 6, 2018, first responders received a 911 call reporting a fire within the Cleveland National Forest, some 460,000 acres in the counties of San Diego, Riverside, and Orange. A quick check of recently deployed fixed field of view cameras confirmed the presence of smoke, immediately followed by pointing the new ALERTWildfire PTZ cameras on Santiago Peak to confirm the fire's location – and with it a significant potential for major destruction.

That fire became known as the Holy Fire, which burned 23,136 acres and was finally contained more than a month later. The new cameras were deployed less than four months before, in partnership with ALERTWildfire, the Orange County Fire Authority, Southern California Edison, and UC San Diego's High Performance Wireless Research and Education Network (HPWREN).

Co-founded in 2000 by Hans-Werner Braun, a research scientist at SDSC, and Frank Vernon, a research geophysicist with Scripps Institution of Oceanography, and initially funded by the NSF, HPWREN has since transitioned to a user-funded system and has undergone a significant expansion and upgrading to cover hundreds of miles throughout southern California's most rugged terrain. Today, the HPWREN cyberinfrastructure connects even more hard-to-reach areas in remote environments and includes a system of internet-based cameras and weather stations to report local weather and environmental conditions including storms, wildfires, and earthquakes, to local authorities.

"Because we were able to determine post-fire that the smoke produced in the incipient phases of the Holy Fire was visible on these cameras prior to the initial 911 call, we plan on using this technology as 'virtual fire towers' on identified 'high-hazard' days to increase situational awareness," said Brian Norton, Division Chief – Special Operations, Orange County Fire Authority, during an HPWREN Users' Group meeting in April 2019. "This will be done in partnership with volunteer fire watch patrols already established in Orange County."

During that meeting, co-founder Vernon stressed the importance of HPWREN across the science, education, and first responder communities, as well as an array of agencies including San Diego Gas and Electric (SDG&E), San Diego County Fire Authority, California State Parks, Orange County Fire Authority, and WIFIRE.

Vernon also described how the role that UC San Diego's Pacific Research Platform (PRP) plays in supporting HPWREN, and how CENIC's 100 Gigabit optical fiber is used to link HPWREN data servers at UC San Diego, San Diego State University, and UC Irvine. Once a wildfire is spotted by the HPWREN-connected cameras, PRP delivers high-resolution weather data to fire modeling workflows in WIFIRE. (Read more about PRP and its expansion on Page 7.

During the latter half of 2018 and the first half of 2019, numerous upgrades were made to the HPWREN network as well as the more recent ALERTWildfire network, which currently has 25 cameras at 24 locations, including recently installed ones on Santiago Peak, Mount Wilson, and the Santa Barbara Mesa. Other improvements in 2018/2019 included upgraded routers and radios, new routing protocols, fiber paths, and high-capacity data transfer nodes.

HPWREN.UCSD.EDU

Earthquake Rupture Modeling at Scale – In the Cloud

Senior Computational Scientist Yifeng Cui directs SDSC's High-Performance Geocomputing Laboratory.

Researchers from SDSC and industry partners Intel Corporation and Amazon recently ran what is believed to be the first petascale-level earthquake simulation on Amazon's AMS Cloud. That simulation is based on an innovative seismic software package developed in 2017 by SDSC researchers and Intel Corporation that led to the fastest seismic simulation of possible large-scale seismic activity in the southern California region.

The findings were presented in June 2019 at the annual International Supercomputing (ISC) High-Performance Conference in Frankfurt, Germany, with over 3,500 international attendees and 160 exhibitors.

Called EDGE for Extreme-Scale Discontinuous Galerkin Environment, that simulation achieved a world record performance of 10.4 PFLOPS (Peta FLoating-point Operations Per Second, or one quadrillion calculations per second). EDGE is the result of a joint initiative that also includes the Southern California Earthquake Center (SCEC), which is one of the largest open research collaborations in geoscience.

Fast forward about two years, from the first EDGE simulation announced in 2017. Today's top public cloud offerings provide an excellent combination of infrastructure flexibility and scalability that EDGE requires for its own modeling scenarios, so researchers focused on integrating cloud-based technologies in this phase of their research.

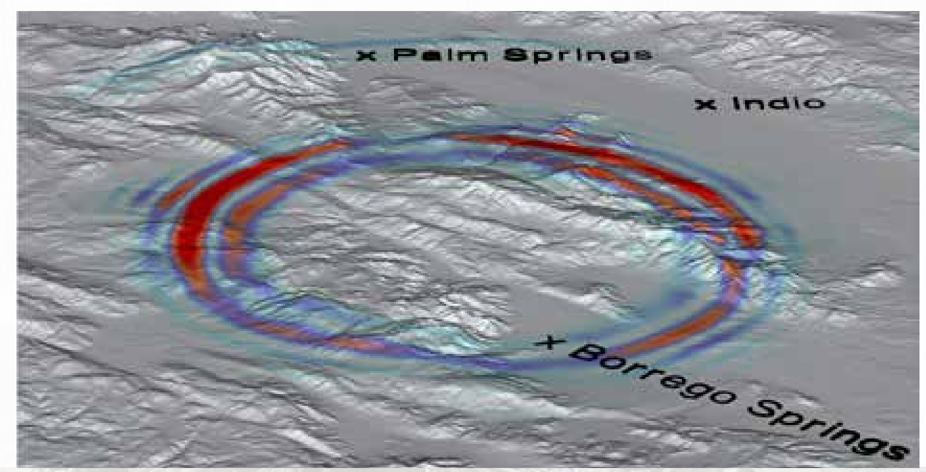
"This work demonstrates the efficient use of application-centric cloud clusters for modern and tightly-coupled scientific computing," said Alex Breuer, lead author of the paper and a postdoctoral researcher with SDSC's HPGeoC group at the time the research was being conducted, "Public cloud services offer elastic multi-petaflops machines, which were four years ago only available through on-premises supercomputing centers."

"This new EDGE project is the first work of its kind that we know of to be conducted at such a large scale in the public cloud," said Yifeng Cui, director of SDSC's High-performance Geo Computing (HPGeoC) Laboratory, principal investigator for SCEC, and co-author of the paper, called 'Petaflop Seismic Simulations in the Public Cloud' that was one of 17 accepted research papers presented at the ISC conference and published in the Springer's Lecture Notes in Computer Science (LNCS) series.

The findings were presented in June 2019 at the International Super Computing (ISC) High-Performance conference in Frankfurt, Germany.

Example of hypothetical seismic wave propagation with mountain topography using the new EDGE software. Shown is the surface of the computational domain covering the San Jacinto fault zone between Anza and Borrego Springs in California. Colors denote the amplitude of the particle velocity, where warmer colors correspond to higher amplitudes.

Credit: Alex Breuer, SDSC.



SDSC INDUSTRY RELATIONS

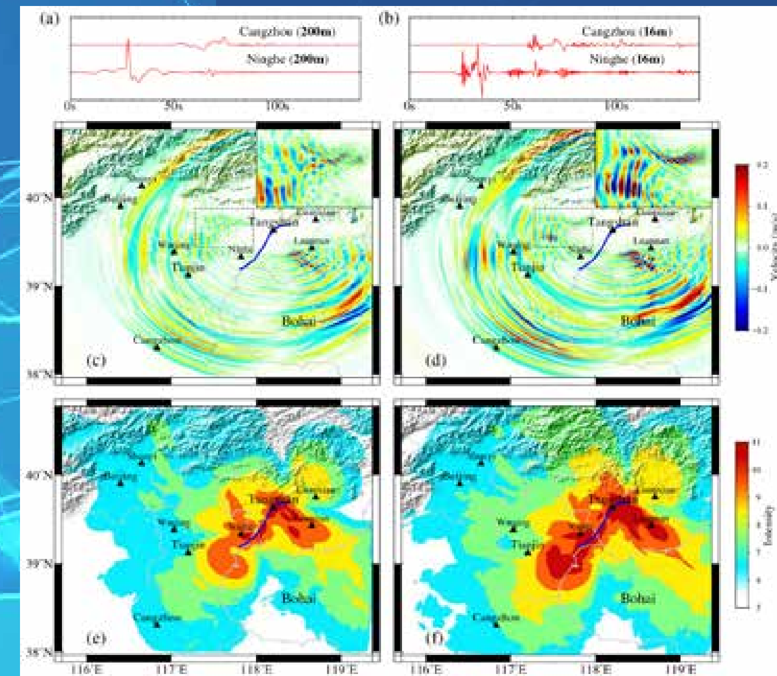
Ron Hawkins is director of Industry Relations for SDSC and manages the Industry Partners Program, which provides member companies with a framework for interacting with SDSC researchers and staff to develop collaborations.

SDSC INDUSTRY PARTNERS PROGRAM
OPTIMIZING FOR IMPACT

A recent MIT study on best practices for industry-university collaborations differentiated between research project outcomes (e.g., providing advice or guidance on use of a technology) and research project impact (e.g., the development of new products or processes). The study concluded that collaborations that result in "tangible impacts" are of much greater value to companies than those that simply provide "promising outcomes."

As a research unit that has traditionally had a strong applied R&D culture, SDSC has embraced the concept of generating impact for its industrial partners and our collaborations during the 2018-2019 reporting period reflected our ongoing focus in this area.

One example is our collaboration with GigaIO, a venture capital-backed startup company developing a high-performance network interconnect fabric based on the industry-standard PCI Express (PCIe) technology. This collaboration was conducted under the auspices of SDSC's Advanced Technology Lab (ATL), which focuses on evaluating emerging hardware and software technologies with industry partners.



Simulated ground motions of the 1976 Tangshan earthquake that resulted in a 12-member team of Chinese researchers being awarded the 2017 Gordon Bell prize. *Image courtesy of Haohuan Fu, Tsinghua University, et al.*

GigaIO is addressing the increasing requirement in high-performance computing (HPC) and artificial intelligence for high-bandwidth, low-latency access to devices such as Graphics Processing Units (GPUs) and Solid State Discs (SSDs) which have traditionally been "captive" units in computer servers. The GigaIO technology permits "dis-aggregating" these devices into "composable" systems for more flexibility, lower cost, and higher performance.

SDSC's ATL worked with GigaIO this year to develop a realistic benchmarking environment for GigaIO's FabreX™ technology, using GigaIO's test lab and SDSC's *Comet* supercomputer. The basis for the benchmarking scenario was the seismic (earthquake) simulation software, AWP-ODC , developed by SDSC Researcher Yifeng Cui and his collaborators. The AWP-ODC code can leverage GPUs and SSDs to improve performance, and therefore was an excellent candidate for a realistic application of GigaIO's technology.

Furthermore, a key aspect of earthquake simulations is visualization of the simulation results for the benefit of both scientists and lay audiences in understanding the reach and destructive impact of an earthquake. Instead of post-processing the simulation results to create a visualization after the simulation finishes, SDSC proposed to implement in situ visualization, which is a technique to display the simulation results in near real-time during a long running simulation. In situ visualization can be a valuable technique for providing quicker access to simulation results, especially for very long running simulations. Under the leadership of SDSC postdoctoral researcher Dawei Mu, a software framework was assembled to perform in situ visualization while running the AWP-ODC code.

This software framework was then deployed on a test environment including servers, GPUs, and SSDs interconnected with the FabreX™ network. A series of simulation runs showed not only the viability of implementing in situ visualization for AWP-ODC but also the favorable performance of the GigaIO-based platform compared to a more conventional GPU server environment such as that found on SDSC's *Comet* supercomputer.

The results of this collaboration generated impact for our partner, GigaIO, in terms of exposing potential features and performance improvements that could be directly incorporated into GigaIO's product. On the flip side, this collaboration generated impact for science by demonstrating a new feature for earthquake simulations (in situ visualization) as well as validating an emerging technology (the FabreX network) for high-performance scientific simulations. The work on in situ visualization done by the SDSC-GigaIO team was subsequently selected for presentation at the 2019 PEARC HPC conference held in Chicago, Illinois, in July 2019.

The above is only one example of impact-producing industrial collaborations during this reporting period. Other collaborations during the year encompassed cybersecurity, high-performance file system technologies for HPC and AI, new pharmaceutical development, and smart manufacturing. Looking forward, the SDSC team anticipates another year of optimizing for impact through the Industry Partners Program.

INDUSTRY.SDSC.EDU

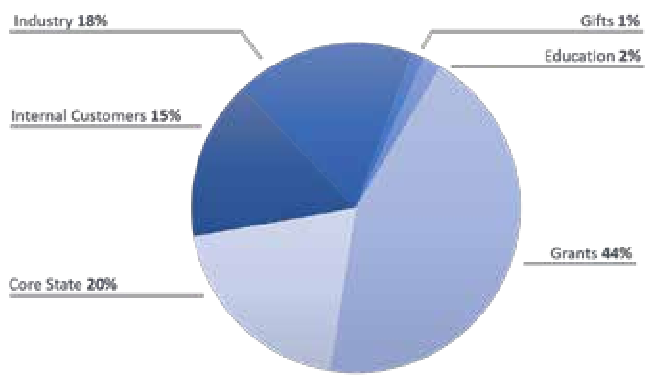
By the Numbers: FY2018/19

Proposal Success Rate

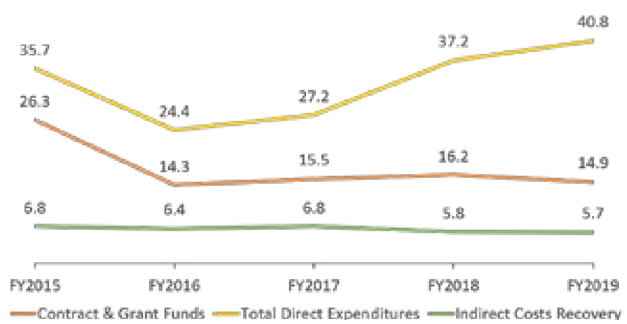
	FY15	FY16	FY17	FY18	FY19
Proposals Submitted	91	73	84	75	67
Proposals Funded	38	33	33	26	31
Success Rate	42%	45%	39%	35%	46%

Number of Sponsored Research Awards

Revenue Categories: $34 Million



Total Expenditures ($ Millions)



In perhaps the most competitive landscape for federal funding in the last two decades, SDSC's overall success rate on federal proposals averages 41% over the last five years compared to the FY2019 national average of about 21% for Computer and Information Science and Engineering proposals at the National Science Foundation (NSF).

During its 34-year history, SDSC revenues have well exceeded $1 billion, a level of sustained funding matched by few academic research units in the country. As of the start of FY2019, SDSC had 112 active awards totaling about $140 million.

Leadership

EXECUTIVE TEAM

Michael Norman

SDSC Director

Shawn Strande

Deputy Director

Ilkay Altintas

Chief Data Science Officer

Chaitanya Baru

(on assignment to NSF)

Associate Director, Data Science and Engineering

Sandeep Chandra

Division Director, Sherlock

(formerly Health Cyberinfrastructure)

Ronald Hawkins

Director, Industry Relations

Christine Kirkpatrick

Division Director, Research Data Services

Samuel "Fritz" Leader

Chief Administrative Officer

Amit Majumdar

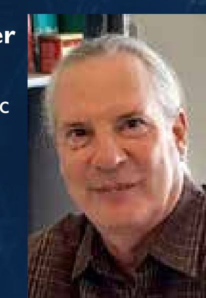
Division Director, Data-Enabled Scientific Computing

Frank Würthwein

Lead, Distributed High-Throughput Computing

Michael Zentner

Division Director, Sustainable Scientific Software



Jan Zverina

Division Director, External Relations

RESEARCH EXPERTS
SDSC Computational & Data Scientists

Ilkay Altintas, Ph.D
Chief Data Science Officer, SDSC
Director, Workflows for Data Science (WorDS) Center of Excellence
Lecturer, Computer Science and Engineering, UC San Diego
Scientific workflows, Kepler Scientific Workflow System
Big data applications, distributed computing
Reproducible science

Michael Baitaluk, Ph.D.
Research Scientist, SDSC
Scientific data modeling and information integration
Systems and molecular biology, bioinformatics

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

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

Laura Carrington, Ph.D.
Director, Performance, Modeling, and Characterization Lab, SDSC
Application performance analysis & optimization,
HPC benchmarking, workload analysis, modeling via machine learning
HPC power modeling and analysis, Chemical engineering

Sandeep Chandra
Executive Director, Sherlock Cloud
Director, Health Cyberinfrastructure Division, SDSC
Compliance (NIST, FISMA, HIPAA), scientific data management
Cloud computing, systems architecture & infrastructure management

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

Amit Chourasia
Senior Visualization Scientist, SDSC
Lead, Visualization Group
Principal Investigator, SeedMeLab
Visualization, computer graphics, data management and sharing

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

Melissa Cragin, Ph.D.
Chief Strategist for Data Initiatives, Research Data Services (RDS), SDSC
Shared data infrastructure, research policy, scholarly communication

Daniel Crawl, Ph.D.
Associate Director, Workflows for Data Science (WorDS) Center of Excellence
Technical Lead, WIFIRE Lab
Spatial and temporal data integration/analysis, interdisciplinary applications

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

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

Diego Davila
Research Scientist, Distributed High-Throughput Computing, SDSC
High-Throughput Computing
Globally distributed compute and data federations

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

Jose M. Duarte, Ph.D.
Scientific Team Lead, RCSB Protein Data Bank
Bioinformatics, structural biology, computational biology

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

Madhusudan Gujral, Ph.D.
Bioinformatics & Genomics Lead, SDSC
Processing & analysis of genomics data, Biomarkers discovery
Genomics software tools development, Data management

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

Edgar Fajardo Hernando
Research Scientist, Distributed High-Throughput Computing, SDSC
High-Throughput Computing, globally distributed compute and data federations

Amit Majumdar, Ph.D.
Division Director, Data-Enabled Scientific Computing, SDSC
Associate Professor, Dept. of Radiation Medicine and Applied Sciences, UC San Diego
Parallel algorithm development
Parallel/scientific code, parallel I/O analysis and optimization
Science gateways, Computational & data cyberinfrastructure software/services

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

Dmitry Mishin, Ph.D.
Research Programmer Analyst, Data-Enabled Scientific Computing, SDSC
Research Programmer Analyst, Calit2
HPC systems, virtual and cloud computing, Kubernetes and containers
Data storage, access, and visualization

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

Viswanath Nandigam
Associate Director, Advanced Cyberinfrastructure Development Lab, SDSC
Data distribution platforms, Scientific data management
Science gateways, Distributed ledger technologies

Mai H. Nguyen, Ph.D.
Lead for Data Analytics, SDSC
Machine learning, Big data analytics, Interdisciplinary applications

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

Francesco Paesani, Ph.D.
Lead, Laboratory for Theoretical and Computational Chemistry, UC San Diego
Theoretical chemistry, Computational chemistry, Physical chemistry

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

Peter Rose, Ph.D.
Director, Structural Bioinformatics Laboratory, SDSC
Lead, Bioinformatics and Biomedical Applications, Data Science Hub, SDSC
Structural biology/bioinformatics, Big data applications
Data mining and machine learning, Reproducible science

Joan Segura, Ph.D.
Scientific Software Developer, RCSB Protein Data Bank
Structural bioinformatics

Igor Sfiligoi
Senior Research Scientist, Distributed High-Throughput Computing, SDSC
Lead Scientific Software Developer and Researcher, SDSC, UC San Diego
High-Throughput Computing
Globally distributed compute and data federations
Cybersecurity in globally distributed systems

James Short, Ph.D.
Lead Scientist, SDSC
Co-Director, Center for Large-scale Data Systems Research (CLDS), SDSC
Director, BlockLAB, SDSC
Market adoption, business application and economics of blockchain
Valuation of data and information markets
Textual analysis of data (Natural Language Processing & Machine Learning)
Corporate data policy and regulatory / legal frameworks

Robert Sinkovits, Ph.D.
Director, Scientific Computing Applications, SDSC
Director, SDSC Training
Co-Director for Extended Collaborative Support, XSEDE
High-performance computing, Software optimization and parallelization
Structural biology, bioinformatics, Immunology

Subhashini Sivagnanam
Senior Computational and Data Science Specialist,
Data-Enabled Scientific Computing Division, SDSC
HPC solutions and applications, Science gateways, Data reproducibility
Computational and data cyberinfrastructure software/services

Shava Smallen
Manager, Cloud and Cluster Development
Principal Investigator, Pacific Rim Application and Grid Middleware Assembly (PRAGMA)
Deputy Manager, XSEDE Requirements Analysis and Capability Delivery (RACD) group
Cyberinfrastructure monitoring and testing
Cloud infrastructure tools, Cluster development tools

Britton Smith, Ph.D.
Assistant Research Scientist, SDSC
Computational astrophysics
Software development

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, UC San Diego
Computational drug design
Personalized cancer medicine
Gene networks analysis
Molecular modeling/molecular dynamics
Neuroscience

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

Nancy Wilkins-Diehr (outgoing)
Associate Director, SDSC
Co-Principal Director, XSEDE at SDSC
Co-Director for Extended Collaborative Support, XSEDE
Principal Investigator, Science Gateways Community Institute
Science gateways
User services
Aerospace engineering

Frank Würthwein, Ph.D.
Lead, Distributed High-Throughput Computing, SDSC
Executive Director, Open Science Grid
Professor of Physics, UC San Diego
High-capacity data cyberinfrastructure
High-energy particle physics

Ilya Zaslavsky, Ph.D.
Director, Spatial Information Systems Laboratory, SDSC
Spatial and temporal data integration/analysis
Geographic information systems
Visual analytics
Geosciences

Michael Zentner, Ph.D.
Director, Sustainable Scientific Software, SDSC
Director, HUBzero Platform for Science and Engineering
Co-Principal Investigator, Network for Computational Nanotechnology
Principal Investigator, Science Gateways Community Institute (SGCI)
Sustainable scientific software
Chemical engineering

Andrea Zonca, Ph.D.
HPC Applications Specialist
Data-intensive computing
Computational astrophysics
Distributed computing with Python
Jupyterhub deployment at scale

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
youtube.com/SanDiegoSupercomputerCenter