

SDSC ANNUAL REPORT FY2020/21

TRANSLATING INNOVATION INTO PRACTICE



SDSC

SAN DIEGO SUPERCOMPUTER CENTER

UNIVERSITY OF CALIFORNIA SAN DIEGO

SAN DIEGO SUPERCOMPUTER CENTER

The San Diego Supercomputer Center (SDSC) is a leader and pioneer in high-performance and data-intensive computing, providing cyberinfrastructure resources, services and expertise to the national research community, academia and industry. Located on the UC San Diego campus, SDSC supports hundreds of multidisciplinary programs spanning a wide variety of domains, from astrophysics and earth sciences to disease research and drug discovery. SDSC's newest National Science Foundation-funded supercomputer, Expanse, supports SDSC's theme of "Computing without Boundaries" with a data-centric architecture, public cloud integration and state-of-the-art GPUs for incorporating experimental facilities and edge computing.

SDSC INFORMATION

Frank Würthwein, 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

Cynthia Dillon
Division Director, External Relations
cdillon@ucsd.edu

TRANSLATING INNOVATION INTO PRACTICE

SDSC Annual Report FY2020/21

(PDF version available online at www.sdsc.edu/pub/)

EDITOR:
Cynthia Dillon

CO-EDITOR:
Kimberly Mann Bruch

CONTRIBUTORS:
Kim Bruch, Cynthia Dillon, Jake Drake, Julie Gallardo, Ron Hawkins, Fritz Leader, Susan Rathbun, Bob Sinkovits, Shawn Strande, Ben Tolo, Nicole Wolter

CREATIVE DIRECTOR:
Ben Tolo

PHOTOGRAPHER:
Owen Stanley

All financial information is for the fiscal year that ended June 30, 2021. Any opinions, conclusions, or recommendations in this publication are those of the author(s) and do not necessarily reflect the views of NSF, other funding organizations, SDSC, or UC San Diego. All brand names and product names are trademarks or registered trademarks of their respective holders.

© 2022 The Regents of the University of California



2 **Director's Letter**
Translating Innovation into Practice



4 **Pi Person of the Year**
Subhashini Sivagnanam

5 **By the Numbers**

52 **Leadership and Research Experts**



WE INFLUENCE

- 6** Taking AI to New Places
- 8** Computing Without Boundaries
- 10** Accelerating Science with Bold National Research Platform
- 12** SDSC Takes a Seat at the Table of Federal Task Force for Advancing Artificial Intelligence
- 14** Comet Blazes a New Trail



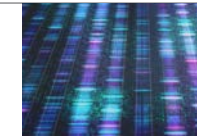
WE PROBLEM SOLVE

- 16** Working to Fight Fires Proactively
- 18** Software Architect's Demo Captures the Attention of CERN
- 20** A Quantum Leap for Molecular Simulations on GPUs
- 22** 'Teaming Up' with San Diego Zoo Wildlife Alliance & ASU to Create Spatial Ecology Gateway
- 24** CAIDA Works with International Team to Inform, Improve Internet Security



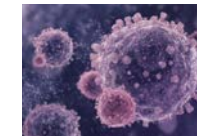
WE STEWARD DATA

- 26** Research Data Services in Focus
- 28** Programmatic Approach to Increasing Revenue Potential of Software Projects
- 30** North, South, East or West, It's Data Collaboration
- 32** Sherlock Continues to Find Solutions for Customers



WE PARTNER

- 34** On the Frontier of Genomics Research
- 36** SDSC, Habana, Supermicro Team Up to Operate Voyager
- 38** Core Scientific Plexus™ Software Extends Reach of SDSC's Expanse Supercomputer
- 39** CloudBank Encompasses Both University and Industry Partnerships
- 40** SDSC Plays Role in NSF's New ICICLE Institute



WE RESEARCH

- 42** New Pathways for Potential RNA Virus Treatment
- 44** Comet Aids in Sickle Cell Research Discovery
- 46** Enabling 3-D Models of Fresh Water Plastic Pollution
- 48** Supercomputer Simulations Show Potential for Carbon-Neutral Fuels



WE EDUCATE

- 50** Education & Outreach at SDSC



“SDSC adopts innovations from industry and academia in the areas of software, hardware, computational & data sciences, and related areas, and translates them into cyberinfrastructure that solves practical problems across any and all scientific domains and societal endeavors.”

– Frank Würthwein

Over the last few months, I've talked about SDSC with literally hundreds of researchers, staff and administrators at UC San Diego, in addition to members of the SDSC community. Invariably, people expected me to articulate the core strength of SDSC in a few words. I have converged on “Translating Innovation into Practice” as my favorite articulation.

SDSC adopts innovations from industry and academia in the areas of software, hardware, computational & data sciences, and related areas, and translates them into cyberinfrastructure that solves practical problems across any and all scientific domains and societal endeavors.

One may argue that this has been SDSC's core competency since inception nearly 40 years ago, and is likely to continue for decades to come. To be a global leader in this endeavor, SDSC is both a services and a research organization. Our research outcomes impact the services we operate, leading to service and infrastructure innovations.

We are seeing a trend towards diversification in hardware architectures and devices. Programmable computational capabilities appear in devices of all kinds, leading to programmable networks, data flow processing and computational storage. These concepts are ready to be integrated into future HPC systems. Simultaneously, there is an explosion in the diversity of compute architectures driven by the market for accelerating artificial intelligence and machine learning (AI/ML).

At the same time, increasing volumes and diversity of data is begging to be turned into insight and solutions, both scientific and societal. Data life cycles need to be managed from acquisition to curation including support for data integration, mining and visualization.

At this point, SDSC is exceptionally well-positioned for growth exploiting these trends. Expanse, our HPC workhorse keeps growing. We added an “industry rack” in 2021, and are about to add two “high-throughput computing” racks in Spring 2022 for a total capacity in excess of 110,000 X86 cores, and hundreds of NVIDIA GPUs, both V100 and A100. Voyager, our new AI/ML optimized HPC system was installed late in 2021. We expect Voyager to support petabyte scale AI/ML. We are expanding the networking infrastructure at SDSC in 2022, and its connectivity to the Wide Area Network in 2023 from 100Gbit/sec to Terabit/sec capacity. A Terabit/sec network allows Petabyte data movements over an extended lunchtime break. While Voyager is already providing Terabit/sec capacity between its nodes, the Prototype National Research Platform (PNRP) to be deployed in 2022, will be able to sink and source data at Terabit/sec across the UC San Diego campus and the nation. Expanse, Voyager and PNRP are all five-year projects with a potential for five-year extensions. These systems and their potential follow-ups are thus resources that we expect to drive translations of innovation into practice for the next decade.

SDSC is so much more than just a set of big machines and the programs around them. In fact, the vast majority of the nearly 270 employees and \$40 million of annual revenue is only peripherally connected to the HPC systems in our data center. The range of activities is so diverse that it is impossible to do it justice in our annual report. Instead, we selected among the myriad of projects and initiatives just a few to give you a flavor of this range.

Among these articles, let me provide some larger context. The WIFIRE activity led by Ilkay Altintas is the exemplar for a larger thrust of “convergence research.” At SDSC, convergence research is the practice of solving complex problems focused on societal needs via deep integration across multiple disciplines. It is translation of innovation into practice at its finest to solve societal challenges. To focus on activities of this kind, we have pivoted the division lead by Altintas, renaming it CICORE for “Cyberinfrastructure and Convergence Research and Education”.

Another innovation is a program we call RevUp, led by Mike Zentner. The idea behind RevUp is to strengthen the entrepreneurial culture at SDSC. RevUp supports principal investigators that are interested in developing software

products sustained via sales of service agreements, rather than or in addition to contracts and grants with funding agencies.

Finally, I'd like to encourage you to read the genomics feature where we provide a large historic arc of SDSC's involvement that started with an industry partnership with Janssen Pharmaceuticals, and connect that history to recent activities in expanding SDSC's engagement with health sciences researchers by building a dbGaP infrastructure into the Expanse Industry rack. Our aspirations here are to accelerate health sciences research, facilitate collaborations between academia and industry, and partner with industry on accelerating research toward new and better diagnoses and treatment of diseases.

I'd like to conclude on a personal note. It has been a remarkable year with lots of challenges, but also lots of inspirational moments. Ultimately, what matters most are the people in the SDSC community, and I feel very fortunate to serve you all as the director.

Frank Würthwein,
SDSC Director

Meet New SDSC Director Frank Würthwein

On Dec. 1, 2021, Frank Würthwein, Ph.D., was appointed executive director of the San Diego Supercomputer Center. He assumed the role initially as interim director last July when the former long-time director, Michael Norman, resumed full-time faculty status in the UC San Diego Department of Physics.

“It's an honor and my great pleasure to be taking on this responsibility, especially now, when SDSC is at the forefront of so many exciting new technologies to accelerate science,” said Würthwein. “I'd like to express my admiration for the SDSC community for its leadership in translating innovation into practice for decades.”

Würthwein is the lead of Distributed High-Throughput Computing at SDSC, a faculty member in the UC San Diego Department of Physics and a founding faculty member of the Halicioğlu Data Science Institute on campus. His research focuses on experimental particle physics, and in particular the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider. He continues to serve, as he has for many years, as executive director of the Open Science Grid (OSG), the premiere national cyberinfrastructure for distributed high-throughput computing.

“Frank is a well-known leader and innovator in the field of scientific computing,” said Norman at the time of Würthwein's appointment. “He has served on SDSC's executive team with distinction and understands what makes SDSC tick. He will do a great job as director.”

Würthwein received his Ph.D. from Cornell University in 1995. After holding appointments at Caltech and MIT, he joined the UC San Diego faculty in 2003. His research interest in particle physics include the search for dark matter, supersymmetry and electroweak symmetry breaking.

As an experimentalist, Würthwein is also interested in instrumentation and data analysis. Over the last few years this has meant developing, deploying and operating a worldwide distributed computing system for high-throughput computing with large data volumes.

Under Würthwein's leadership SDSC is focusing on translating innovation into practice. He is the principal investigator of the recently funded National Research Platform, an innovative, all-in-one system—computing resources, research and education networks, edge computing devices and other instruments—and a testbed for researchers to expedite science and enable transformative discoveries.



Subhashini Sivagnanam Named 2021 Pi Person of the Year

Each year SDSC recognizes an individual who exemplifies research in both science and cyberinfrastructure. For 2021, that Pi Person of the Year is Subhashini Sivagnanam.

As principal investigator (PI) of the Open Science Chain and co-PI of the Neuroscience Gateway (NSG), Sivagnanam provides SDSC's Data Enabled Scientific Computing Division with an array of leadership skills in roles ranging from software architect and high-performance computing (HPC) consultant to student mentor and conference coordinator.

Even while wearing many hats, Sivagnanam contributes to and enables neuroscience research and data provenance research while supporting HPC users. She also is the primary user-facing support resource for the Triton Shared Computing Cluster (TSCC) at UC San Diego.

"I am honored to be named Pi Person of the Year," said Sivagnanam.

According to Amit Majumdar, who leads SDSC's Data Enabled Scientific Computing Division, Sivagnanam originated the idea to apply blockchain technologies to manage research data provenance, winning back-to-back National Science Foundation (NSF) grants to fund the research.

Earlier this year, Sivagnanam was awarded Best Short Paper for the Applications and Software Track at PEARC21.

Sivagnanam's leadership in cyberinfrastructure includes her leadership in implementing Open OnDemand for the Expanse project. She is also a strong contributor to various NSF, National Institutes of Health (NIH) and NSG proposals, the NSF ACCESS proposal and to an NIH Bridge2AI proposal. She serves as PI and Co-PI on several cyberinfrastructure research projects.



opensciencechain.org



nsgportal.org



\$57.3M

ANNUAL REVENUE

\$35.6M

GRANT FUNDING

\$2.2M

INDUSTRY REVENUE INCREASE
(\$5.4M 2019-20; \$7.6M 2020-21)



50

SPONSORED RESEARCH AWARDS

54

SDSC-HOSTED PROGRAMS

638

PUBLICATIONS CITING
SDSC/XSEDE RESOURCES

25+

PETABYTES OF DATA STORED FOR
UC SAN DIEGO & UC USERS



36K

NATIONAL
USERS

71

ACTIVE ALLOCATIONS FOR
UC SAN DIEGO HPC USERS

80

HIGH SCHOOL STUDENTS
MENTORED

368

UC SAN DIEGO
USERS

51

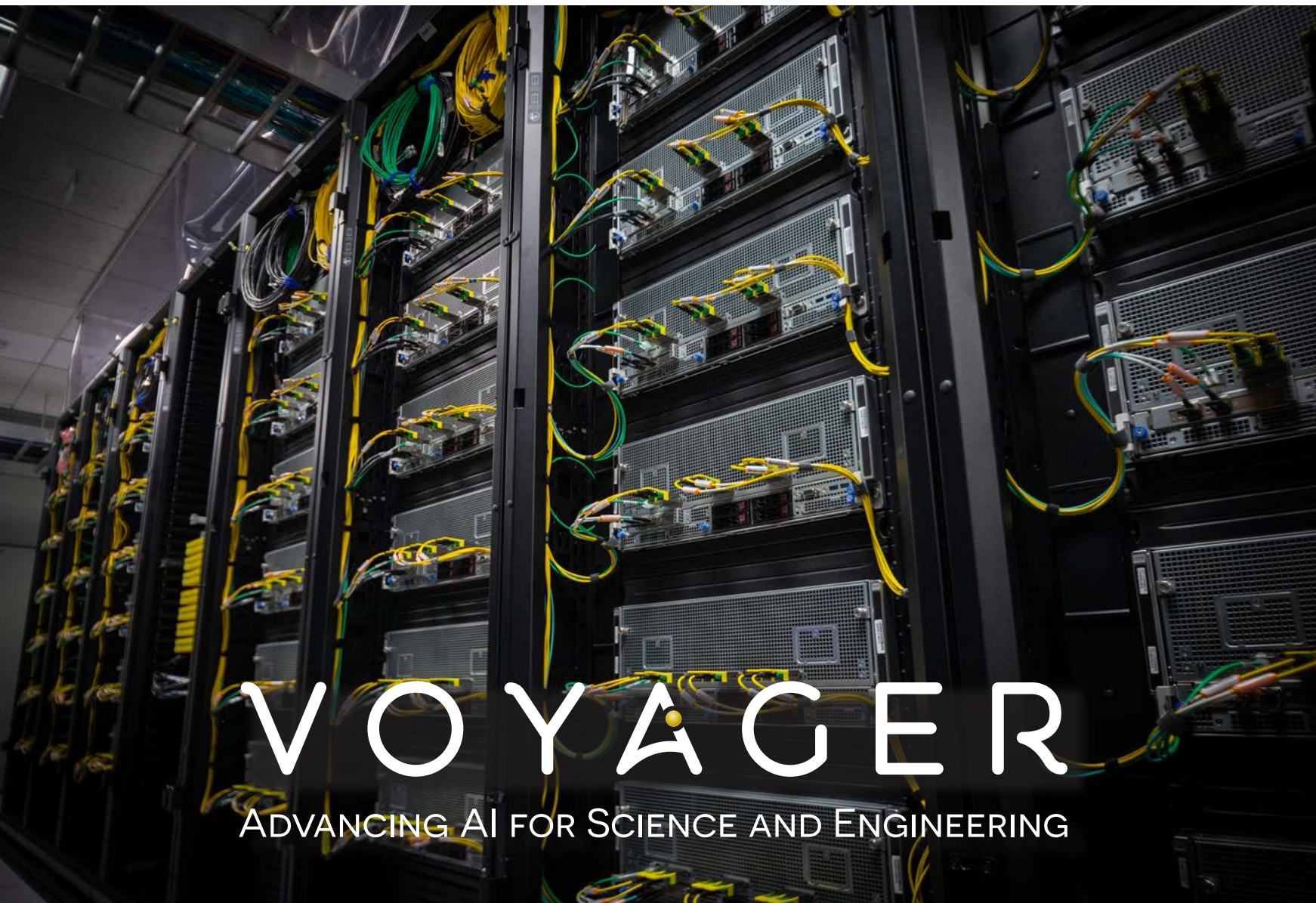
UC SAN DIEGO
RESEARCH GROUPS IN TSCC

269

SDSC EMPLOYEES
& VOLUNTEERS

2,700+

TRAINING & EVENT
PARTICIPANTS



VOYAGER

ADVANCING AI FOR SCIENCE AND ENGINEERING

Taking AI to New Places

From a Voyager Case Study by Intel

Scientists around the world are inundated with petabytes and exabytes of data, which can unlock new insights and discoveries that lead to breakthroughs in health, physics and many other sciences. To sift through and make sense of these mountains of data, researchers have begun turning to artificial intelligence (AI), deep learning (DL) and machine learning (ML). An AI-focused supercomputer called Voyager at the San Diego Supercomputer Center (SDSC) is ready to help scientists around the world discover and develop new methods for applying AI to their fields. Funded by the National Science Foundation (NSF) as an experimental system, Voyager enables scientists to try new approaches and design while serving as a platform for building and optimizing algorithms for accelerated ML model training and inference.

“AI is becoming a discipline itself. Unlike general purpose computing, tools and technologies focused on deep learning are different. This is hardware specially built for AI—Gaudi for training and Goya for inference. We need this hardware to experiment, test and learn in order to advance AI approaches,” said Amit Majumdar, director of the Data Enabled Scientific Computing Division at SDSC and principal investigator for Voyager.

The availability of more data across all scientific domains, driven by new acquisition sources, higher resolution models and federated repositories, is resulting in massive and increasingly complex datasets. To sort and synthesize this data, scientists are applying AI, especially DL techniques and algorithms, using new purpose-built computational methods to extract insights. But there is yet a lot to learn about how to best analyze these massive and complex datasets.

GPUs are typically the go-to architecture for large-scale DL training workloads. But, high-performance computing (HPC) and AI are still in a phase of emerging technologies that present alternatives to existing AI methods. As new solutions appear from companies like Habana Labs (an Intel company), researchers need access to these technologies at scale in order to explore how they can further their insight. Scientists need to understand how to train and deploy their models on these new machines and inform their respective communities of their learnings.

Voyager at the San Diego Supercomputer Center, located at the UC San Diego, offers a platform for AI acceleration research and development.

“We talked to several scientists about what they needed for their research and AI is becoming an important component of their research,” said Majumdar. “When the NSF requested proposals for unique experimental supercomputers, we began architecting Voyager and sought an NSF grant. Voyager is one of the first AI-focused experimental systems to join the NSF ecosystem.”

Voyager is not only one of the first NSF AI-focused supercomputers, but also Habana Labs’ first commercial deployment of its GAUDI® training and GOYA™ inferencing AI processors. Over the years, AI workloads have been adopted and designed for GPU architectures, so many AI solutions today rely on expensive GPUs to train algorithms on large data sets. Unlike these systems, Habana processors were purpose-designed at conception expressly to drive improved AI compute efficiencies in the data center. Habana accelerators enable dramatic performance improvements for training and inferencing.

Voyager comprises 42 training nodes of Supermicro X12 Gaudi Training Servers powered by dual-socket 3rd Gen Intel® Xeon® Scalable processors. Each training node contains eight GAUDI HL-205 training processor cards. Two Supermicro SuperServer nodes are deployed for inferencing, each with two 2nd Gen Intel Xeon Scalable processors and eight GOYA HL-100 inferencing PCIe cards. An alternative to GPU-based AI systems, the combination of Intel Xeon Scalable processors and Habana training and inferencing processors provides a unique and powerful configuration for accelerated training and inference. Using Voyager, scientists can explore how to best take advantage of these computational devices for particular types of workloads.

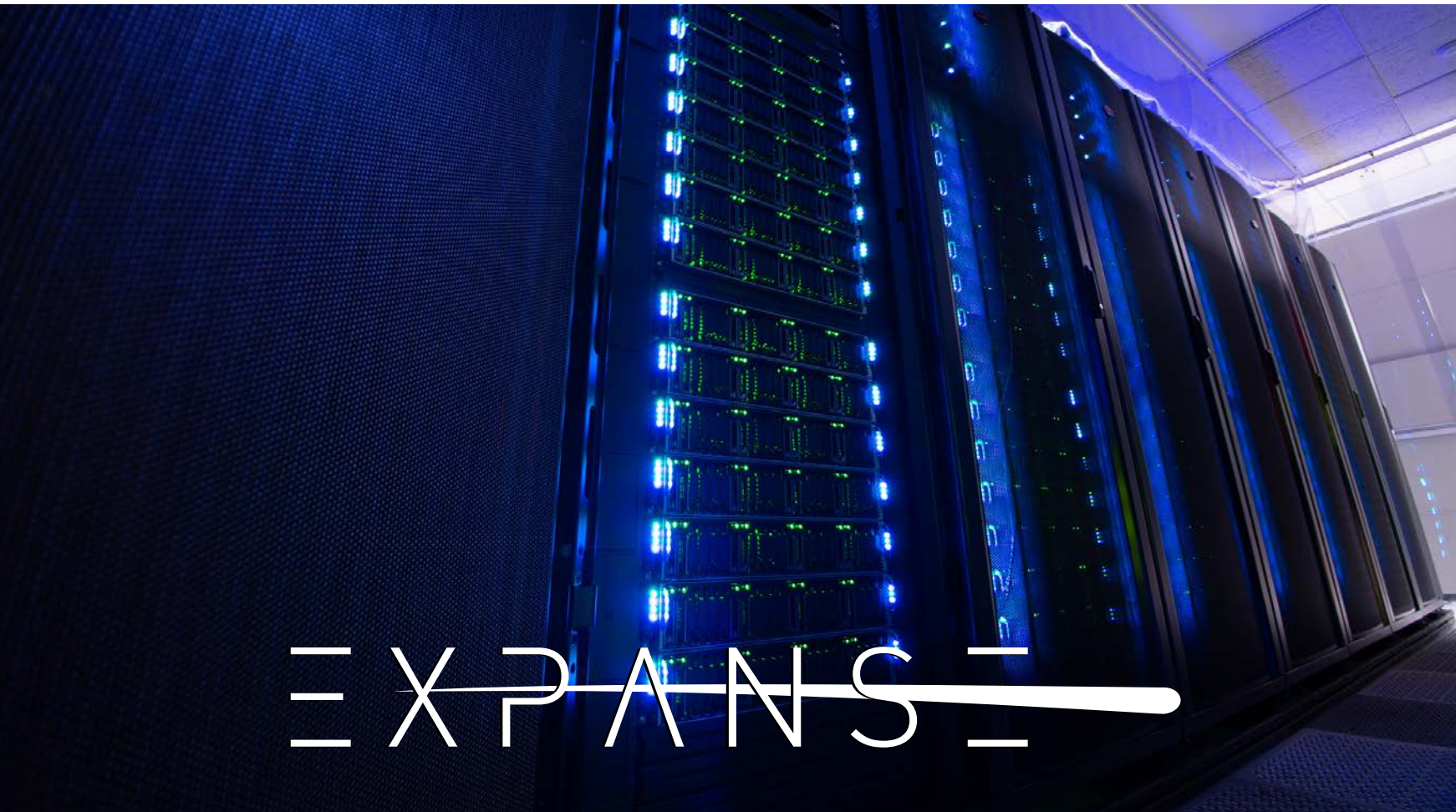
The NSF experimental program for Voyager enables access to a select group of scientists for three years. Several disciplines will be represented, including high-energy physics, biology, genetics, materials science, atmospheric and astronomic sciences, plus others. After three years, the new system will be open to the entire NSF community for accelerated AI projects.



SDSC Builds AI-Focused “Voyager” Supercomputer
www.intel.com/content/www/us/en/customer-spotlight/stories/san-diego-supercomputer-customer-story.html



Preparing for the future today:
Expanding U.S. researchers’ access to advanced computing systems
www.nsf.gov/news/special_reports/announcements/062220.jsp



EXPANS

Computing without Boundaries

In late 2020, SDSC introduced its current petascale supercomputer, Expanse, which supports the center’s theme of “Computing without Boundaries.” The high-performance-computing resource features powerful CPUs, GPUs and a data-centric architecture that supports a wide range of scientific workloads including experimental facilities, edge computing and public clouds.

“The name of our new system says it all,” said former SDSC Director Michael Norman, a UC San Diego astrophysicist who remains the principal investigator (PI) for Expanse. “With innovations in cloud integration and other features such as composable systems, as well as continued support for science gateways and distributed computing via the Open Science Grid (OSG), Expanse allows researchers to push the boundaries of computing and substantially reduce their times to discovery.”

A key innovation of Expanse is its ability to support composable systems, which can be described as the integration of computing elements such as a combination of CPU, GPU and other resources into scientific workflows that may include data acquisition and processing, machine learning and traditional simulation. Expanse also supports integration with public cloud providers, leveraging high-speed networks to ease data movement to and from the cloud, and a familiar scheduler-based approach.

Expanse is designed for modest-scale jobs of just one core to several hundred cores. This includes high-throughput computing jobs via integration with the OSG, which can have tens of thousands of single-core jobs. Those modest-scale jobs are often 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 fields such as earth sciences and biology, depend upon these medium-scale, innovative systems for much of their productive computing.

Expanse’s standard compute nodes are each powered by two 64-core AMD EPYC 7742 processors and contain 256 GB of DDR4 memory, while each GPU node contains four NVIDIA V100s connected via NVLINK, and dual 20-core Intel Xeon 6248 CPUs. Expanse also has four 2 TB large memory nodes. The entire system, integrated by Dell, is organized into 13 SDSC Scalable Compute Units (SSCUs), comprising 56 standard nodes and four GPU nodes, and connected with 100 GB/s HDR InfiniBand.

Remarkably, Expanse delivers more than 90,000 compute cores in a footprint of only 14 racks. Direct liquid cooling (DLC) to the compute nodes provides high core count processors with a cooling solution that improves system reliability and contributes to SDSC’s energy efficient data center.

The new system became allocable in March 2021, and it serves as a key resource within the NSF’s Extreme Science and Engineering Discovery Environment (XSEDE), which comprises some of the most advanced integrated digital resources and services in the world. The NSF award for Expanse runs through September 2025 and is valued at \$10 million for acquisition and deployment, plus an estimated \$12.5 million for operations and maintenance.



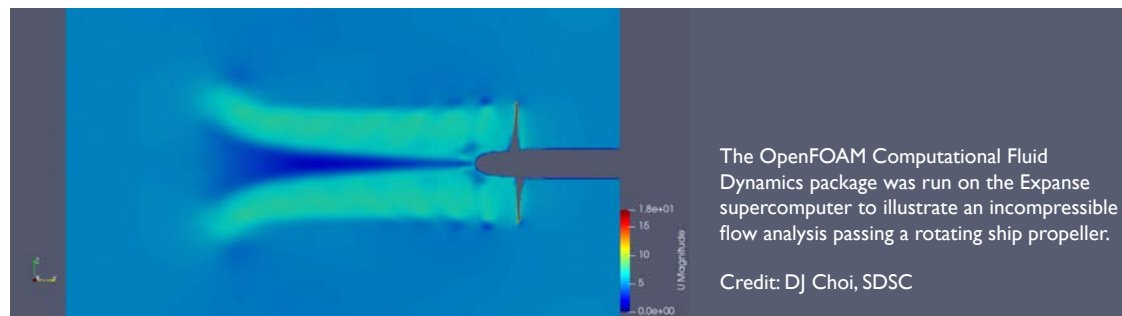
SDSC’s Expanse Supercomputer Formally Enters Production

www.sdsc.edu/News%20Items/PR20201208_Expanse.html



Composing ‘Expanse’: Building Blocks For Future HPC

www.nextplatform.com/2020/06/17/composing-expanse-building-blocks-for-future-hpc/



The OpenFOAM Computational Fluid Dynamics package was run on the Expanse supercomputer to illustrate an incompressible flow analysis passing a rotating ship propeller.

Credit: DJ Choi, SDSC

SDSC Clears Pathways to Discovery for Industrial Partners

By Avani Ranka, Research Experience for High School Students Program, SDSC UC San Diego

SDSC’s Expanse supercomputer has added significant capacity and capability in support of SDSC’s base of scientific users, including its industry partners. With the Industrial Partners Program (IPP) in mind, SDSC previously acquired an “Industry Rack,” which comprises more than 7,000 AMD x86-compatible computing cores and 16 NVIDIA V100 Graphics Processing Units. It also partnered with Infrastructure-as-a-Service provider Core Scientific to implement a portal-driven provisioning and payment system for industrial users of Expanse. Now, SDSC is adding software capabilities to industrial users with the OpenFOAM Computational Fluid Dynamics package.

www.sdsc.edu/News%20Items/PR20210726_OpenFOAM.html



NATIONAL RESEARCH PLATFORM

Designed for Growth & Inclusion



Accelerating Science with Bold National Research Platform

SDSC responded to the NSF’s call for a cyberinfrastructure ecosystem that meets the needs of today’s data-intensive science with the Prototype National Research Platform (NRP). It’s an innovative, all-in-one testbed system—computing resources, research and education networks, edge computing devices and other instruments—for science drivers as diverse as the platform itself. The newly funded project will expedite science and enable transformative discoveries across science themes.

For the first-of-its-kind resource, the NSF awarded SDSC \$5 million over five years, with matched funding for systems operation. The award supports hardware and deployment across three facilities: on the East Coast at the Massachusetts Green High Performance Computing Center (MGHPCC) in Mount Holyoke, MA; in the Midwest at the University of Nebraska–Lincoln (UNL) and on the West Coast at SDSC, as well as five data caches in the Internet2 network backbone.

“Supercomputing continues to be instrumental 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 innovations in

cyberinfrastructure and in building high-performance computational systems specifically designed for scientific research. The NSF sought a national innovator to help improve the crucial resources required to further enhance American prowess in scientific discovery.”

According to Frank Würthwein, director of SDSC and principal investigator (PI) on the NRP project, the funding will enable him and the co-PIs from UNL and MGHPCC to work with the research community to explore NRP’s experimental architecture.

“NRP provides resources and capabilities for diverse science, plus the expertise of systems people and the user community,” said Würthwein. “It’s an open system designed for growth and inclusion—a way for academic institutions to join a national system and, through their participation, enlarge and enrich the HPC ecosystem.”

Key components of the NRP architecture include research and education networks, compute accelerators, heterogenous computing resources (e.g., edge computing near the instrument location), a content delivery network to provide easy access to any data, anytime, from anywhere, and an overall systems software stack that allows for central management of the system, while remaining open for growth. Innovative technical features of the HPC subsystem include a mix of field programmable gate array (FPGA) chips, graphics processing units (GPUs) with memory and storage in a fully integrated extremely low-latency fabric from GigaIO.

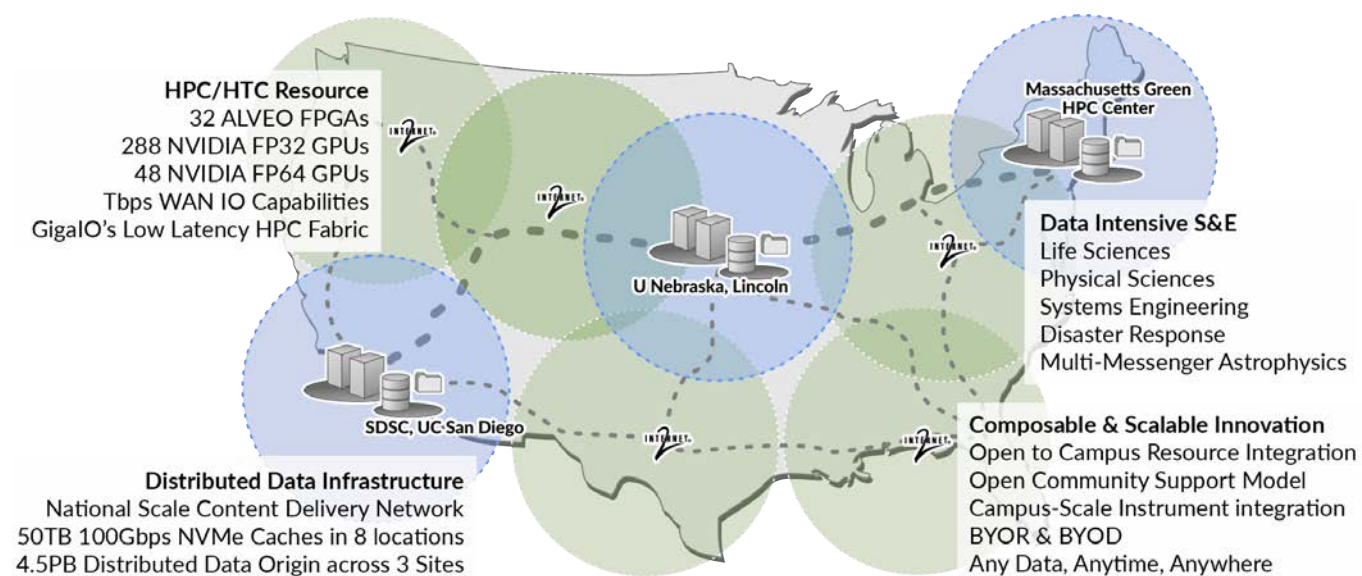
The broader impact of NRP is focused on four themes:

1. empowering underrepresented and/or under-resourced researchers by making them “co-owners” of NRP, via the “Bring Your Own Resource (BYOR)/Bring Your Own Device (BYOD)” program;
2. societal wellbeing, focusing on health and disaster response;
3. STEM education, workforce development and outreach; and
4. enhancing industrial competitiveness.

UNL will lead NRP’s infrastructure operations using Kubernetes to manage remote systems at the three primary sites and Internet2, as well as the BYOR locations. System deployment will take place at SDSC on the UC San Diego campus, where several science drivers from astrophysics, molecular dynamics, health sciences and more will participate.

The NRP project is structured as two distinct phases: the testbed phase, which includes evaluation and expansion of NRP for science and engineering research (years one to three), and the allocations phase, which involves widespread adoption of the NRP resource (years four to five).

“We aspire for our science drivers to not just use the infrastructure but to intellectually impact each other, forming mutual support networks that we enable via managed Slack channels in addition to more standard user support. We consider this in itself a testbed for scalable user support of an HPC system,” said co-PI Mahidhar Tatineni, SDSC.





National Artificial Intelligence Research Resource

SDSC Takes a Seat at the Table of Federal Task Force for Advancing Artificial Intelligence

In June 2021, the Biden Administration announced the National Artificial Intelligence Research Resource (NAIRR) Task Force—a group of 12 individuals from academia, government and industry, including former SDSC Director Michael Norman. NAIRR’s purpose is to support AI researchers’ access to federal data in order to keep the U.S. at the forefront of emerging technology.

“UC San Diego’s faculty have a history of being called to contribute significantly to our nation’s thought leadership on a variety of policy, economic, scientific and social issues,” said UC San Diego Chancellor Pradeep K. Khosla. “Professor Norman’s expertise in the computer simulation of astronomical phenomena using supercomputers, and the development of the numerical methods to carry them out, will provide the NAIRR



Former SDSC Director and Distinguished Professor of Physics Michael Norman has been appointed as a member of the NAIRR Task Force.

taskforce first-hand knowledge to better understand how researchers collaboratively use data analysis and cloud computing in their fields.”

According to the National Artificial Intelligence Initiative Office, the NAIRR is envisioned as a shared computing and data infrastructure to provide AI researchers and students compute resources and high-quality data, along with appropriate educational tools and user support. The goal for NAIRR is to democratize access to resources and tools that fuel AI research and development, expanding the ability of academia, industry and government to explore innovative ideas for advancing AI in a range of scientific fields and disciplines.

Led by officials from the White House Office of Science and Technology Policy (OSTP) and the National Science Foundation (NSF), the task force’s mission is to submit reports to Congress addressing a range of topics that include the appropriate ownership/administration of the NAIRR; requirements for security; assessments of privacy, civil rights and civil liberties requirements; and a plan for sustaining the resource, including through public-private partnerships.

“I am honored to be selected to serve the nation’s interest in this important initiative. I look forward to working with my fellow task force members to discuss needs, capabilities and opportunities for investment in the AI space,” said Norman.

The National AI Initiative Act of 2020 called for the NSF to work with OSTP to form the NAIRR Task Force to investigate the feasibility of establishing a NAIRR, and to develop a roadmap outlining how such a resource could be established and sustained. Throughout its work, the task force will consult a range of experts and stakeholders from government agencies, private industry, academia and civil and disabilities rights organizations. It will also be informed by ongoing interagency efforts.



National Artificial Intelligence Initiative
Overseeing and Implementing the United States National AI Strategy
www.ai.gov



Comet Blazes a New Trail

SDSC’s petascale Comet supercomputer—which can perform nearly three quadrillion operations per second—concluded formal service as a National Science Foundation (NSF) resource and transitioned to exclusive use by Scripps Institution of Oceanography (Scripps) Center for Western Weather and Water Extremes (CW3E). The transition, which occurred during summer 2021, enables CW3E researchers to leverage Comet’s computing capabilities to improve weather and hydrological forecasts with the goal of enhancing the decision-making process associated with reservoir management over California, which could result in increased water supply and reduced flood risk over the region.

“Comet has been a workhorse system for the NSF, and we are thrilled with the opportunity to extend its useful life in support of this important work,” said Michael Norman, former director of SDSC and principal investigator for Comet.

According to Luca Delle Monache, academic program manager of Climate, Atmospheric Science and Physical Oceanography at Scripps, the exclusive access to Comet will allow CW3E to advance several computational projects that would not be possible otherwise.

“The ability to perform computational experiments that just one year ago we couldn’t even think of and to strengthen our collaboration with key partners as the National Weather Service (NWS) is an exciting aspect of this agreement,” he said.

Delle Monache explained that CW3E’s computational projects will span a broad scope from basic to applied research. The projects will range from trying to answer fundamental questions on the formation and evolution of atmospheric rivers, and the interactions between the atmosphere and the ocean, to studying orographic precipitation, which is generated when the moist air mass associated with an atmospheric river is lifted over mountains and condenses into rain or snow. The projects will also include developing high-resolution, sub-seasonal to seasonal predictions; testing and improving operational systems run from the NWS; expanding CW3E near real-time operational capabilities for weather and hydrology with high resolution deterministic prediction and an ensemble with an unprecedented large number of members.

“We will also develop and test new machine learning algorithms and dynamical models leveraging Comet’s several graphics processing unit (GPU) nodes,” said Delle Monache. “Moreover, we will develop and test new data assimilation schemes.”

Delle Monache noted that a CW3E Supercomputing Advisory Group (CSAG) has been established. The group includes leading experts from Scripps, the National Center for Atmospheric Research, University of Arizona, University of Colorado at Boulder, Jet Propulsion Laboratory and the European Centre for Medium-Range Weather Forecasts.

“CSAG is helping us in designing some of the projects mentioned above. In addition to those, we are also exploring the possibility to dedicate substantial computational resources to students and universities from underrepresented groups, to allow for new or expand existing modeling experiments as part of their dissertation research,” said Delle Monache.

According to CW3E’s HPC Lead Pat Mulrooney, compared to the computational capabilities of other supercomputer resources, Comet provides access to 144 NVIDIA GPUs, “not a common feature in the nation’s supercomputer clusters,” Mulrooney said. “Also, Comet has on node solid-state drives (SSDs) and the highest random access memory (RAM) to core processing units (CPU) core ratio.”

In late 2013, the NSF awarded SDSC an initial \$12 million to build and operate Comet, which went into production in 2015. Since then the total NSF funding increased to about \$27 million. Leading up to the end of Comet’s NSF tenure, SDSC and CW3E worked jointly to develop the plan and timeline for supporting CW3E operations, including on-boarding of CW3E users and data.

“Over the last six years, Comet and the team at SDSC have provided a great resource to the science and engineering community, benefiting many researchers and educators across the nation and enabling insights and discoveries across all of science and engineering,” said the NSF’s Office of Advanced Cyberinfrastructure Director Manish Parashar. “NSF would like to express our appreciation to the SDSC team for their commitment to supporting the national research enterprise and inspiring new innovations in research cyberinfrastructure.”



Center for Western Weather and Water Extremes (CW3E)
cw3e.ucsd.edu



National Weather Service (NWS)
www.weather.gov



Working to Fight Fires Proactively

According to scientists, a century of suppressing wildfires has created a dangerous stockpile of flammable vegetation on landscapes. This “fire fuel” has fed the megafires that put human life and property at risk and permanently destroy ecosystems.

To mitigate the threat of these massive blazes, fire managers and responders can set small controllable fires ahead of time to reduce the flammable materials. A new decision support platform called BurnPro3D is designed to help them accomplish this with a proactive approach.

Ilkay Altintas, chief data science officer and director of the WIFIRE Lab at SDSC at UC San Diego, is the principal investigator on the project. The team of researchers includes Yolanda Gil from the University of Southern California (USC), Rod Linn of Los Alamos National Laboratory (LANL) and Kevin Hiers of Tall Timbers Research Station (TTRS) in Florida. The team, which also includes other government, academic and industry partners such as the U.S. Geological Survey (USGS) and the U.S. Forest Service, recently received \$5 million in funding from the National Science Foundation through its Convergence Accelerator program for the dynamic platform that combines the strength of data synthesis and technology.

Leveraging the WIFIRE Commons data-sharing and artificial intelligence (AI) framework, the BurnPro3D platform uses next-generation fire science in prescribed burns for preemptive vegetation treatment at an unprecedented scale.

“The NSF convergence accelerator program is all about innovation for societal impact. We have been developing key infrastructure and partnerships in this area for the last eight years, and more recently working with our BurnPro3D collaborators to include next-generation fire science and AI in various aspects of the project,” said Altintas. “We are looking forward to working with the broader community to maximize the impact of this work.”

To achieve its vision, the BurnPro3D team is developing specific innovations to:

- use technologies to fuse data coming from diverse sources and prepare it for fire modeling,
- integrate physics-based machine learning within next-generation fire models to use deep learning to understand complex processes that drive fire behavior,
- apply constraint optimization methods to address complex tradeoffs in the decision process for the placement and timing of controlled burns and
- employ explainable AI to increase the interpretability of data and models by diverse users all along the decision-making chain.

“USC is thrilled to partner with our strong AI expertise in semantic technologies, optimization and machine learning. Sustainability is a priority for our research, and fire-modeling capabilities are a crucial aspect of planning how to preserve our resources and protect ourselves against climate extremes such as droughts,” said Yolanda Gil, director of New Initiatives in AI and Data Science at USC’s Viterbi School of Engineering.

As the former head of one of the nation’s largest prescribed fire programs, TTRS’ Kevin Hiers has seen firsthand how complex burn planning can be and how it is underserved by fire science initiatives.



“The truth is that more well-planned fire is needed to solve the western wildfire problem, and technology can help us safely plan those good fires. Unfortunately, the gap between the high-tech innovations and fire managers has gotten bigger, which limits applications of new models or AI to solve this problem,” said Hiers. “The BurnPro3D and WIFIRE Commons are designed to bridge that gap. We are excited to work with managers and fire scientists to apply these next generation tools to help safely plan prescribed fires and optimize their placement on the landscape for maximum risk reduction.”

According to Paul Steblein, wildland fire science coordinator for the USGS, this project will help the federal fire science community to explore game-changing innovation information technology.

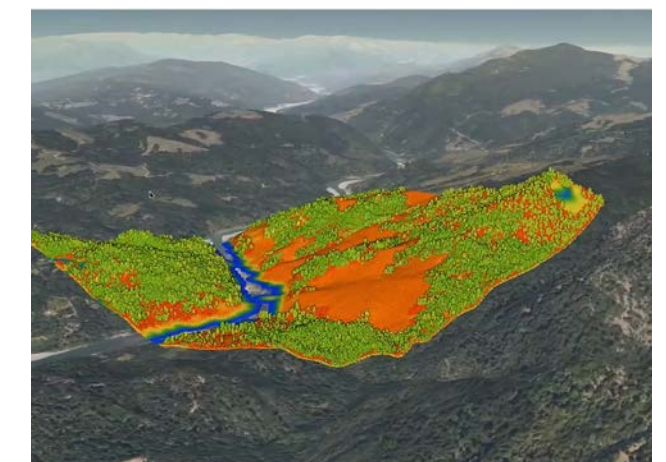
“The BurnPro3D project will help to rapidly improve new models, planning tools and data fusion at a time when prescribed fire and other fuel treatments will be receiving significant investment to reduce wildfire risks,” Steblein said.

LANL’s Rod Linn, the WIFIRE team member who led the development of the modeling tools behind BurnPro3D explained, “The aim of this project is to empower researchers and fire managers to plan fuel treatments and prescribed burns by accessing to next-generation modeling tools which are enabled by AI, data fusion and the IT infrastructure to address site-specific fire management challenges.”

BurnPro3D is one of 10 convergent multidisciplinary research proposals that recently received a collective \$50 million in NSF support.



WIFIRE Lab: BurnPro3D
wifire.ucsd.edu/burnpro3d





CERN Data Center.
Credit: Maximilien Brice/CERN

Software Architect's Demo Captures the Attention of CERN

San Diego Supercomputer Center's Shava Smallen inspired a new cloud computing international partnership

SDSC, a pioneer in advanced computation and data analytics, has teamed up with CERN—the European laboratory for particle physics famously associated with the discovery of the Higgs boson—to bring public cloud computing to researchers worldwide.

After considering other national and international systems, CERN chose SDSC at UC San Diego because of the significant advances it has made in working with researchers through its CloudBank portal. In fact, it was a demonstration by Shava Smallen, co-principal investigator and lead architect for the CloudBank user portal, that initially caught the eye of CERN's cloud team whose members will work with SDSC to adopt the CloudBank user portal at CERN.

"We are excited to collaborate with CERN and support its efforts," Smallen said. "Though the CloudBank portal was originally designed for NSF workflows, we have generalized



Shava Smallen is co-principal investigator and lead architect for the CloudBank user portal.

our framework for multiple funders, and we look forward to CERN's experience further driving improvements that benefit all CloudBank users."

CloudBank is a National Science Foundation (NSF)-funded service co-developed by SDSC at UC San Diego, UC Berkeley and the University of Washington. The team was awarded \$5 million in funding from the NSF in 2019 to help researchers access and use public cloud computing resources.

João Fernandes, senior member of CERN staff and project leader in the CERN IT Department, is tasked with launching "CloudBank Europe," which is supported by the Next Generation Internet (NGI) programme of the European Commission.

"We were motivated by the desire of our research teams to onboard commercial cloud for a set of challenging research use cases. So, we teamed up with SDSC to expand the CloudBank model at CERN, where researchers from all over the world collaborate," Fernandes said. "We have seen increased needs for accessing cloud-based resources—in particular for Machine Learning as a Service (MLaaS), High-Performance Computing as a Service (HPCaaS) and Quantum Computing as a Service (QCaaS). The relative strengths of the cloud vendors wax and wane, so we really wanted to solve the procurement and governance challenges just once, and, at the same time, establish a sustainable cost-optimization process leveraging multiple public cloud service offerings."

Fernandes said that CloudBank provides a model for dealing with procurement and cost-optimization challenges when using public cloud, which his team hopes to use as inspiration for CERN's procurement strategy.

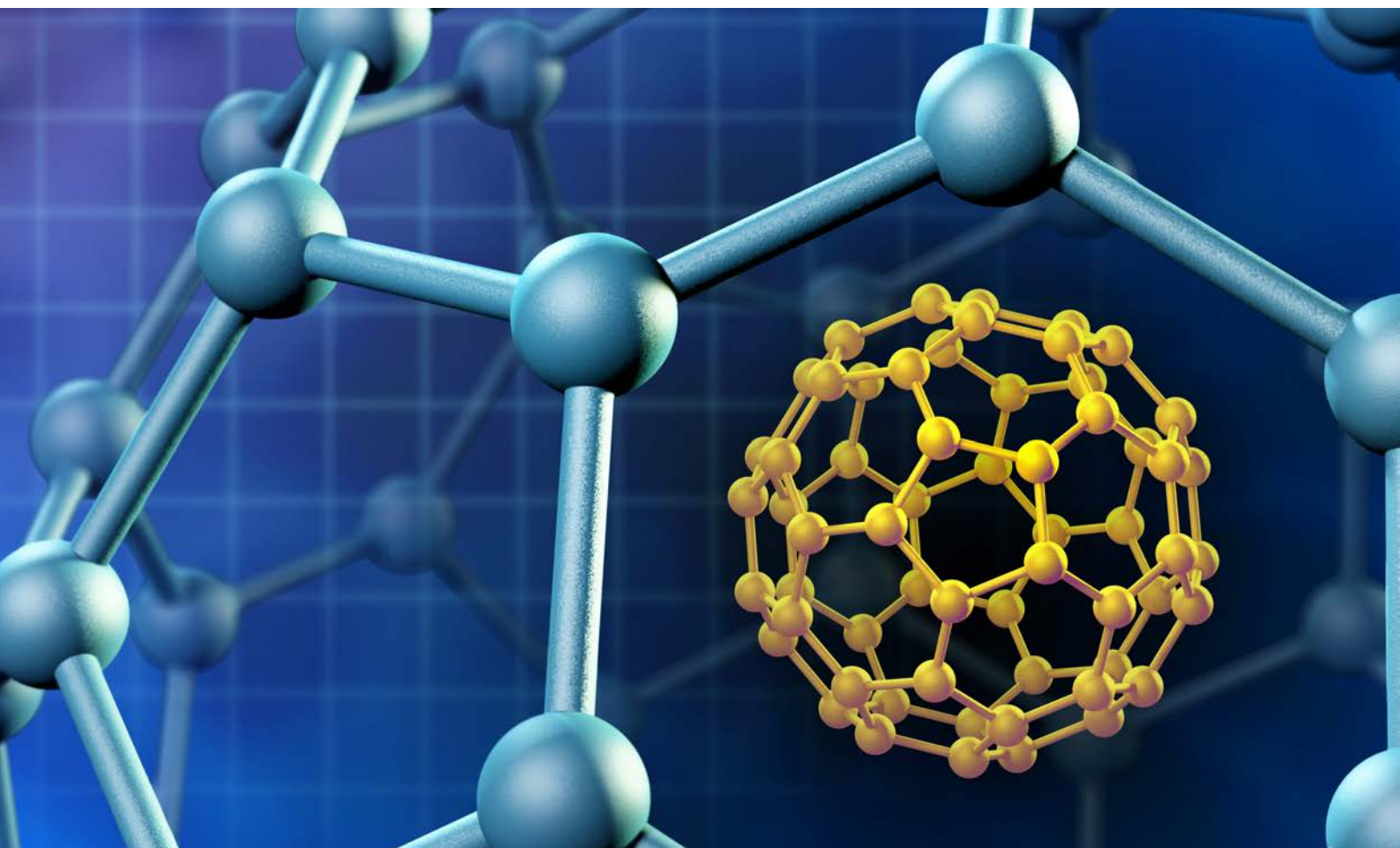
The new international partnership also leverages SDSC's alliance with Strategic Blue, a UK-based Fintech company that helps organizations optimize procurement of cloud services. SDSC and Strategic Blue help researchers in the U.S. adopt their choice of cloud with fewer procurement challenges, while stretching their research dollars as far as possible.



CloudBank
www.cloudbank.org



CloudBank Europe
Next Generation Internet
ngiatlantic.eu/funded-experiments/cloud-bank-eu-ngi



A Quantum Leap for Molecular Simulations on GPUs

Quantum mechanics (QM), a theory that describes the physical properties of nature on the atomic and subatomic scale, is the basis for molecular and materials scientists who develop materials for energy storage and drug discovery purposes.

The challenge with these useful but futuristic products is that the QM calculations to describe the many properties of molecules and the materials they comprise require a lot of computer power.

To tackle this challenge, a small team of postdoctoral scholars, led by SDSC's Andreas Goetz and Kenneth Merz at Michigan State University (MSU), developed software that takes advantage of powerful graphics processing units (GPUs) for these complex QM calculations of molecules, allowing researchers to address problems that would otherwise be computationally intractable in a range of research areas—from computational materials design to drug design.

"Our code can also be used in a high-throughput fashion to generate data for Artificial Intelligence-based molecular design approaches, such as developing improved materials and drugs," noted Merz.

The SDSC and MSU teams jointly designed the research, developed algorithms and engineered the software. They used data collected on SDSC's Expanse supercomputer for their research, the results of which were featured in two peer-reviewed publications: the *Journal of Chemical Theory and Computation* (JCTC) and the *Journal of Chemical Information and Modeling* (JCIM).

"The need of freely available open-source software to enable computational molecular sciences on modern computers—from desktops to supercomputers with thousands of GPUs—inspired our efforts," said Goetz, director of the Computational Chemistry Laboratory at SDSC. "This research provides open-source software for efficient QM calculations and for mixed quantum/classical (QM/MM) simulations of molecular systems on graphics processing units."

In their JCTC paper, "Harnessing the Power of Multi-GPU Acceleration into the Quantum Interaction Computational Kernel Program," the researchers explained supercomputers in a simplified way as made up of many tightly integrated computers that work in parallel together to solve complex computational problems. So, their software, called QUICK, coordinates computations across multiple GPUs to get solutions more quickly into the hands of researchers.

"This is like coordinating a fleet of trucks to deliver time-sensitive goods quickly from one place to another," explained Madushanka Manathunga, postdoctoral scholar at MSU. "We had to dissect the mathematical equations underlying the QM models to understand how we can efficiently parallelize the computations and effectively distribute the work onto multiple GPUs."

In their JCIM paper, "Open-Source Multi-GPU-Accelerated QM/MM Simulations with AMBER and QUICK," the researchers

considered their QM/MM coupling, which uses their faster GPU code for the complex QM calculations of, in this case, an active site of an enzyme together with a simplified MM model for the remainder of the protein beyond the active site region.

"It's exciting to have this multi-GPU QM code QUICK available for QM/MM simulations within the AMBER suite of programs so the community can take advantage of our code base to add new features and to use it to solve a range of chemical and biological problems," Merz said.

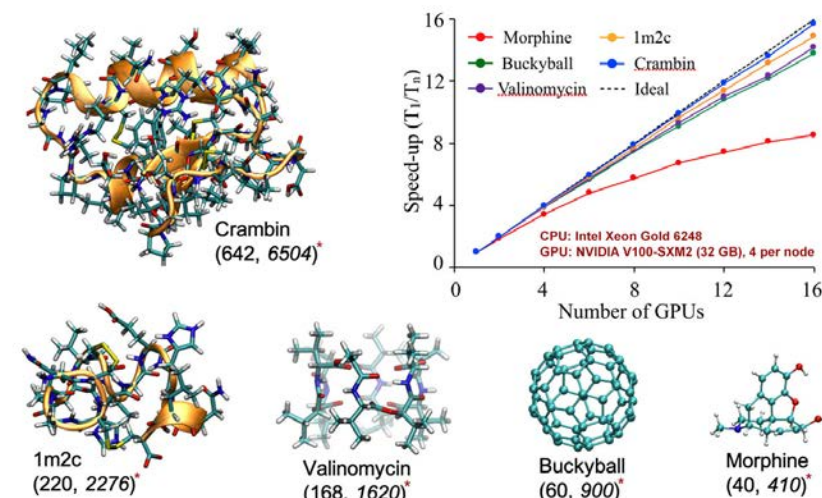
According to the researchers, their software enables so-called ab initio Hartree-Fock and density functional theory (DFT) QM calculations. Due to its favorable ratio of computational cost to accuracy, DFT is widely used in computational molecular sciences.

"The features implemented in QUICK are available in commercial programs, but these require costly license fees," Vinicius Cruzeiro, postdoctoral scholar at SDSC, who co-led the software engineering work and co-wrote the papers with Manathunga, said. "Therefore, it is rewarding to work on a project that aims to bring these features free-of-charge to researchers all over the world, especially to those impacted by paywalls."

Implementing their software on GPUs written in a special programming language, the researchers used the message passing interface (MPI) to distribute work onto different GPUs both within and across compute nodes. Part of this work was initiated last year at the first digital GPU Hackathon co-hosted by SDSC and Nvidia.

"All of this work was done remotely due to the Covid-19 pandemic, but this is not a big problem for software-related work," Goetz said.

Multi-GPU speedup of B3LYP/6-31G** density functional theory nuclear gradient calculations with the QUAntum Interaction Computational Kernel (QUICK) program for a range of organic molecules and proteins. Benchmarks were performed on the SDSC Expanse supercomputer. Image credit: M. Manathunga, A.W. Goetz



* The number of atoms and number of basis functions are given in brackets.



Condor at rest, wings spanned.
Credit: San Diego Zoo Wildlife Alliance

‘Teeming Up’ with San Diego Zoo Wildlife Alliance and ASU to Create New Spatial Ecology Gateway

Spatial ecologists are field experts who study the movements of wildlife animals within a specific geographic area, or their “home range.” Historically, this type of study was limited to a small subset of experts with the right tools. But a group of researchers from San Diego Supercomputer Center (SDSC) at UC San Diego, San Diego Zoo Wildlife Alliance (SDZWA) and Arizona State University (ASU) “teemed up” to make it possible for others to join the herd through the Spatial Ecology Gateway (SEG).

Funded by the National Science Foundation’s (NSF) Office of Advanced Cyberinfrastructure, SEG is envisioned as an innovative computational resource that will enable researchers, students, wildlife managers and others to upload biotelemetry data—typically Global Positioning System (GPS) readings—and construct home ranges that allow them to interpret animals’ use of space. It will allow for exploration of important problems in ecology, including managing habitat impacts such as shrinking home ranges, while letting users focus on the science rather than on mastering the technology helpful to their work.

“A lot of GPS use is tracking data, but most ecologists are not computational researchers. They are more at home with fieldwork that involves putting collars on bears and finding spiders on the rainforest floor,” said Robert Sinkovits, director of Scientific Computing Applications at SDSC and

principal investigator of the project. “SEG will be a tool to make the algorithm accessible to them.”

Spatial ecologists typically work with two-dimensional (2D) data, which is adequate for tracking many animals who move by walking. But not so much for animals like condors and other birds who fly in the vertical dimension.

Wanting to find a way to address the expense of computation in three dimensions (3D), Sinkovits and Mona Wong (SDSC), James Sheppard (SDZWA) and Jesse Lewis (ASU), proposed SEG, which is analogous to the trends in making bioinformatics tools (e.g., Basic Local Alignment Search Tool or BLAST) widely available through the web for anyone with questions or insights to answer problems in genomics.

The collaborators turned to 3D algorithm developer Jeff Tracey at the United States Geological Survey (USGS). Then, they sought the help of a student at SDSC who performed the 3D-code work adapted from a technique by Lewis. As a result, SEG users will have the option to generate 2D or, where applicable, 3D home ranges. Users will also be provided with less computationally intensive tools to perform exploratory analyses.

“Many animals have a major 3D aspect to their movements, such as deep diving whales and soaring birds. Wildlife biologists have been deploying miniaturized biotelemetry devices to track the 3D movements of wild animals for decades, yet only very recently have they been able to investigate the full 3D nature of these ecological ‘Big Data’ because of limitations in mathematical methods and the high computing demands needed to generate realistic 3D models of animal home ranges,” explained Sheppard. “Our initiative will remove these roadblocks by enabling biologists to crunch their animal tracking datasets using powerful 3D home range estimators coupled with the extraordinary computing power of the San Diego Supercomputer Center.”

Sinkovits said that SEG targets a user base of academic ecologists and wildlife managers, but others with interest in addressing ecological challenges can benefit from it, too. For example, applications of SEG may extend to classroom projects, basic research into problems in wildlife ecology, environmental impact studies and mitigation of adverse outcomes, such as habitat fragmentation or increased human-wildlife interaction resulting from new development.

“The Spatial Ecology Gateway will make computationally intensive methods evaluating animal space use available to a wider audience, which can help us better manage and conserve a variety of wildlife species in the U.S. and globally,” said Lewis.

Since the concept of the home range was introduced in the 1940s, ecologists have progressed from very simple to more complex quantitative approaches to studying how animals move within their environments. Although SEG will eventually deploy multiple tools for constructing home ranges, its initial emphasis will be on the Brownian Bridge Movement Model (BBMM)—a model that provides an estimate of the movement path of an animal using discrete location data obtained at relatively short intervals of time. The BBMM has helped revolutionize understanding of the use of space in animal populations.

“It is such an exciting project,” said Sinkovits. “If I retired tomorrow, this is the thing I would still want to do with my free time. It’s interesting, impactful and has real-life applications. It gives us a tool to develop more intelligently around habitat impacts—things like cut-off corridors and turtle tunnels.”

SEG funding by the NSF Office of Advanced Cyberinfrastructure is jointly supported by the Division of Biological Infrastructure within the NSF Biosciences Directorate, and by the Division of Information and Intelligent Systems within the NSF Computer and Information Science and Engineering Directorate.

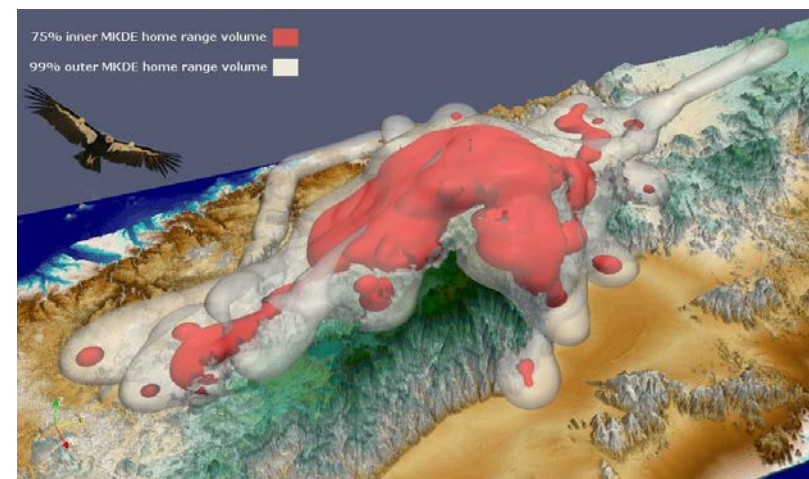


Image map of a condor’s home range. Pink depicts inner home range volume; white depicts outer home range volume. Credit: San Diego Zoo Wildlife Alliance



CAIDA Works with International Team to Inform, Improve Internet Security

A leader and pioneer on the front lines of internet science for nearly three decades, Kimberly “KC” Claffy knows that there have always been security problems on the internet – some baked into the internet architecture. Many technical solutions to these problems have failed to gain traction in the marketplace for fundamentally economic reasons: their complexity and cost, the requirement for pervasive deployment to be effective, combined with a lack of competitive advantage to be an early adopter.

So, as the director of the Center for Applied Internet Data Analysis, or CAIDA, housed at SDSC, Claffy has teamed up with Hervey Allen at the University of Oregon’s Network Startup Resource Center (NSRC), David Clark at MIT’s Computer Science and Artificial Intelligence Laboratory (CSAIL) and Bradley Huffaker at UC San Diego/CAIDA, as well as a team of U.S. and international collaborators on two projects that have been awarded more than \$11 million by the National Science Foundation (NSF). Their aim: to improve internet infrastructure security.

The \$7.8 million for the first project will support the design and prototyping of a distributed but integrated infrastructure to measure internet topology and traffic dynamics, with the intention of improving internet infrastructure security. The team includes more than 30 collaborators altogether – academic and other researchers, industry experts and legal experts in data sharing challenges – who will investigate sustainable production-data acquisition, curation tools, meta-data generation and efficient storage and dissemination to identify security gaps and potential efficiencies.

“All of this is part of stewardship of what is now critical infrastructure, but it is the only critical infrastructure that has no government oversight, and no agency dedicated to measurement in the public interest,” said Claffy. “This project, titled Designing a Global Measurement Infrastructure to Improve Internet Security (GMI3S), serves as a portal to those conversations. It also is an opportunity to transition CAIDA’s work over the last 25 years to serve a growing global research community that can help define, prototype and evaluate instrumentation designed to not only collect data to address grand security challenges, but also to support management, curation and privacy-respecting disclosure of such data.”

According to Clark, the security of the internet is a high priority for the security research community, but that community is greatly hindered by a lack of relevant data.

“Researchers, governments and advocates for society need a more rigorous understanding of the internet ecosystem, a need made more urgent by the rising influence of adversarial actors,” he said. “We cannot secure what we do not understand, and we cannot understand what we do not measure.”

The GMI3S project, however, is not aimed at tackling all internet security problems. “The internet has a layered structure, which, in its simplest form, is a data transport layer on top of which runs a wide range of applications. Our focus is on the internet as a data transport service, and vulnerabilities specific to that layer, such as attacks on internet routing that deflect traffic to bogus destinations, abuses of the domain name system; attacks on the key management system that underpin identity and authentication on the internet; and spoofing of internet addresses to disrupt regions of the internet with untraceable traffic,” explained Huffaker.

Allen added that security challenges at these layers seem to get less publicity than attacks on endpoints (malware, ransomware, etc.), or design features in applications that lead to risky user experiences. “But the challenges at the data transport layer are foundational – they affect the reliable operation of every application that operates over the internet,” he said.

The researchers identified the immediate target for this infrastructure as the research community that measures the internet and tries to improve its security. The intermediate beneficiary will be the operator community and the service providers of the internet while the ultimate beneficiary will be all of society.

“We recognize that better data alone will not improve the security of the internet – this proposal is part of a larger community agenda of research and outreach to industry and policymakers,” said Claffy.

The goal of the second project, Integrated Library for Advancing Network Data Science (ILANDS), which received \$3.5 million in funding over five years, is to understand the internet’s changing character through realistic datasets and longitudinal measurements, as well as new experiments with accessible data for researchers.

The approach of the project is to integrate the research community into the process from the beginning, to align its research goals and optimize NSF’s investment toward achievement of these goals. The approach will have five objectives:

1. shape what data we collect and store;
2. find new users of the infrastructure, especially from underrepresented groups;
3. bring the focused research collaborators together;
4. publish research results and analysis methods and
5. establish a sustainability plan.



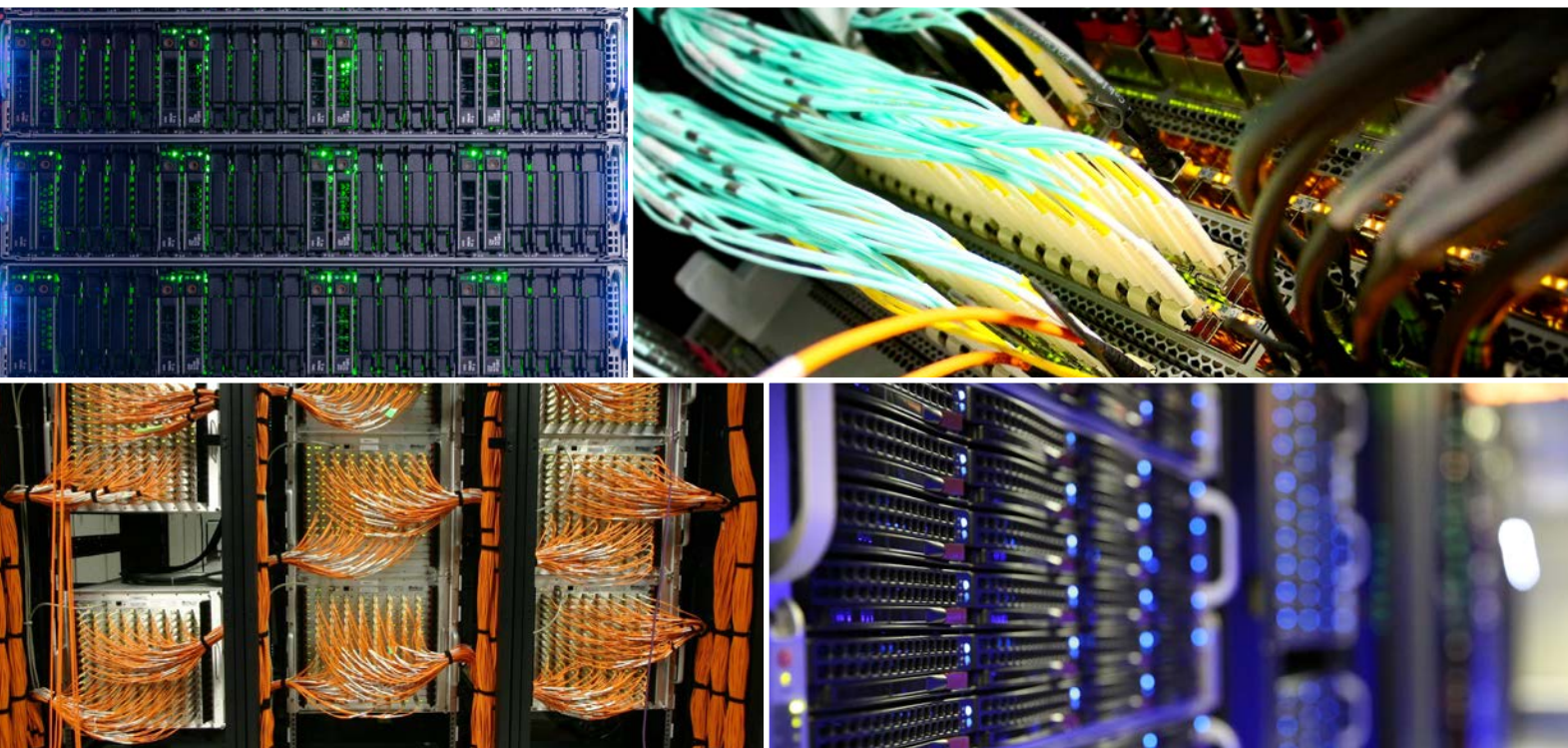
Designing a Global Measurement Infrastructure to Improve Internet Security (GMI3S)

www.caida.org/funding/msri-gmi3s/



Center for Applied Internet Data Analysis (CAIDA)

www.caida.org



Research Data Services in Focus

The SDSC Research Data Services (RDS) Division provides an array of research computing services with a focus on enhancing the advanced computing ecosystem with research data management practices and several data-driven research initiatives.

Over the years, RDS has expanded large-scale and distributed storage offerings for research to more than 20 Petabytes, contributed to and supported the growth of FAIR (Findable, Accessible, Interoperable and Reusable digital objects and data) capacity in the U.S., and provided leadership and worked examples of cloud-based, data-first analysis platforms.

RDS maintains and continually adds to the suite of computing services offered to the national and local UC San Diego communities, including:

- cloud storage and compute platforms and integration services for building private and hybrid clouds,
- data center colocation,
- enterprise networking services,
- research data consulting,
- storage backups,
- high performance, low cost, storage at scale,
- system administration
- and virtualization hosting.

RDS also has built the first storage offering available nationally through Extreme Science and Engineering Discovery Environment (XSEDE) Open Storage Network (OSN) and helped advance electronic notebooks as academic scholarship; researchers are increasingly using Jupyter Notebooks to communicate research methods which allows for showing how data is transformed and how research can be reproduced.

Other ways in which RDS has made a difference:

Operations' COVID-19 Response

As UC San Diego restricted campus access in response to the COVID-19 pandemic, the SDSC operations team continued to provide 24/7 on-site support with new social distancing protocols, temporarily reduced shifts, deep-cleaning and greater sanitation protocols. Operations processed 2,800 support tickets providing exceptional remote hands support, facilitating visits in accordance with campus health policy and other requests for staff and researchers amidst the changing dynamics of building access.

West Big Data Innovation Hub

Last year, the West Hub partnered with the Border Solutions Alliance for the COVID-19 Data Challenge to broaden awareness of pandemic risk levels among border communities. This year, the West Hub has focused on a new curriculum for government and nonprofits to jump start their work with common data sets and data science techniques. The "Hub" is extending its cyberinfrastructure, cloud computing and data science curriculum to educators at Hispanic Serving Institutions through the HSISTEM Hub, as well as The Carpentries curriculum to Tribal Colleges and Universities (TCUs) via a partnership with the American Indian Higher Education Consortium (AIHEC).

GO FAIR U.S

GO FAIR U.S. is headquartered at SDSC and seeks to disseminate, support and coordinate decisions and activities related to FAIR infrastructure, pilots, training and policy in the U.S. Over the past year, the GO FAIR U.S. team has held several events with topics spanning from FAIR Basics, FAIR and Research Software, to FAIR Workflows and more.

Open Storage Network

The OSN team created a distributed platform for data sharing and data reuse into a production service with robust and scalable hardware, software, policy and practices. The team represents a partnership between the nation's Advanced Computing centers and leading innovators including NCSA, PSC, MGHPCC, JHU and others. Rallying around the common goal of low-cost, high performance, distributed storage, the team devised a technical methodology, processes and approach for the most efficient operations possible, concentrated in the OSN Command Center. The five-site OSN deployment is improving and expanding through engagement with sites investing in deploying new OSN pods, such as the new pod coming online at Alabama A&M and a virtual pod at University of Maine.



Making Way for AI

Through its EarthCube Office (ECO) initiative, RDS is working toward models that allow for peer-reviewed notebooks, versus static publications, and searchable notebook directories as an important new resource for data-driven science. This year will bring pilots to define infrastructure for FAIR Digital Objects, a new approach for making data AI-ready, and research related to artificial intelligence (AI) and reproducibility, the intersection of FAIR and AI, and AI-readiness more broadly.

RDS was active in several sponsored projects and secured new funding including the ECO Office and West Hub. For example, RDS Chief Strategist Melissa Cragin and Division Director Christine Kirkpatrick were recently awarded a grant by the National Science Foundation to fund a GO FAIR symposium to seek input from a diverse set of experts and representatives on the most promising areas and pathways for expanding implementation of FAIR principles in research practice and services. Building off of the West Hub, Kirkpatrick is a partner on a new award from the NSF's AccelNet program with SDSC's Information Systems Laboratory Director Ilya Zaslavsky and Sam Fernald of New Mexico State University to better understand the issues around transboundary aquifers.

The ECO also recently re-designed GeoCODES which has significant impact for data and tool discovery in the EarthCube community, and the reach of this work will continue beyond EarthCube to benefit the broader geosciences community.



Programmatic Approach to Increasing Revenue Potential of Software Projects

The San Diego Supercomputer Center (SDSC) Rev-Up program is an innovative resource aimed at revenue generation, project sustainability and community building for internal software initiatives interested in non-commercialization-based revenue.

Teams participating in Rev-Up receive one-on-one time with sustainability experts and may be eligible to receive development support from skilled engineers and makers. They also receive training, coaching and support as they apply revenue-generating strategies through a proven process that involves promotional strategies, implementation guidance, performance evaluation and monitoring analytics.

“The guiding principle of the project is to institutionalize and make revenue-based thinking pervasive within the research enterprises of those SDSC investigators who are interested in developing parallel or replacement funding streams for grant-funded software and service products,” said Rev-Up Founder and Director Michael Zentner, who leads the Sustainable Scientific Software Division at SDSC.

The multi-goal project includes:

- Establishing awareness of principles associated with revenue-generating projects over the entire project lifecycle, using both instructional and consulting modes;
- Helping specific projects in “deep-dive” format to scale and sell the resulting products;
- Serving as a model for non-commercialization revenue generation activities that may be adopted or attended by other campus organizations and possibly other institutions;
- Creating a suite of products that earn sufficient revenue to fund ongoing development and operations; and
- Creating a culture of SDSC investigators focusing on both good research and future revenue.

Zentner explained that over a two-year period, the pilot program will be examined for success based on markers like these:

- Project enrollment in the training sessions;
- Assessment of participating project progress;
- Bringing one new participating project to a stage of readiness for sales; and
- Success in assisting established projects toward increasing sales, including CAIDA, Open Topography, WIFIRE and HUBzero.

“During the third year, generating revenue from one of the early, less established projects will be an additional indicator of success,” explained Zentner.

Rev-Up Assistant Director Claire Stirm noted that at the core of SDSC culture is the center’s leadership and strength in advanced computation in all aspects of “big data” and its support of a wide variety of domains.

“The SDSC Rev-Up program will offer SDSC projects an opportunity to learn sustainability strategies and techniques from Rev-Up instructors at a one-week workshop called On-Ramp,” said Stirm, adding that the workshop builds off the successful Focus Week curriculum developed by Zentner, Juliana Casavan and Nancy Maron, supported by the Science Gateways Community Institute (SGCI).

“On-Ramp expands this foundation curriculum to incorporate more revenue-focused curriculum and hands-on market validation techniques,” said Zentner. “We’ve taken what we learned from hundreds of participants in the SGCI program and paired it with consultation and development services to help accelerate projects that are taking revenue generation seriously.”

SDSC software projects interested in exploring non-commercialization-based revenue strategies can receive coaching from the Rev-Up team for market validation and competitive analysis in their user community, as well as to examine adjacencies for their technologies that may not have been the original intended audience. Additionally, teams that want to apply what they discovered during the coaching engagement can receive development support to enhance their software or receive time from a sales expert to establish a sales funnel for the project.

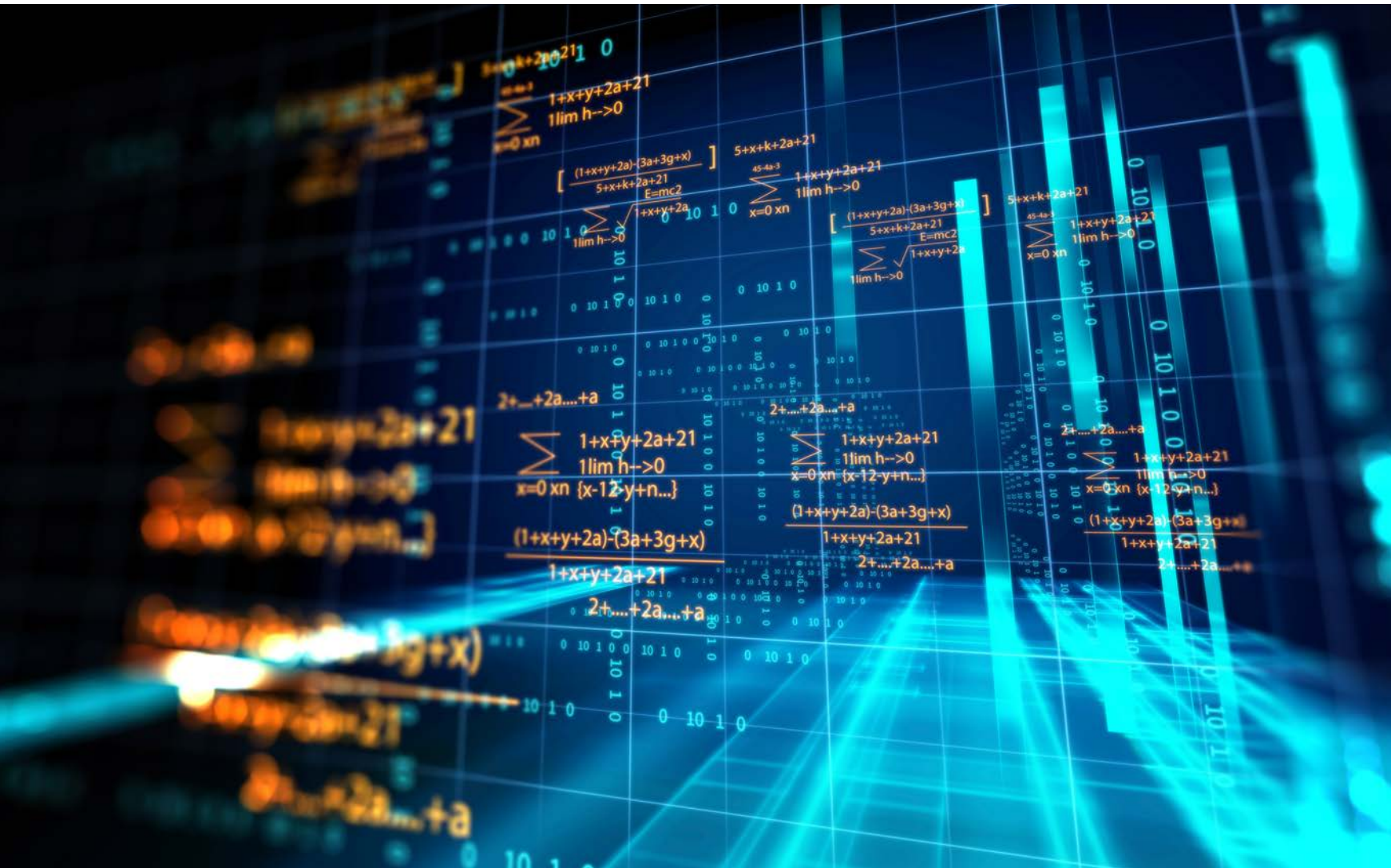
According to Stirm, the Rev-Up team is dynamic, offering a wealth of experience to the SDSC community. Along with Zentner and Stirm, Casavan and Maron oversee the sustainability training, and they also lead sustainability coaching with Mike Shepard. Gene Eberhardt manages sales.



Rev-Up Program
www.sdsc.edu/services/rev-up.html



Sustainable Scientific Software (S3)
www.sdsc.edu/services/s3.html



North, South, East or West, It's Data Collaboration

SDSC's semiannual data technology symposium, Data West, recently addressed the global debate underway over the impact of data – and digital technology more generally – on “open science” and “open data.” Coupled with the connection between today's physical and digital infrastructure design, strategy and implementation, these topics take on special significance as the nation deliberates the largest infrastructure initiative in a generation, as data science continues to expand its influence throughout our economy.

Established in 2016 by SDSC's Center for Large Scale Data Systems Research (CLDS), and chaired by SDSC's James Short, Harvard Business School's Professor Lynda Applegate and MIT's Dr. Mel Horwitch, Data West brings together industry, government and university researchers to envision new ways to design, value and leverage extreme data environments for scientific, business and social benefit.



The data economy, of course, will not be built on data systems alone. “Data is changing our business, social and economic landscape, potentially concentrating market power in a few industry sectors, and thereby diminishing the relative value in others,” said Short. “For individuals, a relentless focus on data raises concerns about trust, privacy and security, as well as equitable distribution of gains from data.”

Another essential element in the data economy is trust. Trust deficits can unravel data markets and undermine civic and social cohesion. “Trust in data, and trusted stakeholders in data partnerships, are a necessary precondition for building our future business platforms and partnerships to sustain innovation and economic growth,” said Applegate.

“Indeed, the growing prominence of data is changing how society views and even defines physical infrastructure, and a better understanding of this relationship is essential for effective infrastructure-related policy, business decision making, and innovation,” said Horwitch.

In its sixth year, Data West has garnered a reputation for inclusiveness, gathering diverse thought leaders, technologists, data scientists and senior general managers to address some of industry's most complex problems. A key aim is to provoke, advance and improve policies and decisions made in the real world.

The success of Data West inspired its complement, Data East in 2019. This expansion allowed for a greater focus on public sector, government and non-profit organizations, including international NGOs such as the UN and on topics including public sector investment in technology, business-government partnerships and evolving public-private partnerships and regulation.

Where are the twin data events headed in the future? According to founding corporate partners and industry co-chairs Raj Patil of AEEC and Steve Orrin of Intel, “The diversity of Data West/ Data East participants, their talent, discussion topics and the positive network effects to educate and cross-pollinate bold ideas more than justify our initial support and our continued enthusiastic involvement with these events. We appreciate the relentless efforts of the conference organizers to make Data West and Data East welcoming platforms for industry case study discussions, plenary panels with government CIOs, industry CEOs, CTOs and CDOs, research institutions and student organizations.”



Data West
www.datawest.org



Data East
www.dataeast.org



Center for Large Scale
Data Systems Research
www.clds.info



Sherlock Continues to Find Solutions for Customers

Over the past year, the Sherlock Division has continued to develop and expand its portfolio of offerings and services to customers, providing them with an array of managed services that meet their secure computing and data management needs. In so doing, the Sherlock team has expanded its partnerships and collaborations.

UC Partners

University of California, Irvine (UCI): The Sherlock Division has partnered with UCI to offer Controlled Unclassified Information (CUI) and Cybersecurity Maturity Model Certification (CMMC) protected data services across UCI's campus. The initial set of projects are supported within the UCI Center for Evidence-Based Corrections and specifically meet the California Department of Justice (CADOJ) requirements.

University of California, Los Angeles (UCLA): Sherlock has partnered with UCLA DGIT to provide an environment in Sherlock Cloud to host the UCLA Screening and Treatment for Anxiety & Depression (STAND) suite of applications. These applications are managed in Sherlock's secure HIPAA-compliant enclave in AWS.

University of California San Diego's (UC San Diego) School of Pharmacy: The Sherlock Division supports UC San Diego's School of Pharmacy by providing a secure enclave within its FISMA-compliant environment on-prem at SDSC for the storage and analysis of Medicare fee-for-service claims data. The research project primarily focuses on pharmacoepidemiology, pharmacoconomics and pharmaceutical outcomes research in a random sample of Medicare beneficiaries.

TemPredict: Sherlock has partnered with researchers from the University of California, San Francisco, UC San Diego, Massachusetts Institute of Technology (MIT) Lincoln Laboratory, the U.S. Army and the U.S. Navy to deploy a HIPAA-compliant platform in AWS to collect physiological data from frontline healthcare workers and the general population. The aim of the study is to complete antibody testing for 10,000 participants and to provide additional support for algorithm development and testing in real-world settings. Sherlock provides secure storage, data processing leveraging time series databases and analytical tools for the TemPredict researcher community.

External Partners

Yale University: The Sherlock Division supports Yale researchers analyzing and running statistical analyses on CUI originating from the U.S. Department of Defense. Yale uses the Division's Sherlock Cloud CUI-compliant environment to process, store and analyze this sensitive data. As the partnership matures, the Sherlock Division hopes to further develop the relationship with Yale by broadening the university's protected data portfolio within Sherlock Cloud to include other forms of sensitive data that requires specialized security and systems protections.

Arizona State University (ASU): The Sherlock Division works with ASU to ensure that CUI stemming from projects with federal agencies are securely protected within its Sherlock Cloud CUI-compliant environment. The initial collaboration includes supporting projects managed out of the ASU Global Security Initiative. While Sherlock is initially focusing on supporting ASU's CUI-specific projects, it is likely that this partnership will extend to include data requiring protections under the CMMC model.

Wilco Labz (Wilco Source): The Sherlock Division developed a partnership with Wilco Source, a Salesforce implementation-focused technology company, to host its product Wilco Docx within Sherlock's secure enclave in Google Cloud Platform (GCP). Wilco is Sherlock's second industry partnership and its first deployment in GCP.

Partnership with the University of California Office of the President Risk Services

The Sherlock Division partnered with industry partner Kwartile and UCOP IT to help migrate the UCOP Risk Services' HIPAA-compliant, Hadoop-based Risk Data Management System (RDMS) from its on-prem instance at SDSC (RDMS 1.0) to AWS (RDMS 2.0). This year-long effort helped Risk Services leapfrog from a dated platform to a more modern, economical and secure cloud-based, serverless data lake platform.

Cybersecurity Maturity Model Certification Partnership with UC San Diego ITS

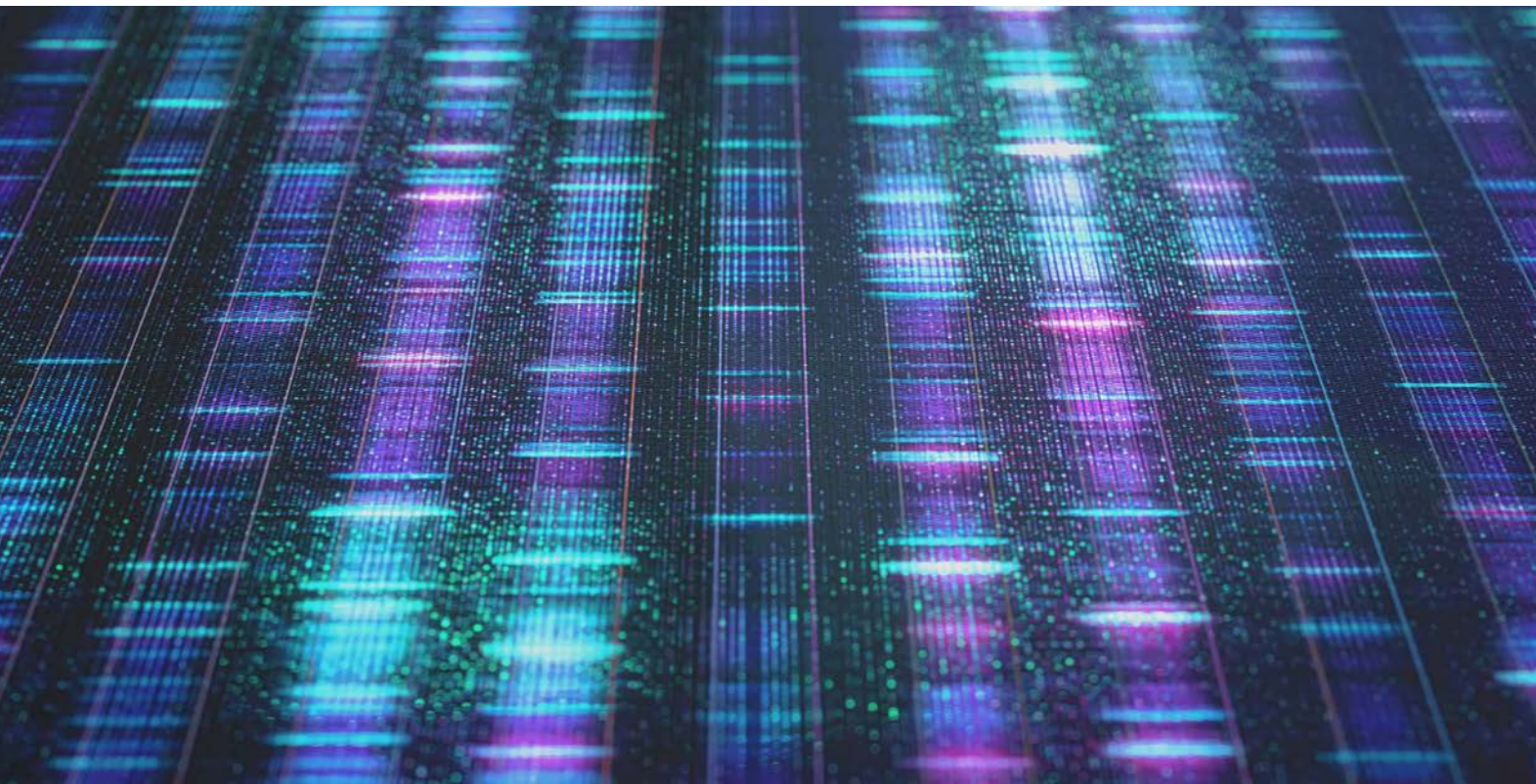
The Sherlock Division, in partnership with UC San Diego Information Technology Services (ITS) and external partners Microsoft, Summit7 and ePlus, has been working to build a Cybersecurity Maturity Model Certification (CMMC) Level 3 environment that will be leveraged by UC San Diego, campuses across the University of California, and academic and research institutions nationwide. Sherlock offers a secure enclave that spans across on-prem, Microsoft Azure Gov Cloud and Office365 GCC High. The efforts to build this environment are underway, and Sherlock anticipates deploying the CMMC Level 3 environment in the coming year.

Cybersecurity Maturity Model Certification (CMMC)

CMMC is a unifying standard for the implementation of cybersecurity across the Defense Industrial Base (DIB). The CMMC framework includes a comprehensive and scalable certification element to verify the implementation of processes and practices associated with the achievement of a cybersecurity maturity level. CMMC is designed to provide increased assurance to the Department that a DIB company can adequately protect sensitive unclassified information, accounting for information flow down to subcontractors in a multi-tier supply chain.

The Department of Defense (DoD) developed CMMC to protect defense contractors from cybersecurity incidents. As such, CMMC identifies a comprehensive set of cybersecurity standards; this set of standards incorporates the CUI NIST 800-171 security controls, and adds additional practices and processes specific to CMMC. CMMC certification is enforced by government entities, and organizations seeking contract awards that require CMMC certification must endure a rigorous certification process that is conducted by third-party auditors.

Sherlock and UC San Diego ITS are partnering with external partners Microsoft, Summit7 and ePlus to build a comprehensive CMMC-certified, private cloud capability that will be leveraged by UC San Diego, campuses across the UC and academic and research institutions nationwide. This CMMC Level-3 compliant environment will comprise of three components: (1) an Azure Government component that will be used for compute and data workloads, (2) an M365 GCC High tenant that will be used for the Office suite of collaborative applications and (3) an on-premise enclave that will support use cases that are better suited operating within the local datacenter. These three integrated tenants together will provide services that will meet a wide range of research use cases. Additionally, the Sherlock and ITS team plan to work with UC San Diego to ensure that research labs campus-wide are compliant with CMMC Level-3. The entire pipeline – from the researchers who generate the CUI on the end-point to the computing environment that will run the analysis – will be configured to be CMMC Level-3 compliant.



On the Frontier of Genomics Research

Next Generation Sequencing technology revolutionized genomics research more than a decade ago. With it came a deluge of raw sequence data that needed to be “assembled” into complete genomes then further analyzed. SDSC arrived early on the genome-analysis scene when in 2012 it commissioned the Gordon supercomputer, designed specifically for data intensive computing tasks such as genome assembly and analysis.

Researchers at UC San Diego and beyond sought out SDSC for assistance with their genomics studies. An early project was conducted in collaboration with Johnson & Johnson’s Janssen R&D division to rapidly analyze 500 whole human genome sequences for a study related to rheumatoid arthritis treatments.

SDSC has also supported numerous other companies involved in genomics research, as well as non-profits including the J. Craig Venter Institute, the La Jolla Institute for Immunology, the Scripps Research Institute and the Salk Institute.

Early genomics studies involved relatively small cohorts of individuals; today multiple programs are focused on population-scale genomics such as the United Kingdom BioBank (UKBB) study, which is amassing Whole Genome Sequence (WGS) data and other data on half a million individuals. SDSC is working with Janssen R&D to implement a high-performance informatics environment capable of supporting targeted research on this massive genomics dataset.

At the other end of the spectrum, application-specific technology such as the Illumina DRAGEN system is providing for quick turnaround analysis of individual samples for clinical and laboratory uses. For example, SDSC has supported UC San Diego’s Department of Health Sciences in rapid turnaround testing of both nasal samples and wastewater for SARS-CoV-2 detection.

FROM CAPABILITY TO RESULTS

SDSC’s Comet and Expanse supercomputers replaced Gordon and were pressed into service for genome analysis. The Triton Shared Computing Cluster (TSCC) has also seen steady growth from genomics research groups. A Campus Cyberinfrastructure grant from the National Science Foundation funded the BioBurst partition of TSCC, providing an Illumina DRAGEN sequencing node and DDN Burst Buffer system, targeted specifically at genome analysis.

SDSC has fielded new storage services such as the Universal Scale Storage (USS) service, which provides researchers with cost-effective, high-performance storage options for petabyte-scale genomic data storage. Security protocols were implemented enabling researchers to work with controlled DNA data such as the National Institute of Health’s (NIH) Database of Genotypes and Phenotypes (dbGaP).

This panoply of computing and storage capabilities and accompanying expertise has afforded SDSC the opportunity to support several UC San Diego research groups:

The Alexandrov Lab

Ludmil Alexandrov’s group focuses its research on developing and leveraging novel machine-learning approaches to elucidate the basic molecular mechanism underlying cancer development and cancer progression, in order to improve cancer treatment and prevention.

SDSC implemented a substantial augmentation of the TSCC cluster, including 36 high-performance compute nodes, a dedicated two petabyte Lustre parallel file system and both Universal Scale Storage and “Project Condo” storage. Another 25 compute and Graphics Processing Unit (GPU) nodes and two petabytes of Lustre storage will be added in 2022 for an even more powerful computational genomics platform.

The Knight Lab

Rob Knight, director of the Center for Microbiome Innovation, uses and develops state-of-the-art computational and experimental techniques to ask fundamental questions about the evolution of the composition of biomolecules, genomes and communities in different ecosystems, including the complex microbial ecosystems of the human body. A professor of Pediatrics, Bioengineering and Computer Science & Engineering, Knight also leads UC San Diego’s SARS-CoV-2 wastewater screening program.

SDSC supports this work with efforts such as porting the popular UniFrac microbiome tool to GPUs, resulting in a three-orders-of-magnitude speed-up. Recently, the SDSC team achieved a milestone with release of a GPU-enabled UniFrac

package in Bioconda, which makes these speed-ups accessible to a larger user community.

Metabolics Workbench

SDSC has a long history as a national and international leader in multi-omics research. In addition to being a hub for the computational horsepower behind genomics research, world-class researchers such as Shankar Subramaniam and Eoin Fahey connect in proteomics, lipidomics and metabolomics. The researchers also maintain RefMet, a reference standard for identifying and describing known metabolites that has enabled cross-study research.

In partnership with SDSC, Subramaniam and Fahey host the Metabolomics Workbench, the NIH-funded, national repository for metabolomics studies.

Sebat Lab

The lab of Professor Jonathan Sebat relies on SDSC’s supercomputers to support research applying computational genomics methods for understanding psychiatric disorders, including autism spectrum disorders and schizophrenia. The overarching focus of the Sebat Lab is to identify rare variants that lead to the observed phenotype and share the findings with physicians.

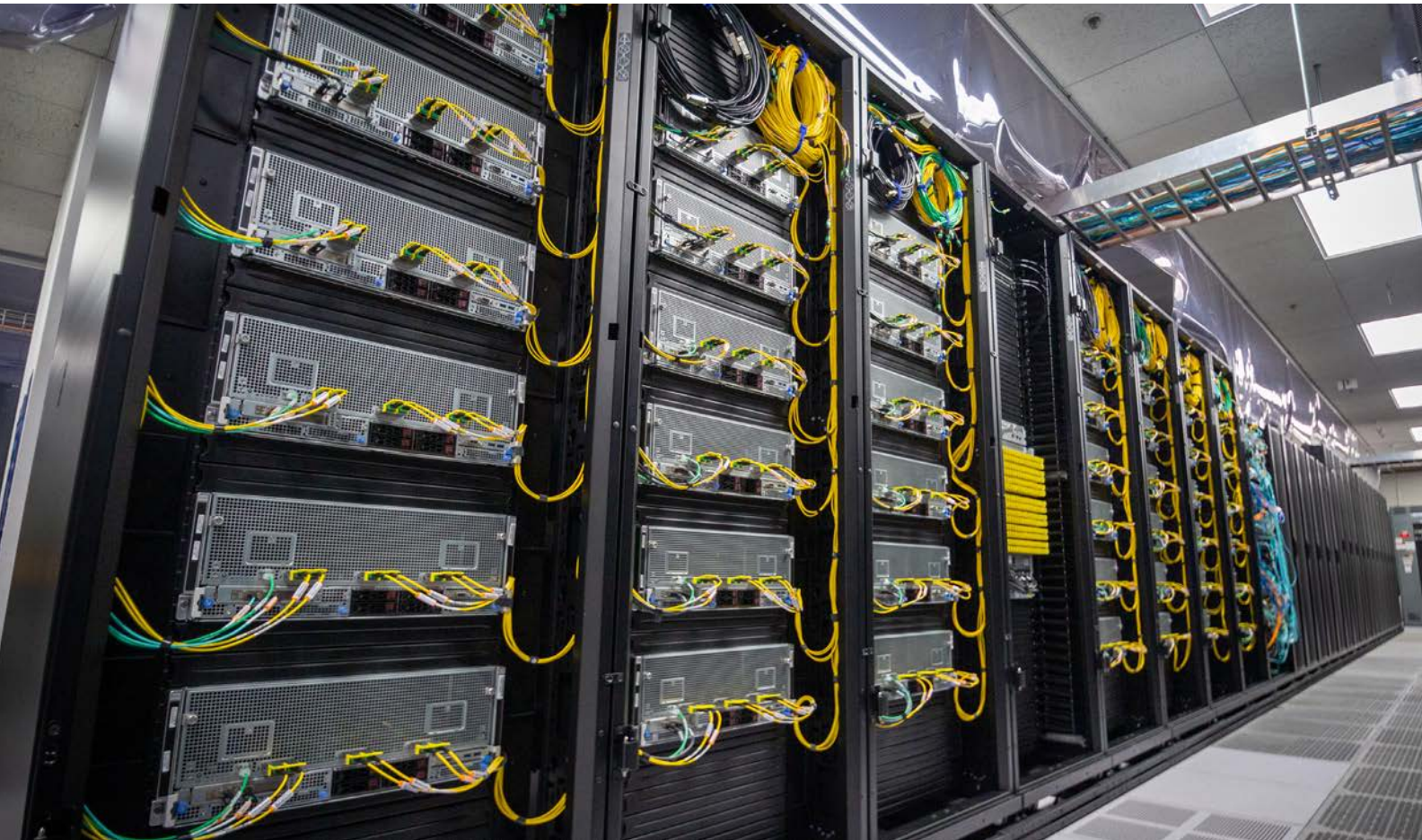
SDSC also has provided bioinformatics support for the lab to develop workflows to process and analyze thousands of high-coverage, whole-genome sequence data sets from ASD patients recruited from San Diego’s Rady Children’s Hospital.

Yeo Lab

The primary research interest of Professor Eugene Yeo’s lab is to understand the importance of RNA processing and the roles that RNA binding proteins (RBPs) play in disease development.

The lab has long utilized TSCC as a computational resource for application of computational methods to study RNA biology and therapeutics.

With the addition of significant new computing capabilities such as Expanse for high-throughput and diverse workloads; Voyager, focused on high-performance computing for experimental artificial intelligence and machine learning; planned upgrades to TSCC; dedicated small cluster and storage support for genomics labs; and commercial cloud integration and bioinformatics expertise, SDSC expects to be an integral part of genomics research on the Torrey Pines Mesa in 2022 and beyond.



SDSC, Habana, Supermicro Team Up to Operate Voyager

The San Diego Supercomputer Center at UC San Diego, a leader and pioneer in high-performance and data-intensive computing, selected the Intel company's Habana Labs artificial intelligence (AI) training and inference accelerators to provide high-performance, high-efficiency AI compute for its experimental Voyager supercomputer – a first-of-its-kind experimental system in the National Science Foundation's resource portfolio, complementing systems focused on capacity production computing, such as SDSC's Expanse system.

Newly installed at SDSC, Voyager will be dedicated to advancing AI research across a range of science and engineering domains. The system build-out, as well as ongoing community support and operations, is funded by a five-year, multimillion-dollar grant from the National Science Foundation (NSF), split between hardware acquisition and operations.

As part of SDSC's mission to support science and engineering and in line with NSF guidelines, SDSC required a cost effective, yet powerful system to accelerate the development of AI algorithms. SDSC chose the combination of Habana® AI processors and Intel® Xeon® Scalable CPUs in Supermicro servers. The Voyager system contains over 42 Supermicro X12 Gaudi® AI Training Systems with 336 Habana Gaudi processors—designed for scaling large supercomputer training applications—and 16 Habana Goya processors to power AI inference models.

We are honored that Habana's AI processors have been selected to power the AI workloads that will run on San Diego Supercomputer Center's Voyager supercomputer," said Eitan Medina, chief business officer at Habana. "This implementation of our Gaudi and Goya products showcases how top academic institutions like SDSC can harness efficiency and performance to effectively address the growing demands of AI research workloads."

Habana Gaudi AI training and Goya inference processors are architected to drive performance and efficiency in AI operations. They will provide data scientists and researchers having access to Voyager with flexibility to customize models with programmable Tensor Processor Cores and kernel libraries, and ease implementation with Habana's SynapseAI® Software platform, which supports popular machine learning frameworks and AI models for applications such as vision, natural language processing and recommendation systems.

Amit Majumdar is the principal investigator (PI), with co-PIs that include SDSC's Mai Nguyen and Robert Sinkovits, as well as Computational Chemist and Biophysicist Rommie Amaro from the Department of Chemistry and Biochemistry, and Physicist Javier Duarte from the Department of Physics in UC San Diego's Division of Physical Sciences. SDSC's Shawn Strande is the project manager.

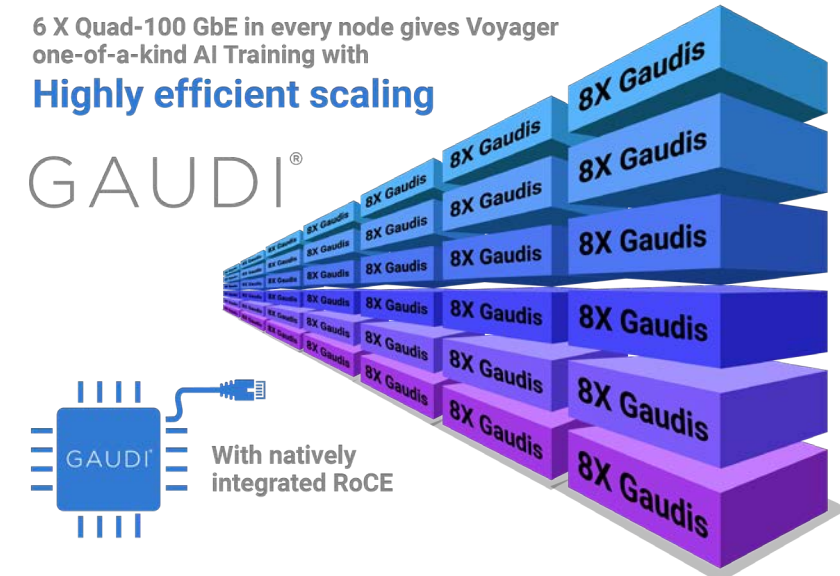
"With innovative solutions optimized for deep learning operations and AI workloads, Habana accelerators are excellent choices to power Voyager's forthcoming AI research," said Majumdar, who leads SDSC's Data Enabled Scientific Computing Division. "We look forward to partnering with

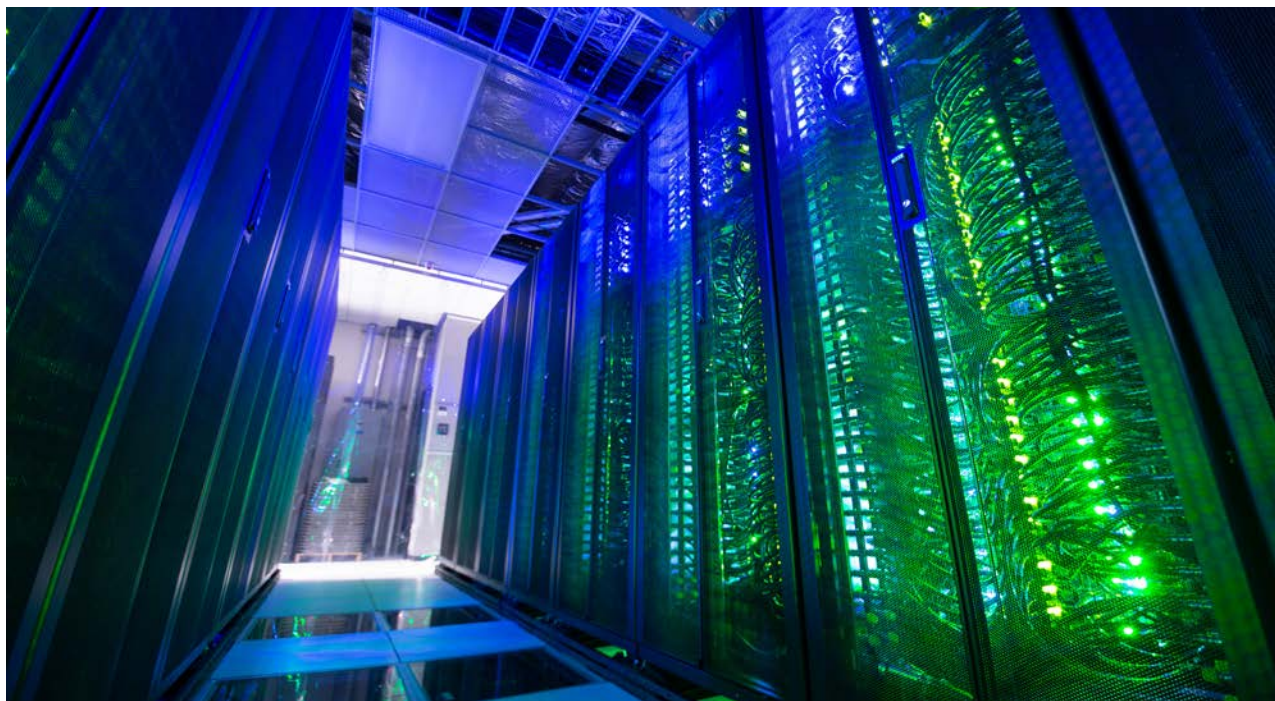
Habana, Intel and Supermicro to bring this uniquely efficient class of compute capabilities to the Voyager program, giving academic researchers access to one of the most capable and innovative AI-focused systems available today."

Technology partner Supermicro, a global leader in enterprise computing, storage, networking solutions and green computing technology, will provide the Habana-based AI systems to SDSC for Voyager. These include the Supermicro X12 Gaudi AI Training System, which includes the newly announced 3rd Gen Intel Xeon Scalable Processors and the Supermicro SuperServer 4029GP-T featuring eight Goya HL-100 PCIe cards for AI inference, paired with dual-socket Intel® Xeon® Scalable Processors.

Combining Supermicro's advanced application-optimized server and storage hardware with Habana's AI training and inference products is an outstanding solution for SDSC's multi-year Voyager AI project," said Ray Pang, vice president of technology and business enablement at Supermicro. "We continue to work closely with leading innovators to deliver solutions for computationally intensive projects worldwide at leading research and HPC environments for science and medical discovery, compute and leading-edge AI solutions."

According to Vice President and General Manager of Intel's High-Performance Computing Group Trish Damkroger, the level of performance and efficiency that Voyager will require is precisely what Intel architectures are designed for. "Our Xeon Scalable processors coupled with Habana AI accelerators will ensure Voyager's users have the HPC and AI capabilities they need to power their game-changing research," she said.





Core Scientific Plexus™ Software Extends Reach of SDSC’s Expanse Supercomputer

SDSC is a leader in advanced computation and data analytics, and it is growing its industrial engagement through new partnerships such as the one it forged this year with Core Scientific—a leading infrastructure and software solutions provider for artificial intelligence and blockchain.

Working under the leadership of SDSC’s former Director Michael Norman and Core Scientific CEO Kevin Turner, former Chief Operating Officer of Microsoft, SDSC and Core Scientific integrated the Core Scientific Plexus™ AI software stack with the Expanse petascale supercomputer, funded by the National Science Foundation.

A key innovation of Expanse is its ability to support composable systems, which can be described as the integration of computing elements such as a combination of central processing unit (CPU), graphics processing unit (GPU) and other resources into scientific workflows that may include data acquisition and processing, machine learning and traditional simulation. While the main Expanse system is funded by the NSF, SDSC recently installed additional capacity to support industrial workloads.

Exposing SDSC’s Expanse supercomputer platform via Core Scientific’s Plexus software stack provides users with a consumption-based high-performance computing (HPC) model that not only supports on-premise infrastructure, but also has

the ability to run HPC workloads in supercomputer centers as well as in any of the four major public cloud providers—all from a single pane of glass.

“We are excited to make the power of the Expanse supercomputer available to industry researchers and other customers who need state-of-the-art supercomputer capabilities,” said Ron Hawkins, SDSC director of Industry Partnerships. “Core Scientific’s Plexus software stack provides end-customers access to our infrastructure in an intuitive, cloud-like point-and-click experience, which is a potential game changer for supercomputer centers.”

Expanse, which entered service in December 2020, consists of nearly 800 AMD EPYC (Rome) based compute nodes with a 12-petabyte parallel file system and Mellanox InfiniBand HDR interconnect, with more than 100,000 cores available to the most demanding HPC workloads.

“SDSC’s Expanse is a unique and innovative supercomputer that supports a wide range of scientific workloads,” said Core Scientific Chief Product Officer of Artificial Intelligence Ian Ferreira. “Being able to democratize this infrastructure via our Plexus software stack will enable researchers and scientists to seamlessly work between CPU and GPU bound class of problems, all via a single, easy-to-use interface.”



CloudBank Encompasses Both University and Industry Partnerships

The public cloud accelerates research and education in at least two important ways: it serves “on demand” compute resources without the associated facilities and maintenance cost, and it provides a marketplace of tools, technologies and ideas that allow for cross-pollination of methodology across teams and fields as well as reproduction of results. This is accomplished through a “pay for what you use” utility model of computation, democratizing access to resources at a scale that formerly would have required an institution-wide investment.

With this technological advance, however, comes added complexity that has slowed adoption of these new resources. To accelerate this adoption process, the NSF issued a Computer and Information Science and Engineering (CISE) Directorate solicitation in December 2018 for a five-year pilot project to establish an entity that would serve as an interface between the CISE research and education community and public cloud computing providers. SDSC and Information Technology Services Division at UC San Diego, the eScience Institute at the University of Washington, the Division of Computing, Data Science, and Society at the University of California Berkeley, and Strategic Blue were the successful respondents to this solicitation and created CloudBank, an entity that delivers managed services that reduce the friction in accessing public clouds and provides curated, researcher-focused resources and support.

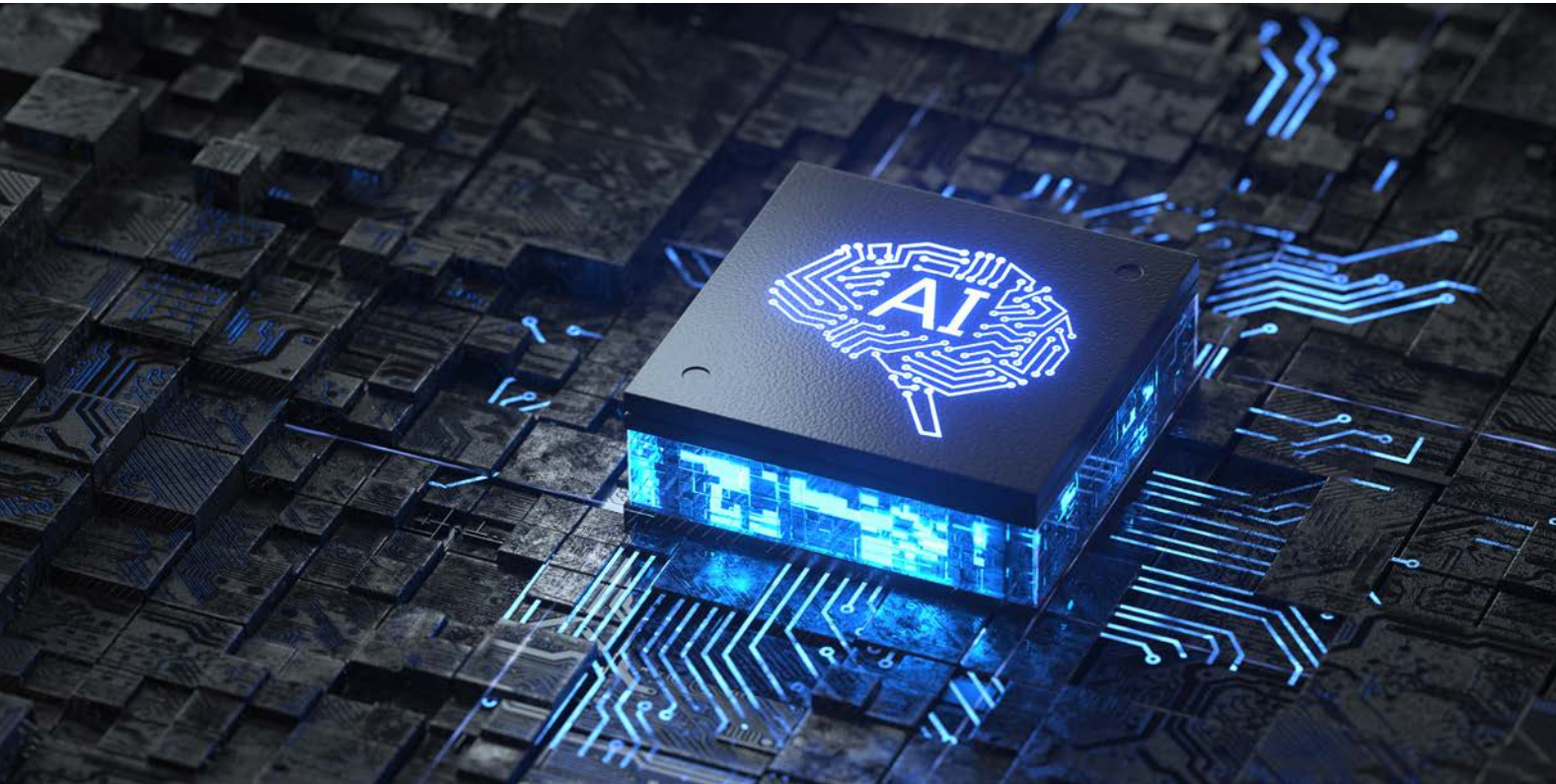
In preparing the proposal, CloudBank surveyed computer science researchers at the partner institutions about the challenges they encountered when using public cloud resources. Their responses revealed three main categories of pain points:

- getting cloud access,
- learning foundational technological concepts and
- billing and managing spend.

CloudBank is focused on removing these sources of friction.

CloudBank went into production in August 2020 and currently supports Amazon Web Services, Google Cloud Platform, IBM Cloud and Microsoft Azure. The key benefits of CloudBank are:

- no indirect costs,
- one-on-one consultation for technological feasibility,
- facilitated cloud access and account management,
- access to multiple cloud platforms with a pay-per-use model,
- spend monitoring and resource usage optimization,
- curated training materials targeted for researchers and educators,
- tools for running scalable classes in the cloud and
- dedicated help desk support.



SDSC Plays Role in NSF's New ICICLE Institute

While some areas of science know first-hand how cool artificial intelligence (AI) is, not all domains have had access to the technology. But AI is on the verge of becoming accessible to more researchers with the AI Institute for Intelligent Cyberinfrastructure with Computational Learning in the Environment, or ICICLE. The new institute will focus on next-generation intelligent cyberinfrastructure that makes using AI as easy as plugging an appliance into an electrical outlet.

One of 11 new initiatives collectively funded with \$220 million from the National Science Foundation (NSF), ICICLE, with its \$20-million share, is envisioned to result in AI-based technologies that bring about a range of advances—from helping better understand movement and behavior of animals, to creating solutions to improve agriculture and food supply chains, and supporting underrepresented students in kindergarten through post-doctoral STEM education to improve equity in AI research. It will advance “plug-and-play” AI across a wide range of domains, promoting the democratization of AI.

ICICLE specifically will develop intelligent cyberinfrastructure (CI) with transparent and high-performance execution on diverse and heterogeneous environments; for example, animals

in forests, crops in farmlands or foods stored in warehouses. It brings together a multidisciplinary team of scientists and engineers, led by the Ohio State University in partnership with the San Diego Supercomputer Center (SDSC), University of California San Diego, University of California Davis, Case Western Reserve University, IC-FOODS, Indiana University, Iowa State University, Ohio Supercomputer Center, Rensselaer Polytechnic Institute, Texas Advanced Computing Center, University of Utah, University of Delaware and University of Wisconsin-Madison.

Initially, ICICLE is tasked with addressing complex societal challenges in the areas of smart foodsheds, precision agriculture and animal ecology, which will drive the institute’s research and workforce development agenda. This includes:

- empowering plug-and-play AI by advancing five foundational areas—knowledge graphs, model commons, adaptive AI, federated learning and conversational AI;
- providing a robust cyberinfrastructure capable of propelling AI-driven science (CI4AI), solving the challenges arising from heterogeneity in applications, software and hardware, and disseminating the CI4AI innovations to use-inspired science domains;

- creating new AI techniques for the adaptation/ optimization of various CI components (AI4CI), enabling a virtuous cycle to advance both AI and CI;
- developing novel techniques to address cross-cutting issues including privacy, accountability and data integrity for CI and AI; and
- providing a geographically distributed and heterogeneous system consisting of software, data and applications orchestrated by a common application programming interface and execution middleware. Overall, ICICLE will shore up next-generation AI cyberinfrastructure.

“It will transform today’s AI landscape,” said Division Director of Data Enabled Scientific Computing at SDSC Amit Majumdar, principal investigator (PI) of the SDSC sub-award, who is charged with CI development and integration with research facilities. “SDSC and UC San Diego researchers are involved in various research topics including education and workforce development, knowledge graphs, adaptive AI, high-performance model training and more.”

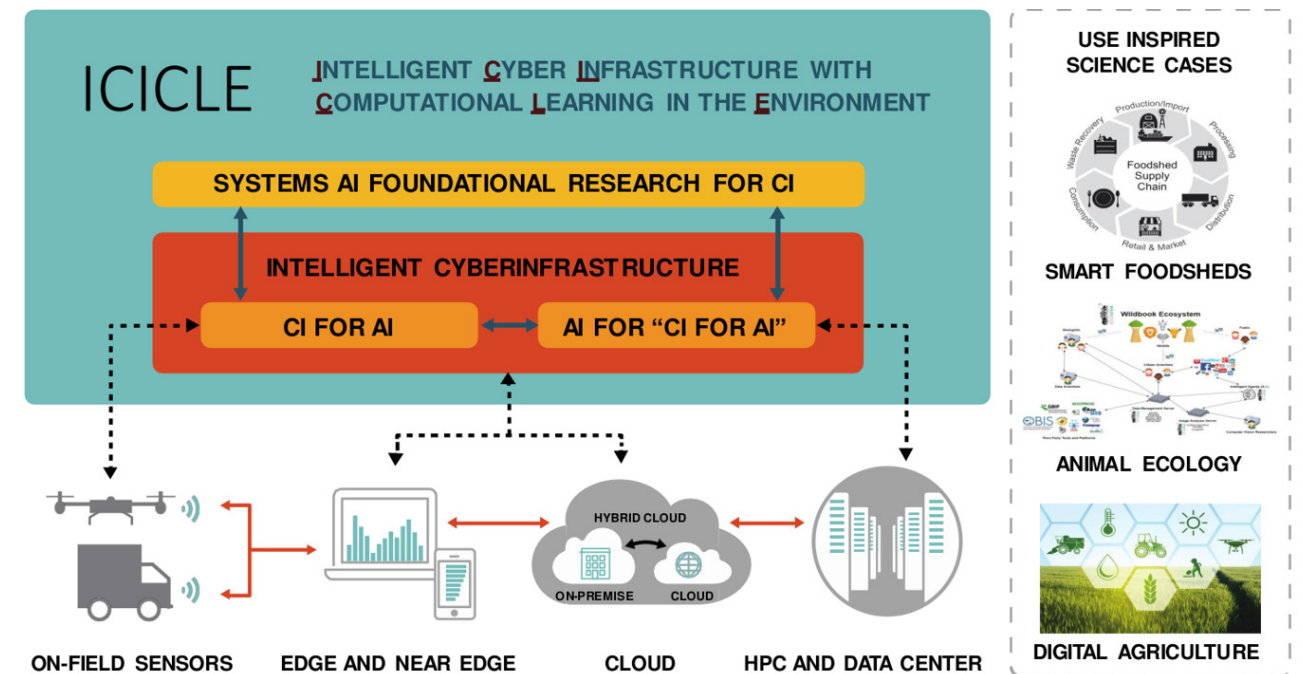
Additionally, ICICLE’s advanced and integrated edge, cloud and high-performance computing (HPC) hardware and software CI components simplify the use of AI, making it easier to address new areas of inquiry. In this way, ICICLE focuses on research in AI, innovation through AI and accelerates the application of AI.

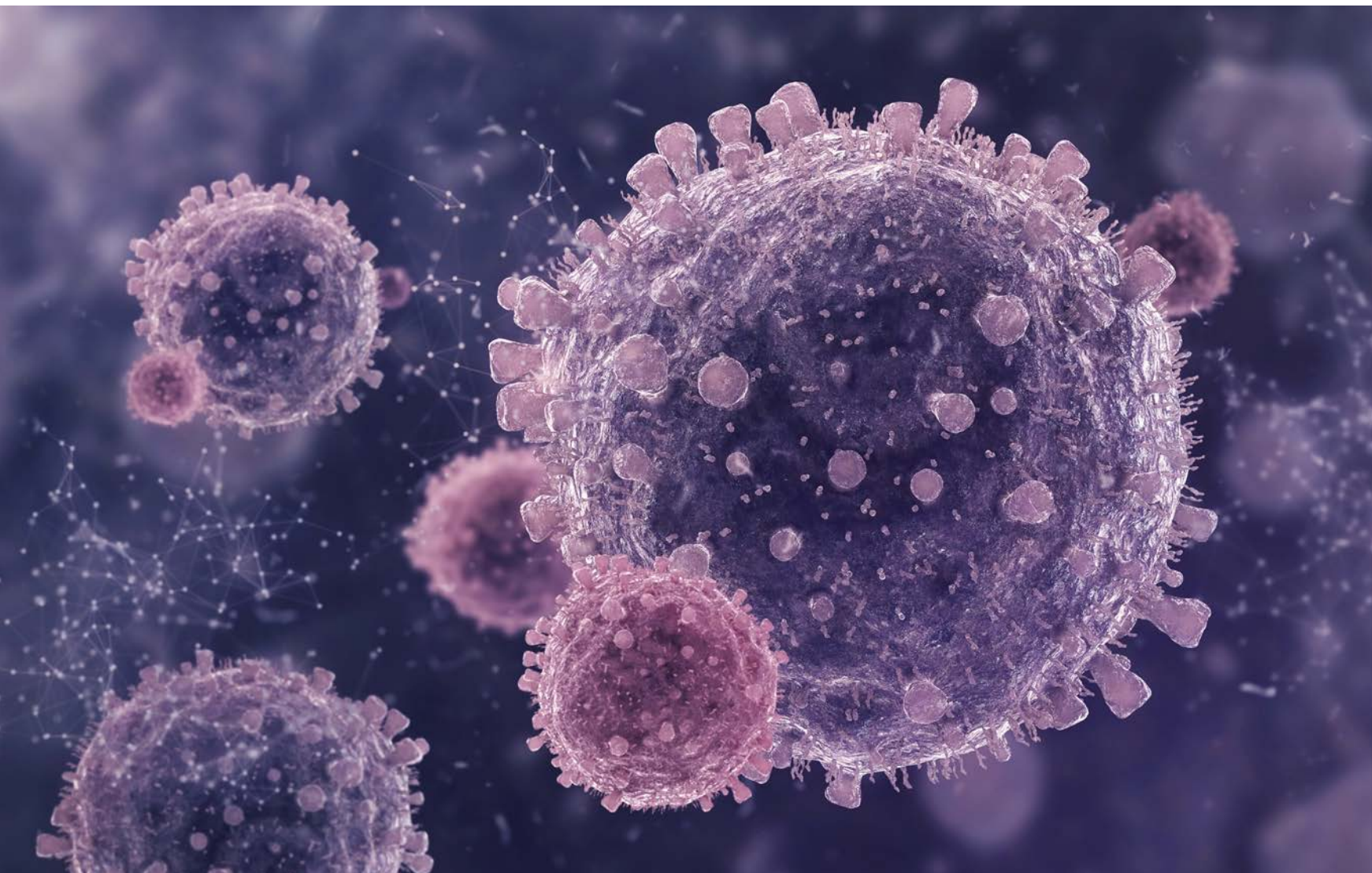
Notably, ICICLE is building a diverse STEM workforce through innovative approaches to education, training and broadening

participation in computing that ensure sustained measurable outcomes and impact on a national scale—supporting a pipeline from middle- and high-school students to practitioners. As a nexus of collaboration, ICICLE promotes technology transfer to industry and other stakeholders, as well as data sharing and coordination across other NSF AI institutes and federal agencies. As a national resource for research, development, technology transfer, workforce development and education, ICICLE is creating a widely usable, smarter, more robust and diverse, resilient and effective CI4AI and AI4CI ecosystem.

“ICICLE will enable a transparent and trustworthy national infrastructure for an AI-enabled future, addressing pressing societal problems and enabling decisions for national priorities,” said PI and Professor and Distinguished Scholar of Computer Science at the Ohio State University Dhabaleswar Panda. “ICICLE will grow new generations of workforce and incubate innovative companies, with sustained diversity and inclusion at all levels.”

The 11 new AI Research Institutes add to the seven AI Institutes funded with \$140 million by NSF in 2020. The AI Institutes, led by NSF in partnership with the U.S. Department of Agriculture National Institute of Food and Agriculture (USDA-NIFA), U.S. Department of Homeland Security (DHS), Google, Amazon, Intel and Accenture, will act as connections in a broader nationwide network that will pursue transformational advances in a range of economic sectors and science and engineering fields.





New Pathways for Potential RNA Virus Treatment

Similar to how humans encode their genome using DNA, many viruses have a genetic makeup of RNA molecules. These RNA-based genomes contain potential sites where inhibitors can attach and deactivate the virus. Part of the challenge in drug development is that variations or mutations in the viral genome may prevent the inhibitors from attaching.

So, researchers at the University of New Hampshire (UNH) looked to use Comet at the San Diego Supercomputer Center (SDSC) and Stampede2 at the Texas Advanced Computing Center (TACC), both part of NSF's Extreme Science and Engineering Environment (XSEDE), to identify new inhibitor binding/unbinding pathways in an RNA-based virus. Their findings, published in the Journal of Physical Chemistry Letters, could be beneficial in understanding how these inhibitors react and potentially help develop a new generation of drugs to target viruses with high death rates, such as HIV-1, Zika, Ebola and SARS-CoV2, the virus that causes COVID-19.

"When we first started this research, we never anticipated that we'd be in the midst of a pandemic caused by an RNA virus," said Harish Vashisth, associate professor of chemical engineering at UNH. "As these types of viruses emerge, our findings will hopefully offer an enhanced understanding of how viral RNAs interact with inhibitors and be used to design better treatments."

In their study, Vashisth and his team created molecular dynamics simulations using Comet and Stampede2 supercomputers to look specifically at an RNA fragment from the HIV-1 virus and its interaction with acetylpromazine, a small molecule that is known to interfere with the virus replication process.

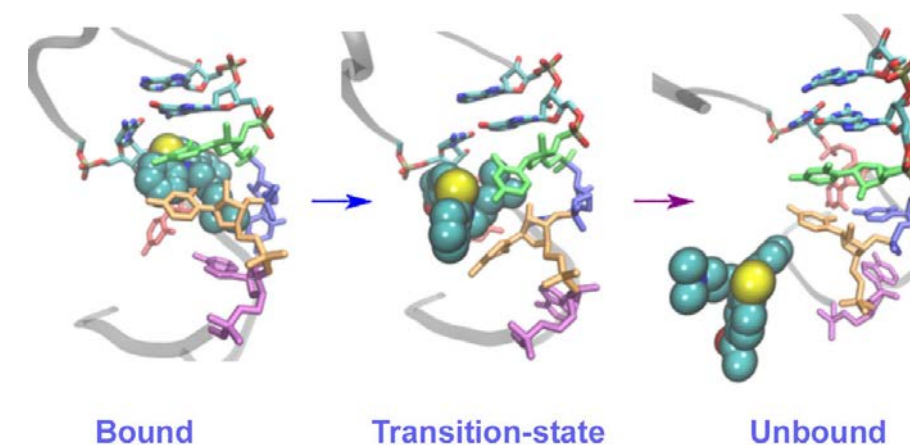
"Comet and Stampede2 provided exceptional platforms to enhance the impact of our work carried out via existing grants," said Vashisth, who has been using XSEDE resources for almost a decade. "The XSEDE initiative has been a game changer in biomolecular simulations by providing access to a large number of pre-installed software tools, training for students and principal investigators such as myself, data analysis and visualization tools, and long-term data storage."

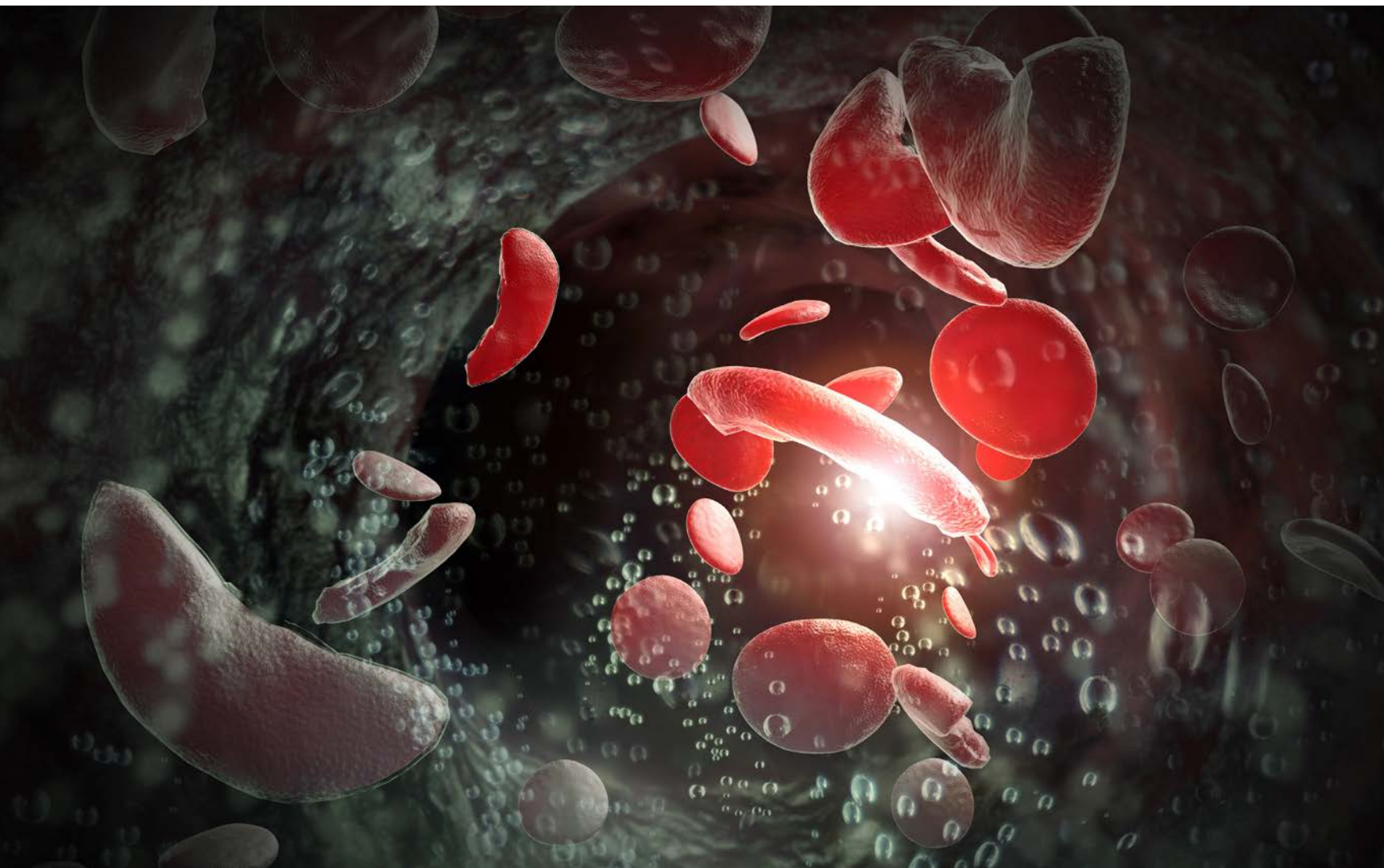
The scientists focused on the structural elements from the HIV-1 RNA genome because they are considered a good model for studying the same processes across a wide range of RNA viruses. These simulations enabled them to discover the pathways of the inhibitor unbinding from the viral RNA in several rare events – which are often difficult to observe experimentally – that unexpectedly showed a coordinated movement in many parts of the binding pocket that are the building blocks of RNA.

Comet and Stampede2 enabled the researchers to run hundreds of simulations at the same time to observe what are called rare base-flipping events involved in the inhibitor binding/unbinding process that provided the new details of the underlying mechanism of this process.

"Our hope is that this adds new possibilities to a field traditionally focused on static biomolecular structures and leads to new medications," Vashisth said.

Structural changes in RNA on drug binding/unbinding.
Credit: Lev Levintov, University of New Hampshire





Comet Aids in Sickle Cell Research Discovery

Although there has yet to be a cure for sickle cell disease, researchers this year used National Science Foundation’s Extreme Science and Engineering Discovery Environment (XSEDE) allocations on Comet at SDSC to create detailed simulations that showed how these stiff red blood cells flow through blood vessels, deforming and colliding along the way.

The study, which was published in the *Physical Review Fluids* journal, revealed new information about the ways in which sickle cells collide with one another as well as healthy cells and blood vessel walls.

“Our new simulations showed how the motions of sickle cells near vessel walls generate large forces,” explained the study’s Principal Investigator Michael Graham, a professor of chemical and biological engineering at the University of Wisconsin-Madison, who

collaborated with Wilbur Lam, a physician and biomedical engineer at Emory University and the Georgia Institute of Technology. “The forces caused by these sickle cells impact healthy cells on the blood vessel walls, which in turns causes inflammation.”

“We are grateful for being able to use supercomputers like Comet to help us better illustrate and hopefully come closer to a cure for sickle cell disease.”

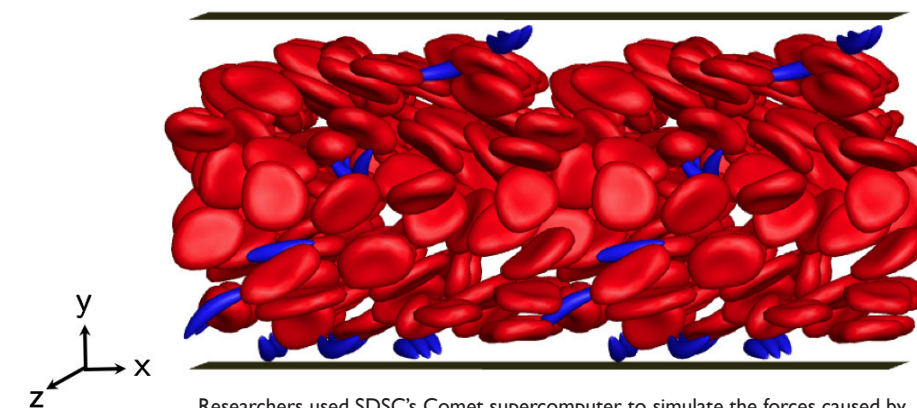
– Wilbur Lam

In particular, Graham said the group’s simulations provided evidence regarding damage to healthy cells lining blood vessels, but the origin of this damage had not been well understood. According to Graham, large-scale shared computing resources such as Comet open doors to doing simulations that could not be done with the resources of an individual research group.

“We are grateful for being able to use supercomputers like Comet to help us better illustrate and hopefully come closer to a cure for sickle cell disease,” said Lam, who is also a pediatric hematologist at

the Aflac Cancer and Blood Disorders Center of Children’s Healthcare of Atlanta, which has the largest pediatric hematology program in the United States.

Last year, Lam’s dedication to sickle cell disease research was recognized by the Atlanta Business Chronicle as a finalist for the Health Care Innovator/Researcher Award in the Annual Health Care Heroes Awards. In addition to his collaboration with Graham on these latest supercomputer simulations, Lam developed a non-invasive test for anemia that enables anyone at risk for that condition, including sickle cell disease patients, to send an image of their fingernails to measure their hemoglobin levels and determine whether they need to seek medical care.



Researchers used SDSC’s Comet supercomputer to simulate the forces caused by sickle cells upon vessel walls. This snapshot shows a front view of healthy red blood cells (red) and sickle red blood cells (blue) as they are distributed during flow. Credit: Michael Graham, University of Wisconsin – Madison, et al.



Enabling 3-D Models of Fresh Water Plastic Pollution

The transport of nine types of plastics floating in Lake Erie was modeled in two studies that used SDSC's Comet supercomputer located at UC San Diego. The studies compared a two-dimensional model with a new Great Lakes microplastic dataset and then developed the first ever three-dimensional mass estimate for plastic in Lake Erie. The studies were published in the *Marine Pollution Bulletin* and the *Journal of Great Lakes Research* by researchers from Pennsylvania State University (Penn State) at Behrend and the Rochester Institute of Technology (RIT).

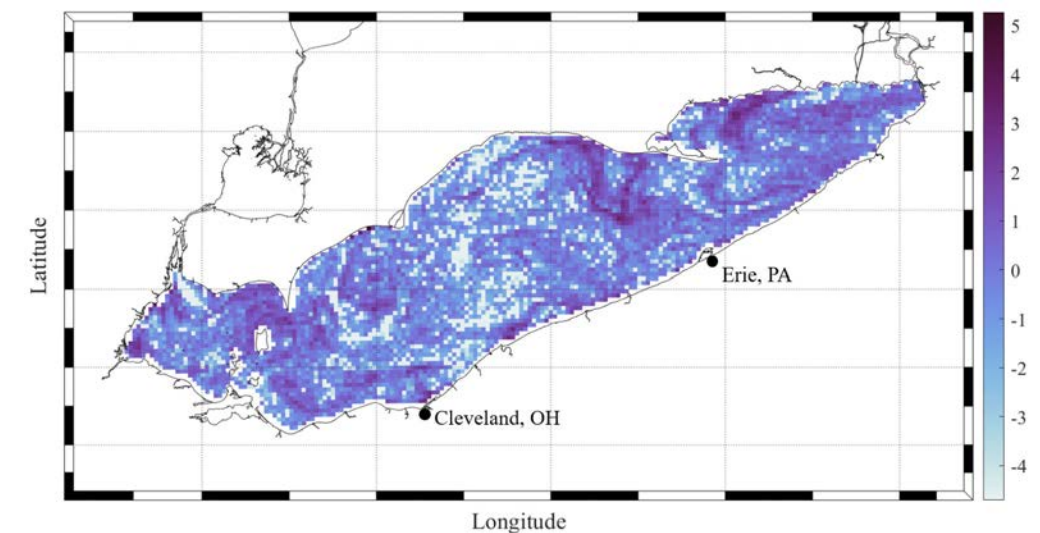
"Comet allowed us to run three years of hydrodynamic simulations of Lake Erie that drive the transport of plastic particles within Lake Erie," said Matthew Hoffman, an RIT associate professor of mathematical sciences. "We don't have the computing infrastructure at my university to run these simulations efficiently, so having access to Comet enabled us to run state-of-the-art code that is not possible on our local desktops and laptops."

Hoffman worked with Penn State-Behrend Chemistry Professor Sherri Mason, who led the collection of the sampling data that the modeling simulated. Mason's laboratory, which is located on the edge of Lake Erie, made sampling quite convenient.

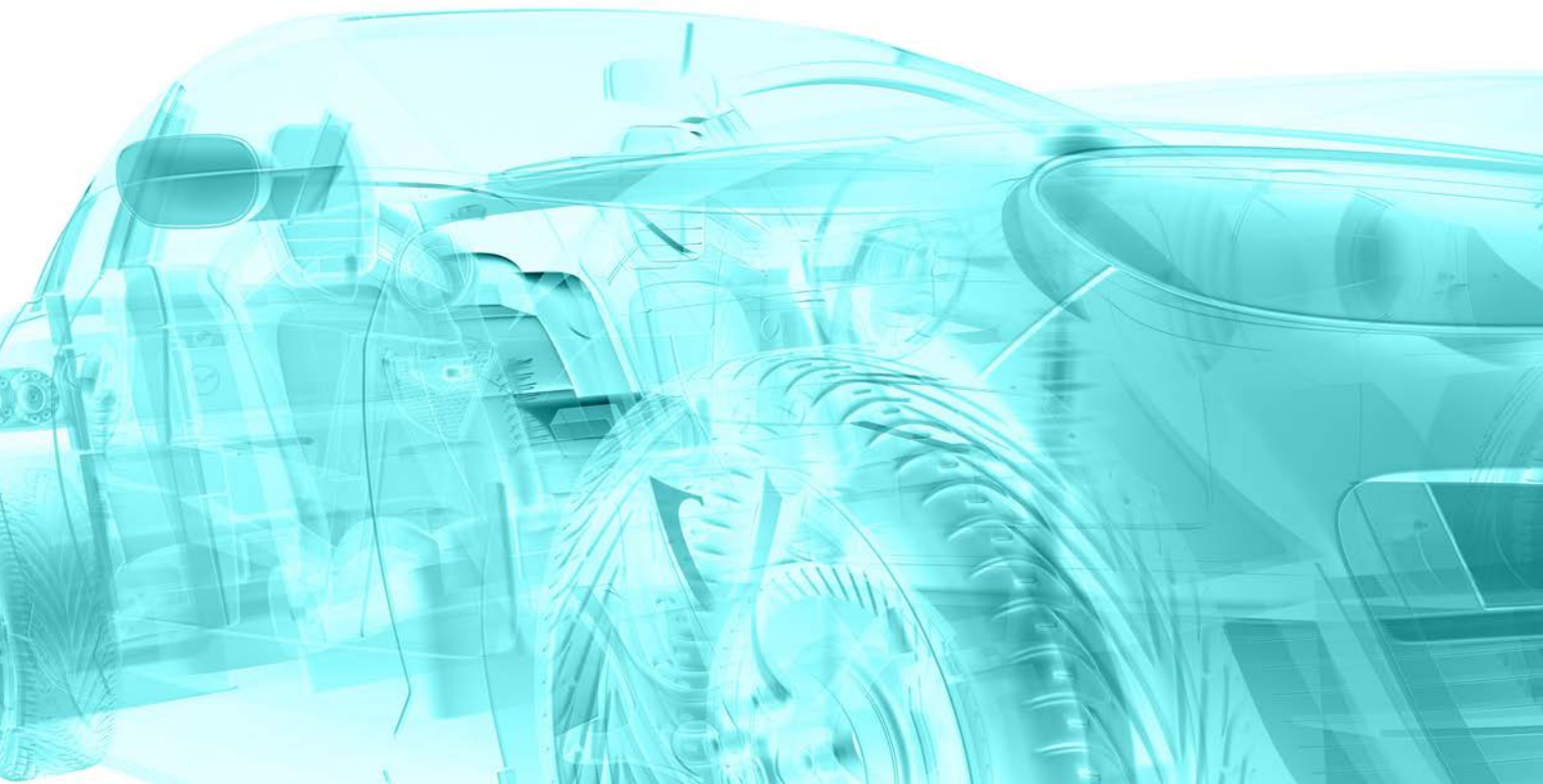
"Undergraduate students were heavily involved in this process," added Mason. "We used three different vessels (Niagara, Sea Dragon, and Sara B) to collect our samples, taking advantage of every opportunity afforded to us. We would hop aboard, collect our samples, then take them back to the lab for analysis. Such fieldwork is important to better understand how plastic is not only impacting our lake, but freshwater lakes around the world."

There were more than six metric tons found on the surface and approximately 300 metric tons in the volume of the lake. The nine plastics involved with the study were polystyrene (PS), polyamide (PA; nylon), polymethylene methacrylate (PMMA; acrylic), polyethylene terephthalate (PET; polyester), polyoxymethylene (POM), polyvinyl chloride (PVC), polyethylene (PE), polypropylene (PP), and expanded polystyrene (EPS).

This research was funded by The Burning River Foundation. Researchers were provided access to Comet via the National Science Foundation's Extreme Science and Engineering Discovery Environment (XSEDE) program under allocation TG-OCE150006.



According to the National Oceanic and Atmospheric Administration (NOAA), 75 percent of predicted worldwide plastic waste involves nine specific polymer types. These kinds of plastic were the focus of this model of Lake Erie completed using XSEDE allocations on SDSC's Comet supercomputer. The plastic concentration is shown on a logarithmic scale, with light blue indicating lower concentrations and dark purple higher concentrations. Credit: M. Hoffman (Rochester Institute of Technology), S. Mason (Pennsylvania State University at Behrend)



Supercomputer Simulations Show Potential for Carbon-Neutral Fuels

When it comes to potential, porous carbon materials and students share it in common. The porous carbon materials show potential in wide-ranging applications, while university students show promise in their participation on research teams, exploring things such as the technology needed for reduced-carbon or carbon-free chemical fuels for vehicles.

Rylan Rowsey, a first-year graduate student in chemistry at UC San Diego, attended Montana State University (MSU) as an undergraduate who used Comet to study zeolite-templated carbon (ZTC) as a gas storage material.

“The computational chemistry skills and tools I learned are invaluable to my scientific growth as a researcher, and now that I am beginning graduate school, I feel that I am coming in prepared with a unique skill set,” said Rowsey.

The skills and tools to which Rowsey referred were applied in the computational modeling of ZTC structural maquettes (small models of surface sites of interest) to evaluate methane adsorption as a function of chemical composition. The U.S. Department of Energy-supported study was led by Nicholas Stadie, an assistant professor at MSU and Robert Szilagy, an associate professor now at the University of British Columbia, and published in the *Journal of Physical Chemistry A*.

“We clearly identified the preference of methane toward nitrogen-substituted sites,” said Szilagy. “The estimated binding energies and the experimentally measured heat of adsorption values are in good agreement; thus, validating the use of computational chemistry.”

Computer-aided rationalized design of materials provides an efficient way to develop new or optimize existing energy storage and conversion technologies. Instead of a traditional trial-and-error approach, computer modeling of chemical processes reduces product-to-market times, minimizes chemical waste and maximizes the use of human resources.

Specifically, methane storage and conversion to energy or to chemical feedstocks without greenhouse gas emissions is a long overdue petrochemical target. Computational materials design allows for simulation of every step of methane storage or conversion at the atomic-scale, which provides insights into how to mitigate emissions and chemical waste.

Stadie and Szilagy demonstrated that experimentalists can use simulations to guide their strategies in synthesizing nitrogen-doped carbon materials for methane storage and delivery. The computational results clearly documented the benefit of focusing on nitrogen-doping of carbon materials for methane storage.

“We did not expect to see the consistent preference for methane toward nitrogen-substituted carbon maquettes, over both the unsubstituted and the boron-substituted models,” said Szilagy.

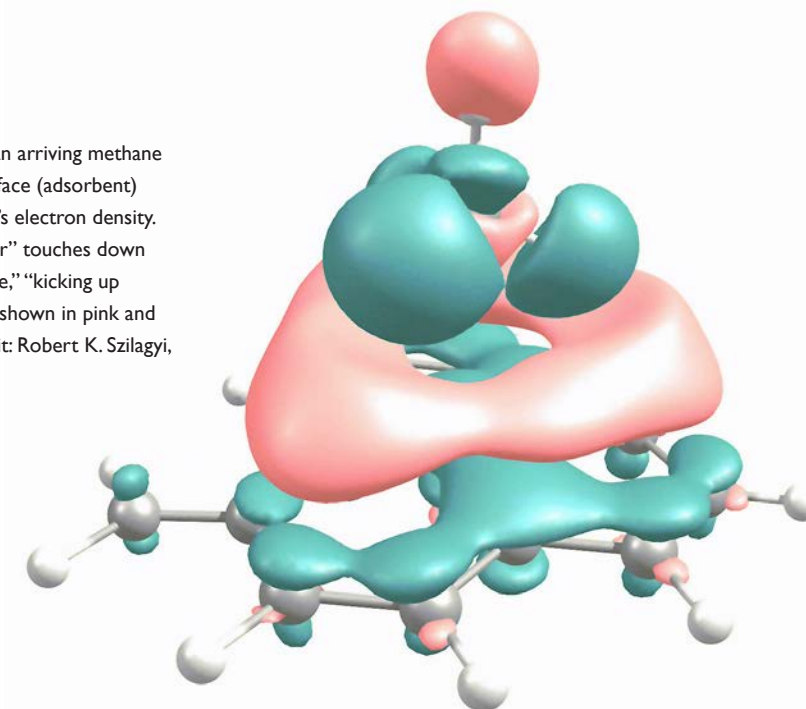
“The calculated energy differences were just the size of effect we were expecting,” added Stadie. “This effect has very important implications for methane storage at ambient temperature.”

While the team had access to local resources, accessing Comet allowed for the speedy completion of intense computational studies. The team was able to run potential energy surface mapping where each grid point was calculated at high level of theory without truncations or simplifications.

“The additional computational resources allowed us to complete a more thorough study, including computational control and blank simulations, which are mandatory for experimental work,” said Szilagy.

“The experimentalist’s approach to computational chemistry has great potential to elevate the validity of theoretical models,” continued Stadie. “The successful combination of theory and experiment enhances the impact of computations.”

This image illustrates how an arriving methane (adsorbate) and carbon surface (adsorbent) mutually modify each other’s electron density. Here, the methane = “lander” touches down at the carbon “moon surface,” “kicking up dust” (i.e., electron density, shown in pink and green) in the process. Credit: Robert K. Szilagy, Montana State University





Education & Outreach at SDSC

SDSC’s education and outreach programs serve students from middle school through graduate school. Additionally, SDSC offers high-performance online courses that attract participants from all over the world.

Research Experience for High School Students

SDSC’s Research Experience for High School Students (REHS) program celebrated its 12th year in 2021. The program increases awareness of computational science, technical writing and related fields of research among students in the greater San Diego area. Conducted remotely starting in 2020 due to the COVID-19 pandemic, the 2021 program paired nearly 70 students with eight SDSC mentors, who guided students as they gained practical experience and exposure to career options and work-readiness skills. The program concluded with a virtual “Project Showcase,” during which students shared their research projects with peers, other mentors, family and friends. To date, more than 460 students have participated in the REHS program.

Mentor Assistance Program

Each year, high school students interested in pursuing a career in scientific 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 students to gain access to UC San Diego faculty, postdoctoral fellows, doctoral candidates and staff to mentor them in their specific field of interest. Mentors are recruited from biology, chemistry, aerospace engineering, network architectures, pharmaceutical sciences, physics, social sciences and other fields of study.

ForMIDABLE

An offshoot of UC San Diego’s Anita Borg Leadership and Engagement (ABLE) program, this new eight-week program introduces middle school students from six pilot schools to STEM careers through hands-on workshops, invited speakers, tutorials and robotics demonstrations.

Pi Wars

Once again, SDSC participated in Pi Wars – an international, challenge-based robotics competition in which teams build Raspberry Pi-controlled robots and then compete in non-destructive autonomous and remote-controlled challenges. The competition encompassed several teams of middle and high school students. Because the program was virtual, to keep the students engaged, the SDSC education team led several online talks, demonstrations and virtual visits with undergraduate robotics teams.

High Performance Computing (HPC) Students Program

The HPC Students Program focuses on organizing, coordinating and supporting club activities; purchasing/loaning tool and cluster hardware to the club and sponsoring students to travel to the annual Supercomputing Conference (SC). This program also hosts the HPC User Training classes in collaboration with the club, where participants are taught about the architecture of HPC clusters, and they learn to run scientific applications on those systems. The program also organizes and awards Co-Curricular Record (CCR) credits to SDSC interns while assisting principal investigators to create new CCRs.

An additional initiative for undergraduate students is the annual Student Cluster Competition (SCC), which encompasses student involvement at the annual SC.

Online Data Science & Big Data Courses

UC San Diego offers a four-part Data Science Series via edX’s MicroMasters® program with instructors from the computer science and engineering department and SDSC. In partnership with Coursera, SDSC created a series of massive open online courses (MOOCs) 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, students apply their acquired skills to a real-world big data problem. To date, the courses have reached more than one million students around the world – from Uruguay to the Ivory Coast to Bangladesh. A subset of students pays for a certificate of completion.

HPC Training Webinars & Workshops

A vast array of HPC training opportunities were offered over the past year at SDSC. Specifically, 19 events were held that focused on familiarizing researchers with HPC systems such as Expanse, Voyager and the Triton Shared Computing Cluster. SDSC’s HPC programs were attended by more than 1,600 participants, and workshop 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.

Cyberinfrastructure-Enabled Machine Learning (CIML) Summer Institute

This year’s CIML Summer Institute introduced machine learning (ML) researchers, developers and educators to the techniques and methods needed to migrate their ML applications from smaller, locally run resources, such as laptops and workstations, to large-scale HPC systems, such as the SDSC Expanse supercomputer. Participants accelerated their learning process through highly interactive classes with hands-on tutorials on SDSC’s Expanse.

SDSC Summer Institute

Aimed at researchers in both academia and industry, the week-long workshop focused on a broad spectrum of introductory-to-intermediate topics in HPC and Data Science. The 2021 institute gave participants an opportunity to accelerate their learning process through highly interactive classes with hands-on tutorials on the Expanse supercomputer. This year’s Summer Institute continued SDSC’s strategy of bringing advanced cyberinfrastructure to the long tail of science and provided resources to a larger number of modest-sized computational research projects that advance, in aggregate, a tremendous amount of scientific progress.



SDSC Education & Training
www.sdsc.edu/education_and_training/index.html



SDSC StudentTech
education.sdsc.edu/studenttech/



SDSC Advanced Computing Training
sdsc.edu/education_and_training/training_hpc.html

SDSC EXECUTIVE TEAM

Frank Würthwein, Ph.D.

Director, SDSC
Professor, UC San Diego Department of Physics
Professor, The Halicioğlu Data Science Institute

Ilkay Altintas, Ph.D.

Chief Data Science Officer
Division Director, Cyberinfrastructure and Convergence
Research and Education (CICORE)
Director, Workflows for Data Science (WorDS)
Director, WIFIRE Lab

Sandeep Chandra

Division Director, Sherlock Cloud Solutions & Services

Cynthia Dillon

Division Director, External Relations
Director, Communications

Ronald Hawkins

Director, Industry Partnerships

Christine Kirkpatrick

Division Director, Research Data Services
Principal Investigator (PI), EarthCube Office (ECO)
Co-PI, West Big Data Innovation Hub
Co-PI, Open Storage Network

Samuel 'Fritz' Leader

Chief Administrative Officer
Division Director, Business Services

Amit Majumdar, Ph.D.

Division Director, Data-Enabled Scientific Computing
Associate Professor, Department of Radiation Medicine and
Applied Sciences
PI, Voyager

Shawn Strande

Deputy Director, SDSC

Michael Zentner, Ph.D.

Division Director, Sustainable Scientific Software

RESEARCH EXPERTS

Chaitan Baru, Ph.D.

Senior Science Advisor, Office of Integrative Activities, Office
of the Director, NSF
Distinguished Scientist, SDSC
Director, Center for Large-scale Data Systems research (CLDS)
Associate Director, Data Science and Engineering
Associate Director, Data Initiatives

James Bordner, Ph.D.

Computational Scientist

Hans-Werner Braun

Research Scientist
Internet Hall of Fame Inductee

Dong Ju Choi, Ph.D.

Senior Computational Scientist
Assistant Clinical Professor, Department of Radiation Medicine
and Applied Sciences, UC San Diego

Amit Chourasia

Senior Visualization Scientist
PI, Stream Encode Explore and Disseminate
My Experiments

Kimberly Claffy, Ph.D.

Director, Center for Applied Internet Data Analysis (CAIDA)
Research Scientist
Internet Hall of Fame Inductee

Melissa Cragin, Ph.D.

Chief Strategist for Data Initiatives, Research Data Services (RDS)

Daniel Crawl, Ph.D.

Associate Director, Workflows for Data Science

Yifeng Cui, Ph.D.

Director, High-Performance GeoComputing Laboratory
Director, Intel Parallel Computing Center
PI, Southern California Earthquake Center

Alberto Dainotti, Ph.D.

Assistant Research Scientist, CAIDA

Diego Davila

Computer Scientist

Jose M. Duarte, Ph.D.

Assistant Project Scientist, RCSB Protein Data Bank

Anthony Gamst, Ph.D.

Director, Computational and Applied Statistics Laboratory

Andreas Goetz, Ph.D.

Assistant Research Scientist, PI, Co-Director, GPU Education Center

Madhusudan Gujral, Ph.D.

Bioinformatics Programmer Analyst

Amarnath Gupta, Ph.D.

Research Scientist, Advanced Query Processing Laboratory, Data
and Knowledge Systems

Valentina Kouznetsov, Ph.D.

Associate Project Scientist

Alexander Marder, Ph.D.

Research Scientist, CAIDA

Mark Miller, Ph.D.

PI, Biology

Dmitry Mishin, Ph.D.

Applications Developer

Ka Pui Mok, Ph.D.

Research Scientist, CAIDA

David Nadeau, Ph.D. (outgoing)

Senior Visualization Researcher

Viswanath Nandigam

Associate Director, Advanced Cyberinfrastructure Development Lab
PI, Open Topography

Mai H. Nguyen, Ph.D.

Lead Data Scientist, Data Analytics

Michael Norman, Ph.D.

Former Director, SDSC
Distinguished Professor, Physics, UC San Diego
Director, Laboratory for Computational Astrophysics, UC San Diego

Dmitri Pekurovsky, Ph.D.

Senior Computational Scientist

Wayne Pfeiffer, Ph.D.

Distinguished Scientist

Peter Rose, Ph.D.

Director, Structural Bioinformatics Laboratory
Lead, Bioinformatics and Biomedical Applications, Data Science Hub

Joan Segura, Ph.D.

Scientific Software Developer, RCSB Protein Data Bank

Igor Sfiligoi

Senior Research Scientist, Distributed High-Throughput Computing
Lead Scientific Software Developer and Researcher

James Short, Ph.D.

Lead Scientist
Co-Director, Center for Large-scale Data Systems Research (CLDS)
Director, BlockLAB

Robert Sinkovits, Ph.D.

Director, Scientific Computing Applications

Subhashini Sivagnanam

Lead, CyberInfrastructure Solutions and Services
PI, Open Science Chain
Co-PI, Neuroscience Gateway
2021 Pi Person of the Year

Shava Smallen

Manager, Cloud and Cluster Development
Lead Software Architect and Co-PI, CloudBank
Co-Deputy Manager, XSEDE Cyberinfrastructure Integration (XCI)
Deputy Manager, XSEDE Requirements Analysis and Capability
Delivery (RACD)
Steering Committee Co-Chair, Pacific Rim Application and Grid
Middleware Assembly (PRAGMA)

Mahidhar Tatineni, Ph.D.

Lead, User Support
Research Programmer Analyst

Mary Thomas, Ph.D.

Computational Data Scientist
Lead, HPC Training
Co-PI, CC* Compute: Triton Stratus

Igor Tsigelny, Ph.D.

Research Scientist, SDSC and Department of Neurosciences, UC
San Diego

David Valentine, Ph.D.

Research Programmer, Spatial Information Systems Laboratory

Tanya Wolfson

Senior Staff Member, Computational and Applied Statistics Laboratory

Kenneth Yoshimoto

Researcher, Computational and Data Science

Choonhan Youn, Ph.D.

Scientific Researcher

Ilya Zaslavsky, Ph.D.

Director, Spatial Information Systems Laboratory


Andrea Zonca, Ph.D.

Specialist, HPC Applications

SDSC@UC San Diego

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


www.sdsc.edu


 email/info@sdsc.edu

 [twitter/SDSC_UCSD](https://twitter.com/SDSC_UCSD)

 [instagram.com/SDSC_UCSD](https://www.instagram.com/SDSC_UCSD)

 [facebook/SanDiegoSupercomputerCenter](https://www.facebook.com/SanDiegoSupercomputerCenter)

 [linkedin.com/company/san-diego-supercomputer-center](https://www.linkedin.com/company/san-diego-supercomputer-center)

 [youtube.com/SanDiegoSupercomputerCenter](https://www.youtube.com/SanDiegoSupercomputerCenter)