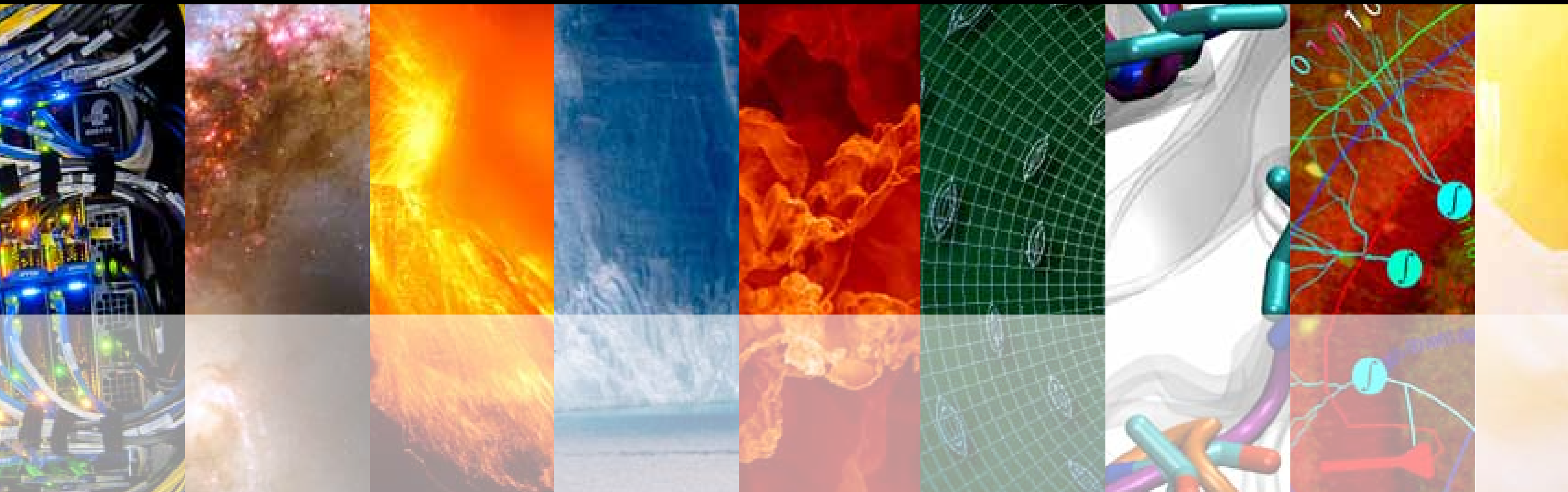
SAN DIEGO SUPERCOMPUTER CENTER at UC SAN DIEGO



DATA to DISCOVERY

Research and Development for UC and the State of California

SDSC

ANNUAL REPORT 2012

SAN DIEGO SUPERCOMPUTER CENTER

As an Organized Research Unit of UC San Diego, SDSC is considered a leader in data-intensive computing and all aspects of "Big Data," which includes data integration, performance modeling, data mining, software development, workflow automation, and more. SDSC supports hundreds of multidisciplinary programs spanning a wide variety of domains, from earth sciences and biology to astrophysics, bioinformatics, and health IT. With its two newest supercomputer systems, Trestles and Gordon, SDSC is a partner in XSEDE (Extreme Science and Engineering Discovery Environment), the most advanced collection of integrated digital resources and services in the world.

SDSC INFORMATION
Michael L. Norman, Director

San Diego Supercomputer Center
University of California, San Diego
9500 Gilman Drive MC 0505
La Jolla, CA 92093-0505
Phone: 858-534-5000
Fax: 858-534-5152

info@sdsc.edu
www.sdsc.edu

Warren Froelich
Director, External Relations wfroelich@sdsc.edu
858-822-3622

DATA to DISCOVERY
SDSC Annual Report 2012
(PDF version available online at the SDSC website)

EDITOR:
Warren Froelich

CO-EDITOR:
Jan Zverina
Director, Communications and Media Relations

CONTRIBUTORS:
Diane Baxter, Cassie Ferguson, Warren Froelich, Ron Hawkins, Alma Palazzolo, Susan Rathbun, Mahidhar Tatineni, Jan Zverina

CREATIVE DIRECTOR:
Ben Tolo

DESIGN:
Ben Tolo

In memory of Allan E. Snavely, 1962-2012

Bridging the Third and Fourth Paradigms for "Big Data" Science

DIRECTOR'S LETTER

Science has entered an historical crossroads, a paradigm shift in the way researchers discover and innovate. If high-performance computing and computational science during the past three or four decades provided the framework for the "third paradigm" of science—following centuries of experimental and theoretical science—then today we've entered the "fourth paradigm," built around the need to harness what many have called "Big Data" or the "data deluge."

From all corners of the globe, a veritable tidal wave of data is being created by academic centers, commercial laboratories, government scientists, and observational tools such as satellites, oceanographic sensors, astronomical observatories, and personal websites. The planet is awash in data, so much so that it threatens to flood our ability to preserve, analyze, comprehend, and apply in a meaningful way. Yet many researchers believe that humanity's future well-being may hinge on the mysteries hidden in all this data.

The San Diego Supercomputer Center (SDSC)—an Organized Research Unit at the University of California, San Diego—is convinced that the challenges posed by the fourth paradigm of science are so critical that it's now a central part of our mission. This means we are focusing on developing technologies and infrastructure to help make sense of all this data, built around a deep expertise that includes data integration, data management, data mining, software development, workflow automation, and more.

In this annual report, we highlight our historical strengths in high-performance computing and scientific computation—the third paradigm—but we also place a spotlight on our pioneering efforts in the fourth paradigm. The National Science Foundation (NSF) recognizes that these two methodologies of modern science do not exist in isolation but complement one another, and are often practiced in the same labs. NSF refers to this cyberinfrastructure-enabled scientific research as "computation and data-enabled science and engineering." This report will show how SDSC is providing a bridge that connects these two contemporary paradigms of scientific discovery in the 21st century.

Within the last couple of years, SDSC has built an impressive suite of data cyberinfrastructure services and expertise dedicated to Big Data problems, with those resources available to UC and national researchers. Perhaps our most visible example is embodied by *Gordon*, funded by a $20 million grant from the NSF as the nation's first "data-intensive" supercomputer. Launched in January 2012, *Gordon* is a unique flash memory-based system capable of storing 100,000 entire human genomes, while operating hundreds of times faster than conventional computers to study genetic data. More than 100 UC researchers have projects on *Gordon*.

Providing a secure and stable environment to store short- and long-term research and archival data is another essential component of Big Data services. The SDSC Cloud, which went into production in late 2011, is the first large-scale academic deployment of cloud storage in the world. *Data Oasis*, operational in 2012, provides a high-performance Lustre-based parallel file system with four petabytes of storage and a 100 gigabyte-per-second (GB/s) connection for scratch and medium-term storage.

SDSC also has made half of its 19,000 square-foot data center available as a recharge-based facility to users across the UC system, now including more than 90 groups spanning eight UC campuses, at an estimated annual system-wide utility cost-savings of about $822,000.

These state-of-the-art Big Data services continue to place SDSC at the hub of academic research across UC and beyond. Today, SDSC's pioneering data-intensive computing and high-performance storage technologies are opening new doors for discovery in areas vital to the future of Californians and the rest of the nation. This report spotlights but a few key research collaborations with UC researchers that encompass the following disciplines: the designing of nanoparticles and novel materials for industry and medicine; delving inside the brain to treat disorders; modeling ocean and coastal impacts from climate change; tackling turbulence problems for airplane design; probing the earth's interior to understand earthquakes, volcanoes, and other seismic disruptions; and others.

Another area of critical interest is the study of genes, with a focus on genomic medicine. Next-generation sequencing of DNA and RNA is profoundly transforming biology and medicine, providing insights into our origins and diseases, with the potential of tailoring treatments to individual patients, or so-called personalized medicine. However, obtaining insights from the sequencer data deluge requires complex software and increasingly powerful data-intensive computers.

For example, SDSC is collaborating with David Haussler at UC Santa Cruz to store the Cancer Genomics Hub (CGHub), a large-scale data repository and user portal for the National Cancer Institute's cancer genome research programs, with the goal of targeting anti-cancer drugs to specific tumors and individual patients based on their genetic signatures.

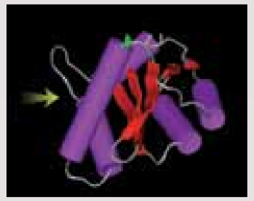
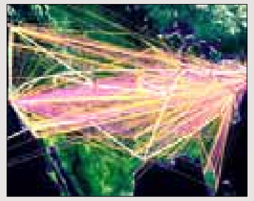
Genetic data creates many additional requirements for sharing and computing. The iDASH project (integrating Data for Analysis, Anonymization, and Sharing), under the leadership of Lucila Ohno-Machado and funded by the National Institutes of Health (NIH) Roadmap for Bioinformatics and Computational Biology, is providing biomedical and behavior scientists with access to a sophisticated, secure privacy-preserving data cyberinfrastructure needed to build and analyze their data.

SDSC also is providing a secure environment to store patient records for the UC Riverside School of Medicine and UC Davis, and is fully compliant with regulations in the Health Insurance Portability and Accountability Act of 1996 (HIPAA). The Center is hosting and managing collected data as part of the National Children's Study, the largest long-term examination of children's health and development ever conducted in the United States.

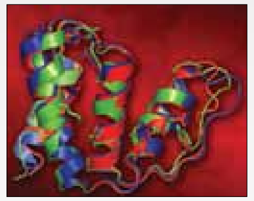
The revolution in genomics, bioinformatics, and other "omics" is but one example of how SDSC is leveraging its resources and expertise across Big Data, for science and society, UC researchers, and others. Potential applications run the gamut from astronomy to zoology. After you've reviewed this annual report, I hope you'll understand why we're excited about this new era at the crossroads of two scientific paradigms, and why we're devoting so much time and energy into this effort.

Michael L. Norman
SDSC Director

More than a Quarter Century of "Turning Data into Discovery" for the University of California





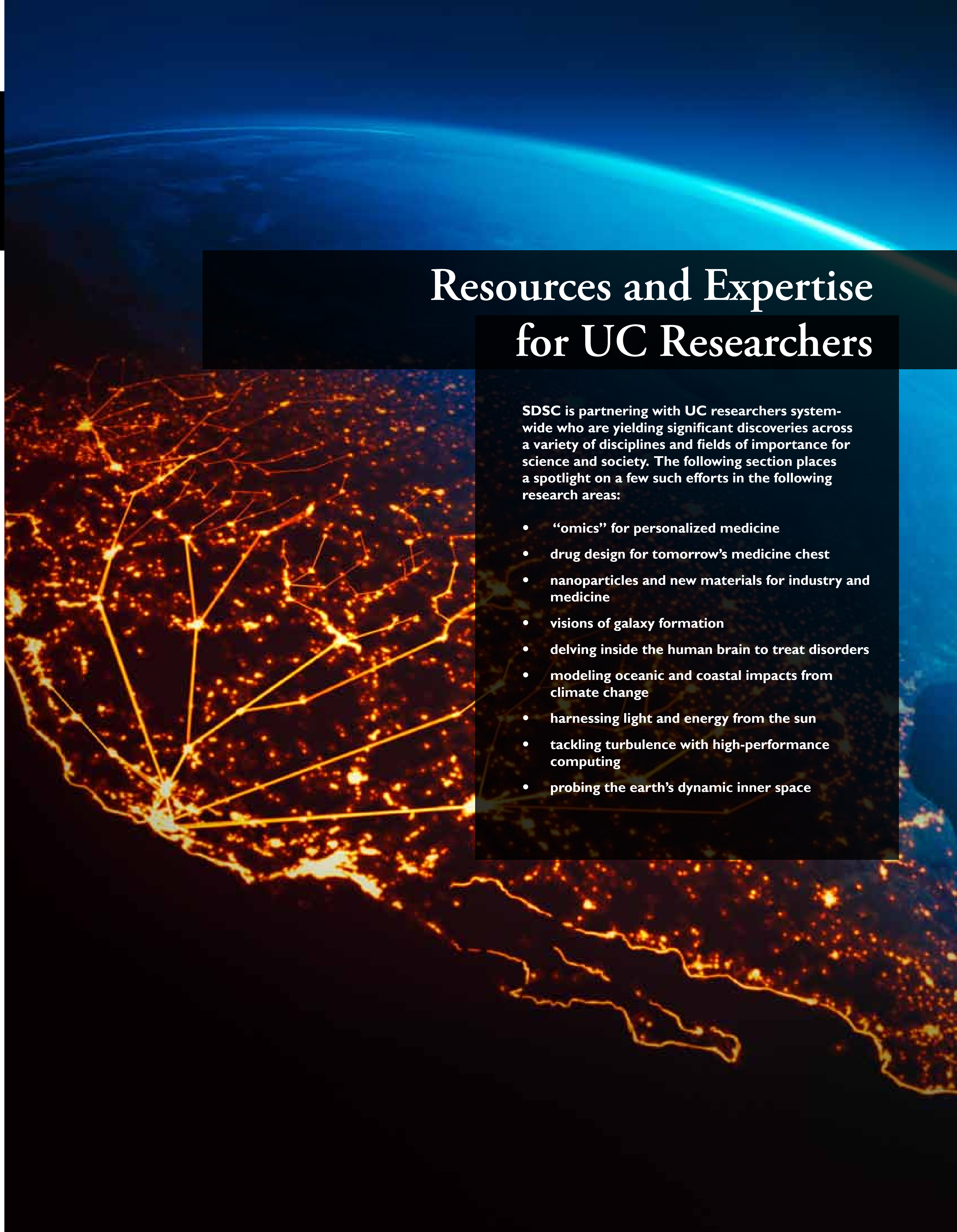
Since opening its doors on November 14, 1985, scientific and technological advances made possible and/or created by SDSC staff and resources have made a major mark in academia, industry, and society-at-large—"turning data into discovery" a phrase that has become synonymous with SDSC. The Center has brought together researchers at UC San Diego and across UC, the nation, and the world, in partnerships and collaborations that now are the hallmark of today's scientific enterprise, building innovative and leading-edge cyberinfrastructure for this effort.

As such, the Center has provided an historical and unique competitive edge for the University of California, extending its impact, ability to attract funding, national rankings, and international reach through collaborations that require 21st century tools to address and solve 21st century science problems.

Many of these SDSC collaborations with UC faculty yielded landmark discoveries, were reported in prestigious scientific journals, and garnered local, state, and national headlines. The lengthy list includes:

- determination of the relative free energies of binding for different chemical inhibitors at the same molecular site—a pioneering effort by the UC San Francisco research team of Paul Bash, Peter Kollman, and Robert Langridge, considered one of the first steps in the rational design of drugs;
- a computer model that confirmed the importance of fossil fuel combustion in loading the atmosphere with carbon dioxide, with data collected by Charles Keeling from UC San Diego's Scripps Institution of Oceanography;
- discovery of the three-dimensional structure for protein kinase, considered the body's transistor, with the aid of a CRAY supercomputer at SDSC and a stereoscopic visualization system in SDSC's Advanced Scientific Visualization Lab (VizLab), funded by the State of California;
- first design of a molecule that moves in a straight line on a flat surface, by UC Riverside Professor Ludwig Bartels, considered a major step in the construction of molecular machines for a variety of applications, from computers to medicine;
- molecular dynamic simulations conducted by J. Andrew McCammon at UC San Diego that led to the development of a new anti-HIV drug;
- simulations by astrophysicist Richard Klein of UC Berkeley and Lawrence Livermore National Laboratory and his colleagues to support a model for how stars are formed;
- a world record established for size, performance, and fidelity for computer weather simulations, modeling the kind of "virtual weather" that society depends on for accurate weather forecasts;
- the largest and most detailed simulation of a major earthquake along the Southern San Andreas Fault—"The Big One"—a collaboration of 33 earthquake scientists and computer scientists from UC and elsewhere.

Today, SDSC is collaborating with some 1,600 UC researchers on more than 150 projects, using the Center's resources and expertise to help solve some of the most significant "grand challenges" facing science and society today as it continues its long-held quest "to turn data into discovery."



Resources and Expertise for UC Researchers

SDSC is partnering with UC researchers system-wide who are yielding significant discoveries across a variety of disciplines and fields of importance for science and society. The following section places a spotlight on a few such efforts in the following research areas:

- **"omics" for personalized medicine**
- **drug design for tomorrow's medicine chest**
- **nanoparticles and new materials for industry and medicine**
- **visions of galaxy formation**
- **delving inside the human brain to treat disorders**
- **modeling oceanic and coastal impacts from climate change**
- **harnessing light and energy from the sun**
- **tackling turbulence with high-performance computing**
- **probing the earth's dynamic inner space**

Mining "Omics" for Personalized Medicine

Since the 90's, researchers in the biological sciences have flocked to the so-called "omics" fields, genomics, proteomics, and metabolomics, among others, untangling the intertwined workings of DNA and proteins. Technological advances have led to high-throughput biology which allows rapid DNA sequencing, imaging, gene expression microarrays, and even genome-wide screening. The future of medicine is often proposed to be personalized treatment and prevention based on a patient's "omics" profile.

Such research generates huge data, a quantity that exceeds Moore's Law 100-fold just in terms of the amount of DNA being sequenced per day, said Mark Miller, principal investigator in SDSC's research, education, and development group. Gene sequencing is not like feeding a string into a machine that reads it – the DNA is chopped into small pieces, sequenced, then, in a computationally intensive process, reassembled to create a genetic map of the locations of genes, proteins, their functions, and the relationships between them. Such research requires a large amount of computing power and massive data storage.

Cancer Genome Hub

SDSC provides such an infrastructure, a high-speed network, data storage, and high-speed processing, for the new Cancer Genomics Hub (CGHub), managed by UC Santa Cruz. The CGHub serves as a large-scale data repository and user portal for the National Cancer Institute's (NCI) cancer genome research programs that will allow researchers to catalog all the genetic abnormalities found in different types of cancers, and to find connections between specific genetic changes and how patients respond to different treatments.

David Haussler, distinguished professor of biomolecular engineering at UCSC and leader of the CGHub project, designed and oversees the storage and computing infrastructure for the repository, which has an automated query and download interface for large-scale, high-speed use. Eventually, it will also include an interactive web-based interface to allow researchers to browse and query the system and download custom datasets from three major NCI cancer genome sequencing programs: The Cancer Genome Atlas (TCGA), Therapeutically Applicable Research to Generate Effective Treatments (TARGET), and the Cancer Genome Characterization Initiative (CGCI).

"TCGA is allowing us for the first time to look at cancer in full molecular detail," Haussler said. "Cancer is a disease caused by disruption of DNA molecules within the cell. When life starts, every cell in the body has the same DNA. In the course of a person's lifetime, however, some cells may accumulate changes in their DNA that cause them to go rogue and multiply without control, creating the disease we call cancer. For the first time now, we are able to look into an individual patient's cancer cells and see all the genetic disruptions, among which are the molecular drivers of that person's cancer."

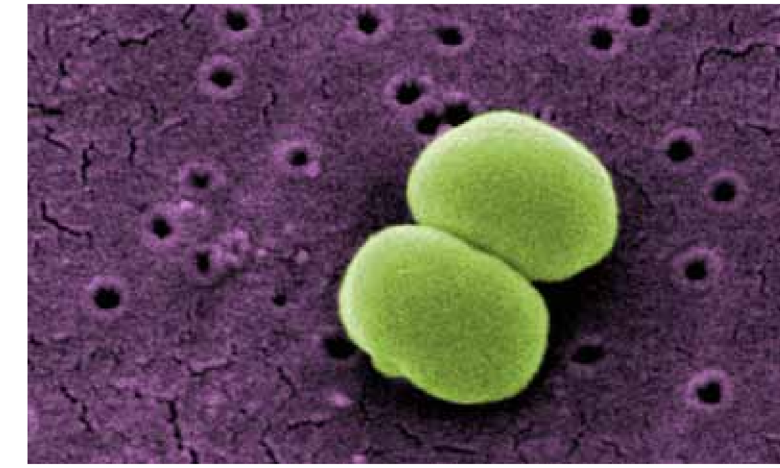
TCGA currently generates about 10 terabytes of data each month. By comparison, the Hubble Space Telescope amassed about 45 terabytes of data in its first 20 years of operation. TCGA's output will increase tenfold or more over the next two years. Over the next four years, if the project produces a terabyte of DNA and RNA data from each of more than 10,000 patients, it will have produced 10 petabytes of data. And TCGA is just the beginning of the data deluge, Haussler said, noting that 10,000 cases is a small fraction of the 1.5 million new cancer cases diagnosed every year in the United States alone.

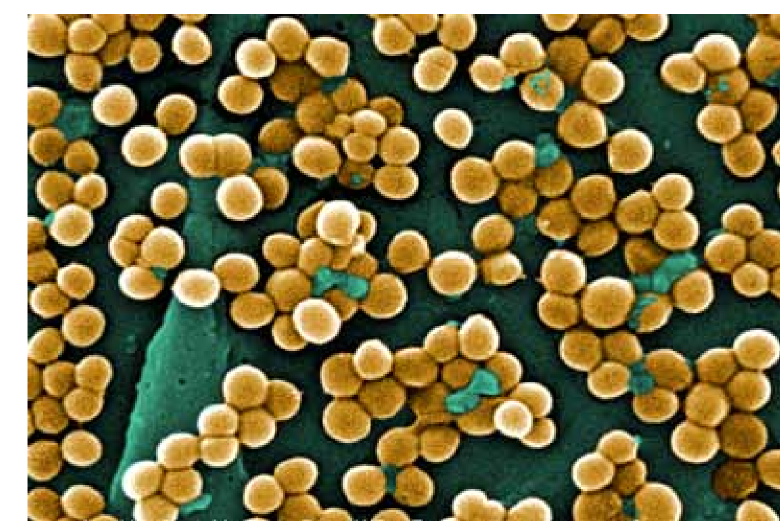
Efficient Genome Assembly

In order to submit data to a data resource such as the one curated by Haussler's group, it first needs to be organized. Sequencing technologies allow researchers to generate incredible amounts of data – billions of short gene sequences – so much that the algorithms struggle computationally to prepare a virtual scaffolding upon which to assemble the sequence into complete genomes.

Xifeng Yan, the Venkatesh Narayanamurti Chair of Computer Science at UC Santa Barbara, has been developing new algorithms to streamline the creation of such scaffolding, known as a De Bruijn graph. Using SDSC's *Trestles* system he demonstrated his algorithm, MSP for minimum substring partitioning, required five times less memory than previous methods.

MSP will be one part of a "pipeline" or a group of software that assembles entire genomes, each piece of the software doing one part of the job. Yan and his group are focusing on genomes that have not yet been assembled, such as that of Cladonema, a small jellyfish.





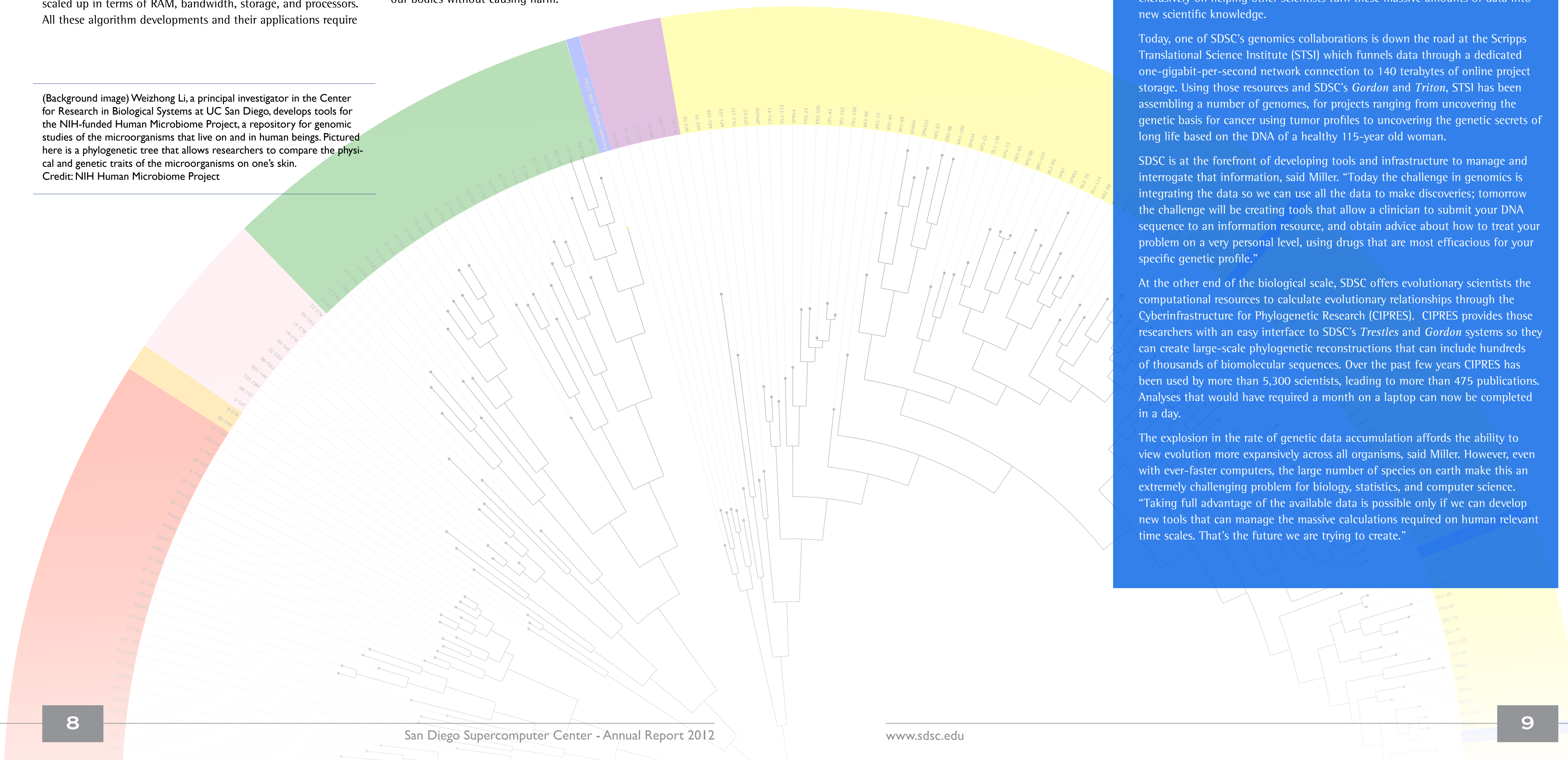
Staphylococcus epidermidis (top), Clostridium difficile (middle), and Staphylococcus aureus (bottom) are among the organisms that colonize human beings. The Human Microbiome Project aims to characterize the genomics of those organisms with the ultimate goal of finding out how changes in the microbiome are related to disease. Weizhong Li, a principal investigator in the Center for Research in Biological Systems (CRBS) at UC San Diego, develops tools for the project. Credit: Segrid McAllister, Janice Carr, Pete Wardell, and Jeff Hageman; Centers for Disease Control

Discovering Your Bugs

Once the DNA from an organism or from a clinical or environmental sample has been sequenced, the next step is for scientists to use computational tools to perform sequence and genomics analyses. Weizhong Li, a principal investigator in the Center for Research in Biological Systems (CRBS) at UC San Diego, develops such tools for the NIH-funded Human Microbiome Project (HMP), using SDSC's *Triton*. HMP aims to characterize the microbial communities found at several different sites on the human body, including nasal passages, oral cavities, skin, gastro-intestinal tract, and urogenital tract, and to analyze the role of these microbes in human health and disease.

"Computationally, the greatest challenge in analyzing metagenomics data is robust computer resources that can be reasonably scaled up in terms of RAM, bandwidth, storage, and processors. All these algorithm developments and their applications require extremely high-end computational resources that are reliable, scalable, fast, and also affordable," said Li. "*Triton* has enabled us to perform challenging tasks in several large projects."

The data for the project is generated by technology that produces millions to billions of sequences at once. Li's goal is to create the tools used to analyze that sequence for various kinds of genomic, metagenomic, and transcriptomic studies of microorganisms. So far his group has developed cd-hit, a program widely used for sequence clustering; fr-hit, a sequence mapping program for metagenomics; and MGAviewer to analyze metagenomics alignment. These tools help to reveal not only the identity and composition of the sample but, more importantly, also the variations between samples and references that might signal why some microganisms cause disease while others inhabit our bodies without causing harm.



(Background image) Weizhong Li, a principal investigator in the Center for Research in Biological Systems at UC San Diego, develops tools for the NIH-funded Human Microbiome Project, a repository for genomic studies of the microorganisms that live on and in human beings. Pictured here is a phylogenetic tree that allows researchers to compare the physical and genetic traits of the microorganisms on one's skin.
Credit: NIH Human Microbiome Project

Mark Miller, principal investigator in SDSC's Research, Education, and Development group, said SDSC is at the forefront of developing tools to manage and interrogate the outpouring of data from the "omics" fields.

Assembling Resources for Genomics

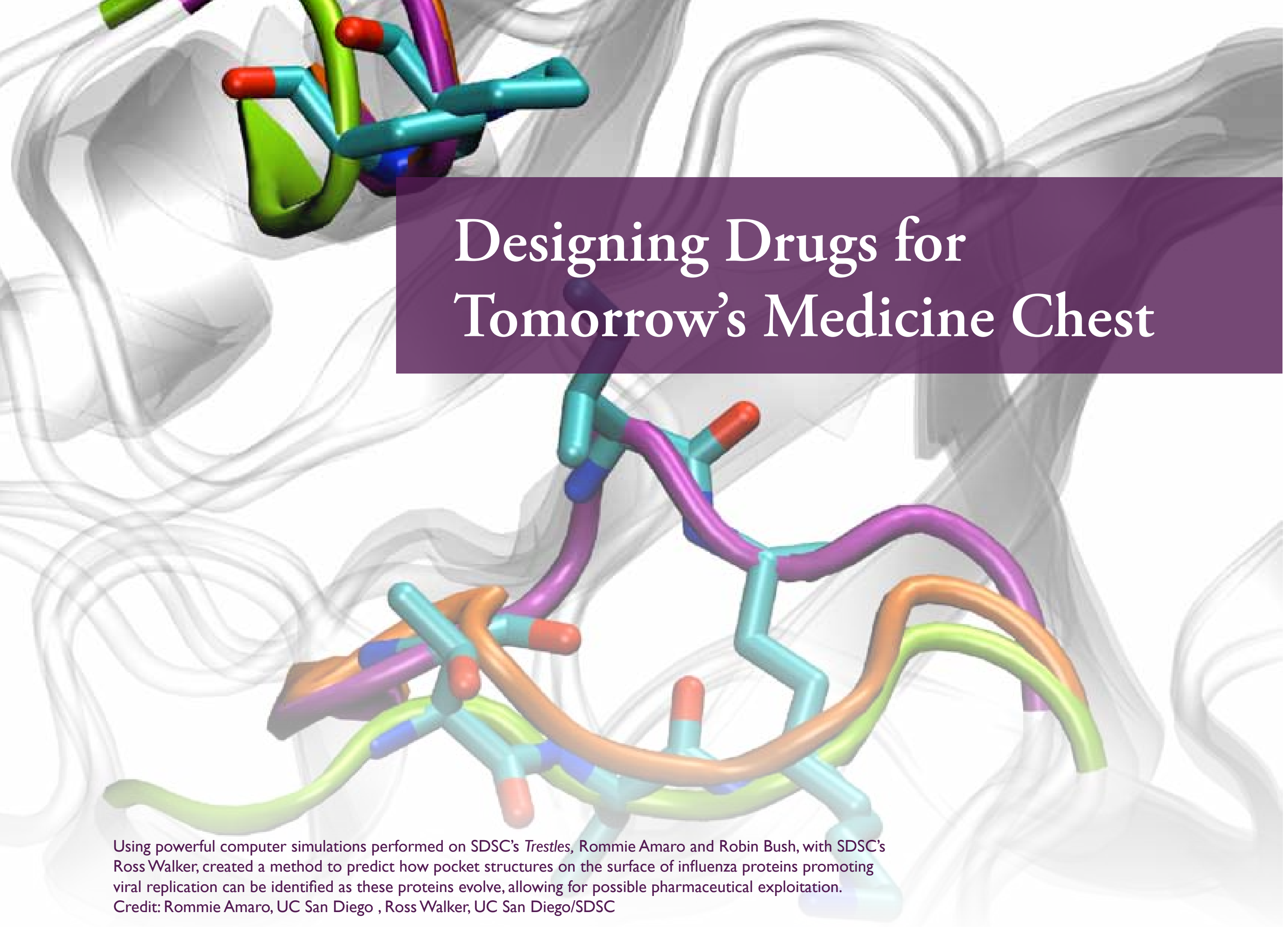
When Mark Miller, principal investigator in SDSC's Research, Education, and Development group, started out as a scientist, he followed his interests in biochemistry and structural biology and dabbled in genetics. Genomics, he admits, was not yet a field in its own right. Computationally-intensive biology was limited to molecular simulations, and the concept of data-intensive computing did not even exist. This was a far cry from today's world where DNA sequences, mass spectra, and many other types of biological data are produced at the rate of petabytes per year. Over the past few years, Miller's work has focused almost exclusively on helping other scientists turn these massive amounts of data into new scientific knowledge.

Today, one of SDSC's genomics collaborations is down the road at the Scripps Translational Science Institute (STSI) which funnels data through a dedicated one-gigabit-per-second network connection to 140 terabytes of online project storage. Using those resources and SDSC's *Gordon* and *Triton*, STSI has been assembling a number of genomes, for projects ranging from uncovering the genetic basis for cancer using tumor profiles to uncovering the genetic secrets of long life based on the DNA of a healthy 115-year old woman.

SDSC is at the forefront of developing tools and infrastructure to manage and interrogate that information, said Miller. "Today the challenge in genomics is integrating the data so we can use all the data to make discoveries; tomorrow the challenge will be creating tools that allow a clinician to submit your DNA sequence to an information resource, and obtain advice about how to treat your problem on a very personal level, using drugs that are most efficacious for your specific genetic profile."

At the other end of the biological scale, SDSC offers evolutionary scientists the computational resources to calculate evolutionary relationships through the Cyberinfrastructure for Phylogenetic Research (CIPRES). CIPRES provides those researchers with an easy interface to SDSC's *Trestles* and *Gordon* systems so they can create large-scale phylogenetic reconstructions that can include hundreds of thousands of biomolecular sequences. Over the past few years CIPRES has been used by more than 5,300 scientists, leading to more than 475 publications. Analyses that would have required a month on a laptop can now be completed in a day.

The explosion in the rate of genetic data accumulation affords the ability to view evolution more expansively across all organisms, said Miller. However, even with ever-faster computers, the large number of species on earth make this an extremely challenging problem for biology, statistics, and computer science. "Taking full advantage of the available data is possible only if we can develop new tools that can manage the massive calculations required on human relevant time scales. That's the future we are trying to create."



Designing Drugs for Tomorrow's Medicine Chest

Using powerful computer simulations performed on SDSC's *Trestles*, Rommie Amaro and Robin Bush, with SDSC's Ross Walker, created a method to predict how pocket structures on the surface of influenza proteins promoting viral replication can be identified as these proteins evolve, allowing for possible pharmaceutical exploitation. Credit: Rommie Amaro, UC San Diego , Ross Walker, UC San Diego/SDSC

The medicines of tomorrow will be generated at the interface of chemistry, biology, physics and pharmacology. One of the areas generating the most interest is the development and application of computational techniques to mining biological systems for possible drug discovery.

Ensemble Approach

Scientists at that interface have been able to model increasingly complex biological systems, including the identity of targets within a virus or bacteria where a drug might enter and shut it down. Rommie Amaro, assistant professor of chemistry and biochemistry at UC San Diego, dubbed "The Protein Explorer" in a recent issue of *The Scientist*, applies physics-based methods to her models, which allows her to increase their complexity and to offer more opportunities to find those targets.

Amaro's work using SDSC's *Trestles* and *Gordon* systems, which includes investigations of African sleeping sickness, bird flu, and disease-causing Yersinia bacteria, has helped generate potential drug targets. Those targets are then screened against libraries of compounds that might fit those targets. Once the potential drugs have been found, Amaro works with experimentalists who try out the compounds in the lab.

Steps to Stopping HIV

Many biological processes involve a molecule hopping through a series of ligand interactions, each step a chemical handshake that alters the molecule in some way to prepare it for the next step. Myungshim Kang, a postdoc in assistant chemistry professor Chia-en Angelina Chang's lab at UC Riverside, models ligands starting from the initial stages of the binding process to the final steps where the complexes stabilize.

Using SDSC's *Trestles* system, Kang focused on HIV-1 protease and its two neutral inhibitors, a major target for AIDS treatments due to the protease's role in the virus' replication cycle. Without the protease, HIV is unable to create the building blocks it needs to assemble additional infectious viruses. An HIV-stopping drug targeting reproduction has to mimic the natural substrates of HIV-1 protease in shape and action to take its place.

Kang first performed the Brownian dynamics simulations with the coarse-grained model of the system of HIV-1 protease and its inhibitors, creating a picture of the diffusional movements and the resulting encounter complexes. This provides the most likely formations that she then applies to a molecular dynamics simulation. This two-step approach reveals how the ligands and the protease interact and lock together with the second step detailing the critical roles of water and hydrogen bonding as the complexes approach their most stable states.

Getting the Gist of Water

Determining the role of water is an essential, elusive key to a molecular understanding of biological processes. When a ligand approaches a binding site on another molecule, chances are it needs to push water molecules aside in order to dock. These water molecules have their own surface energy and entropy and affect the overall thermodynamics of binding.

In a paper featured on the cover of *The Journal of Chemical Physics*, Michael Gilson, professor of pharmacy at UC San Diego, and Crystal Nguyen, a postdoc in his lab, and their colleague Thomas Kurtzman Young, assistant professor of chemistry at the City University of New York, simulated a small synthetic receptor to observe the properties of water around its binding sites. They created a version of inhomogenous solvation theory (GIST) that improves on earlier versions by placing the model into a three-dimensional grid and accounting for the displacement of water.

Using SDSC's *Trestles* system, they simulated an artificial pumpkin-shaped receptor, CB7—and the water within its rounded central cavity—and found an unlikely area where water contributes to the ability of the receptor to tightly bind ligands.

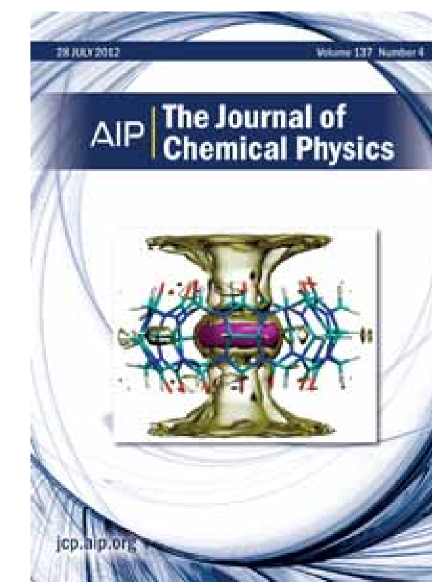
The GIST analysis is currently being applied to more complex protein structures and can ultimately serve as a tool for guiding the design of ligands. It may be the key to solving a common problem in structure-based drug design of whether to incorporate water, to plan on its displacement, or to avoid it altogether.

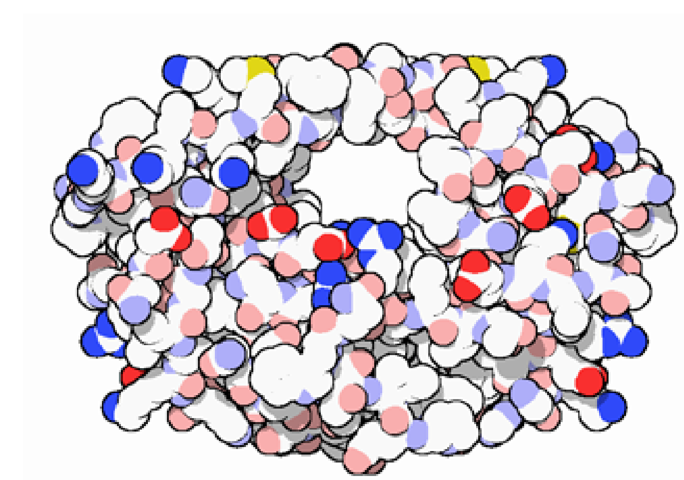
Designer Enzymes

Natural enzymes are large, complex proteins that control the reactions that sustain life. They play a central role in transforming food into the essential nutrients that provide energy, among other critical functions. A pioneer in applying computational methods to organic chemistry, Kendall Houk, UC Los Angeles' Saul Winstein professor of organic chemistry, uses large-scale computational studies to study enzymes including the creation of "designer enzymes," those that will catalyze non-natural reactions. In other words, they build synthetic enzymes that do jobs that naturally occurring enzymes will not do.

Winstein and his research group used SDSC's *Gordon* and *Trestles* systems to perform quantum mechanical calculations that allow him to explore pericyclic reactions and organocatalysis, molecular dynamics, organometallic chemistry, enzyme design, and biological catalysis.

In a paper featured on the cover of *The Journal of Chemical Physics*, Michael Gilson, professor of pharmacy at UC San Diego, and his colleagues simulated a small synthetic receptor to observe the properties of water around its binding sites. Credit: Michael Gilson, Crystal Nguyen, UC San Diego and Thomas Kurtzman Young, City University of New York





HIV-1 protease as illustrated by David Goodsell, UC San Diego, for the RCSB Protein Data Bank. Using SDSC's *Trestles* system, Myungshim Kang, a postdoc in assistant chemistry professor Chia-en Angelina Chang's lab at UC Riverside, focused on HIV-1 protease and its two neutral inhibitors, a major target for AIDS treatments due to the protease's role in the virus' replication cycle. An HIV-stopping drug targeting reproduction has to mimic the natural substrates of HIV-1 protease in shape and action to take its place. Credit: David Goodsell, UC San Diego, and the RCSB Protein Data Bank

Developing Nanoparticles and Novel Materials for Industry and Medicine

Transverse wireframe view of a bone microstructure with small cavities, or lacunae, each containing an osteocyte—the most abundant cell found in bone. Researchers study bone microstructures for clues to better predict bone fracture in patients, including those with osteoporosis. Credit: Maria-Grazia Ascenzi, UCLA

Nanoparticles, tiny technological wonders typically ranging in size from one to 100 nanometers in diameter, are now found in almost every sector: energy, information technology, homeland security, medicine, and transportation, as well as an array of consumer goods ranging from food and clothing to sporting goods and vehicle components.

Today, about 800 everyday commercial products rely on nanoscale materials and processes, according to the National Nanotechnology Initiative (NNI), a multi-agency U.S. government program that coordinates nanotechnology efforts.

Nano-bioengineering of enzymes, for example, is aimed at efficiently converting cellulose into ethanol for fuel from wood chips, corn stalks, and varieties of unfertilized perennial grasses. Nanotechnology is being used in new kinds of batteries—including those for cars—as they become more efficient, lighter weight, and more quickly recharged. Researchers are developing wires containing carbon nanotubes that have much lower resistance than the high-tension wires currently used in our electric power grid, thereby reducing transmission power losses.

Most of these stronger, lighter, more resilient, or energy-efficient materials are being conceived and tested in the lab with the help of simulations made possible by high-performance computing, like those housed at SDSC.

Novel Uses for Glassy Nanostructures

At UC Merced, Assistant Professor Lilian Dávila and her team have been combining simulations, theory, and experiments to quantify size-dependent properties of silica (SiO_2) glass helical nanostructures and silica nanowires, due to their unique properties and potential applications that range from optical communications to chemical and biological sensors for microsurgery. Dávila's group has been using SDSC's *Triton* supercomputer—along with a powerful and innovative 3D-visualization system with accelerated molecular dynamics (MD) simulation capabilities, and Finite Element Method (FEM) analysis—to conduct highly detailed modeling and simulations of these flexible nanostructure materials.

This research is advancing our fundamental understanding in materials science and engineering, says Dávila, who was awarded a UC President's Postdoctoral Fellow at UC Merced and the Lawrence Livermore National Laboratory (LLNL) before establishing the first computational materials science and engineering research group at UC Merced. "We also are promoting collaborations and diverse educational programs in computational simulation and nanotechnology, and contributing to efforts by the National Science Foundation's Materials Genome Initiative to meet pressing societal challenges in energy, environment, and sustainability."

Going "Green" at the Nano-level

In water-based solutions, metal-oxides can be synthesized to form highly stable clusters with nanometer-scale dimensions. These clusters are of great interest as precursors for "greener" routes to the synthesis of oxide thin films—that is, via low energy-cost routes using reagents of high geologic abundance and low toxicity. Such films find applications in a wide variety of technologies, ranging from microelectronics to catalysis.

UC Berkeley Professor Mark Asta and postdoctoral research Benjamin Hanken are using SDSC's *Trestles* cluster to employ the methodology of *ab-initio* molecular dynamics simulations to the study of the structure and dynamics of oxide clusters in water. These simulations are being performed in close collaboration with Professor William Casey's experimental group at UC Davis. The research, funded through the National Science Foundation's (NSF) Center for Sustainable Materials Chemistry, employs computational simulations to augment experimental data, and will ultimately provide a detailed understanding of the stability and reactivity of pure and mixed metal-oxide clusters within water.

Understanding Our Ever-Changing Bones

Bone is an organ with the complexity of a body system that changes throughout life. The role of the microstructure of human bone with respect to bone aging and its higher risk to fracture is becoming better understood, thanks to advances in supercomputer simulations and new modeling techniques. Maria-Grazia Ascenzi, director of the bone micro-biomechanics laboratory in the Rehab Building at UCLA, is using SDSC's *Trestles* supercomputer to further investigate this role using both proven and experimental techniques in response to scientists throughout the clinical and research communities who now regard bone as a multi-level structure.

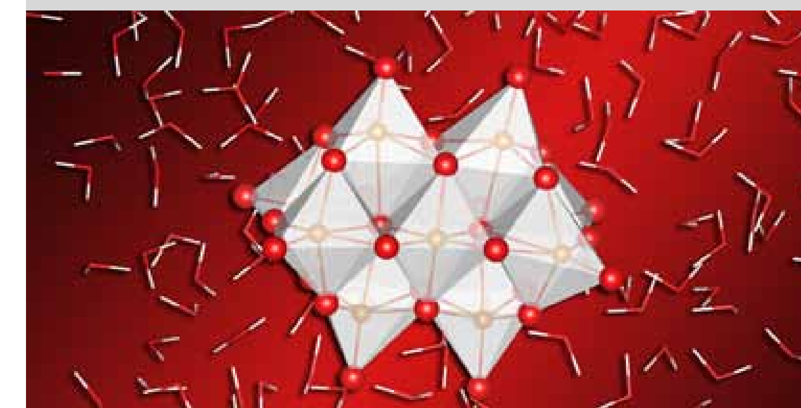
Ascenzi is using *Trestles* to develop a larger pool of modeling data that synthesizes various elements of human bone biology obtained from micro-and macro-mechanical testing, microdissection, and a variety of imaging technologies that include optical microscopy under ordinary and polarized light, electron microscopy, confocal microscopy, high resolution micro-X-rays, and more. By applying theoretical mathematics to biological contexts to explain biomechanical behavior observed in the laboratory, Ascenzi aims to provide more effective clinical interventions by translating the clinical factors of fracture risk into mechanical evaluation of a patient's bone tissue, while advocating further development of advanced engineering modeling techniques and their application to biological composite materials.

This kind of research may help scientists develop new 'smart' materials, which can be changed in a controlled manner by external stimuli and used in place of fractured areas to help individuals maintain or even improve their mobility.

Optimizing Simulation Algorithms

First-principles simulations provide a very detailed computational model of materials properties. They are based on fundamental (quantum) theory and therefore do not rely on empirical data, or on the results of experiments. This gives them the potential ability to predict materials properties by running simulations on computers, rather than by synthesizing materials and measuring their properties in a laboratory. Francois Gygi, a professor with UC Davis' Department of Computer Science, has been using SDSC's *Trestles* supercomputer to further develop first-principles simulation algorithms to extend the range of properties that can be predicted through computer-based simulations, and to accelerate such simulations by using larger and more data-intensive supercomputers efficiently.

As computer architectures continually evolve, algorithm design must be adapted to maximize the efficiency of simulation software. *Trestles* was used in this project as being representative of a new trend of computer architecture in which each node includes a large number of processor cores.



A configuration of the modeled metal-oxide cluster immersed in circa 200 water molecules. Credit: Benjamin Hanken, UC Davis

Creating Visions of Galaxies

The Hubble Space Telescope captured the merger the Antennae Galaxies hundreds of millions of years ago. Researchers at UC Riverside have been running simulations on *Trestles* to capture new details about the evolution of such mergers over billions of years. Credit: NASA, ESA, and the Hubble Heritage Team (STScI/AURA)-ESA/ Hubble Collaboration

Astrophysicists use computational simulations to investigate the cosmos in ways that cannot be studied using observations alone. Such models allow researchers to investigate how stars interact, to find hidden galaxies, and even to study the evolution of the universe over time.

Stellar Velocity During Galaxy Mergers

As galaxies travel through the universe, inevitably they collide and occasionally merge. As they approach one another, they begin to interact gravitationally in such a manner that they eventually unite into a single galaxy. Gabriela Canalizo, associate professor of physics and astronomy at UC Riverside studies the changing velocities of stars during galaxy mergers by running simulations that include cooling, stellar formation and the interactions of galaxy centers and supernovae.

One of the key measurements for understanding a galaxy merger is the velocity dispersion, or the degree of "randomness" of the stellar velocities. In a system with low velocity dispersion, the stars move in the same direction with very similar speeds, whereas stars in a system with high velocity dispersion move in different directions at very different speeds. Canalizo and Nathaniel Stickley, a graduate student in her research group, used SDSC's *Gordon* system to generate simulations of a merger in five-million year intervals over four-to-six billion years. From that model Stickley has been working to interpret the stellar velocity dispersion during a merger for a paper to be submitted to the *Astrophysical Journal*.

"This evolution cannot be observed in the actual universe because we can't watch galaxies merge in real time," said Stickley. "We would have to wait billions of years from the beginning to the end of a merger, which clearly isn't possible." The velocity dispersion studied by Stickley and Canalizo is an important quantity in galactic astronomy because it correlates strongly with the gravitational potential present in the galaxy, which is something that cannot be observed directly.

Stellar Feedback in Dwarf Galaxies

Indirect observations are commonplace in astrophysics, where phenomena are often discovered and measured by the forces they exert on on their surroundings. Observable only by the bending of light from other objects, dwarf galaxies, which can be found orbiting large galaxies and are considered to be their building blocks, should be common... yet they are difficult to detect. These small galaxies contain just a few billion stars compared to the 200-400 billion in the Milky Way.

James Bullock, professor of physics and astronomy at UC Irvine, and his colleagues are using SDSC's *Gordon* system to create a zoom-in cosmological simulation of dwarf galaxies. His model will investigate the role stellar physics in the supernova explosions and stellar winds play during that process. By creating a more realistic stellar feedback model they can more easily connect stellar physics with galaxy formation models.

(Above images) Gabriela Canalizo, associate professor of physics and astronomy at UC Riverside and her postdoc Nathaniel Stickley study the changing velocities of stars during galaxy mergers. This simulated sequence allows the scientists to measure the velocity dispersion, or the degree of "randomness" of the stellar velocities. Credit: Nathaniel Stickley and Gabriela Canalizo, UC Riverside

(Background image) The Bolshoi simulation generated the most accurate and detailed large cosmological simulation run to date, a truly massive amount of data spanning billions of year. To even examine a slice of the data such as this requires the services of a high performance system such as *Gordon*. Credit: Visualization by Stefan Gottlober, Leibniz-Institut fuer Astrophysik Potsdam (AIP)

Supersized Snapshot

Last year, Joel Primack, a distinguished professor of physics at UC Santa Cruz and director of the UC Systemwide High-Performance Astro-Computing Center (UC-HiPACC) and his colleagues including Anatoly Klypin, professor of astronomy at New Mexico State University, created the most realistic simulation of the evolving from the Big Bang to now. The simulation, named Bolshoi, ran on NASA Ames' *Pleiades* cluster for 18 days straight on over 23,552 processors, consumed millions of compute hours, and generated data on a cosmological scale. A subsequent model, called BigBolshoi/MultiDark of a larger area was also generated, though at a lower resolution than Bolshoi.

Bolshoi and BigBolshoi/MultiDark provide researchers with a powerful new tool for understanding the evolution of both the visible and invisible elements of the universe. However, the models are so large that investigating them requires wielding the power of a resources powerful enough to look at a slice of the model universe. Primack and his colleagues used SDSC's *Gordon* system to take snapshots of the data they had created. Their goal is to measure the cosmic web—the large-scale structure of the universe—using a method that incorporates velocities as well as locations. They also plan to investigate the shapes and orientations of dark matter halos, which play a role in the formation and evolution of galaxies.

Managing Astronomy's "Data Deluge"

Today's telescopes and digital sky surveys are generating huge amounts of new data about the universe—in some cases, by the second. Indeed, astronomers were among the first researchers to face the onrushing tidal wave of the data deluge, spurring investigations on how best to capture and analyze this information for discovery.

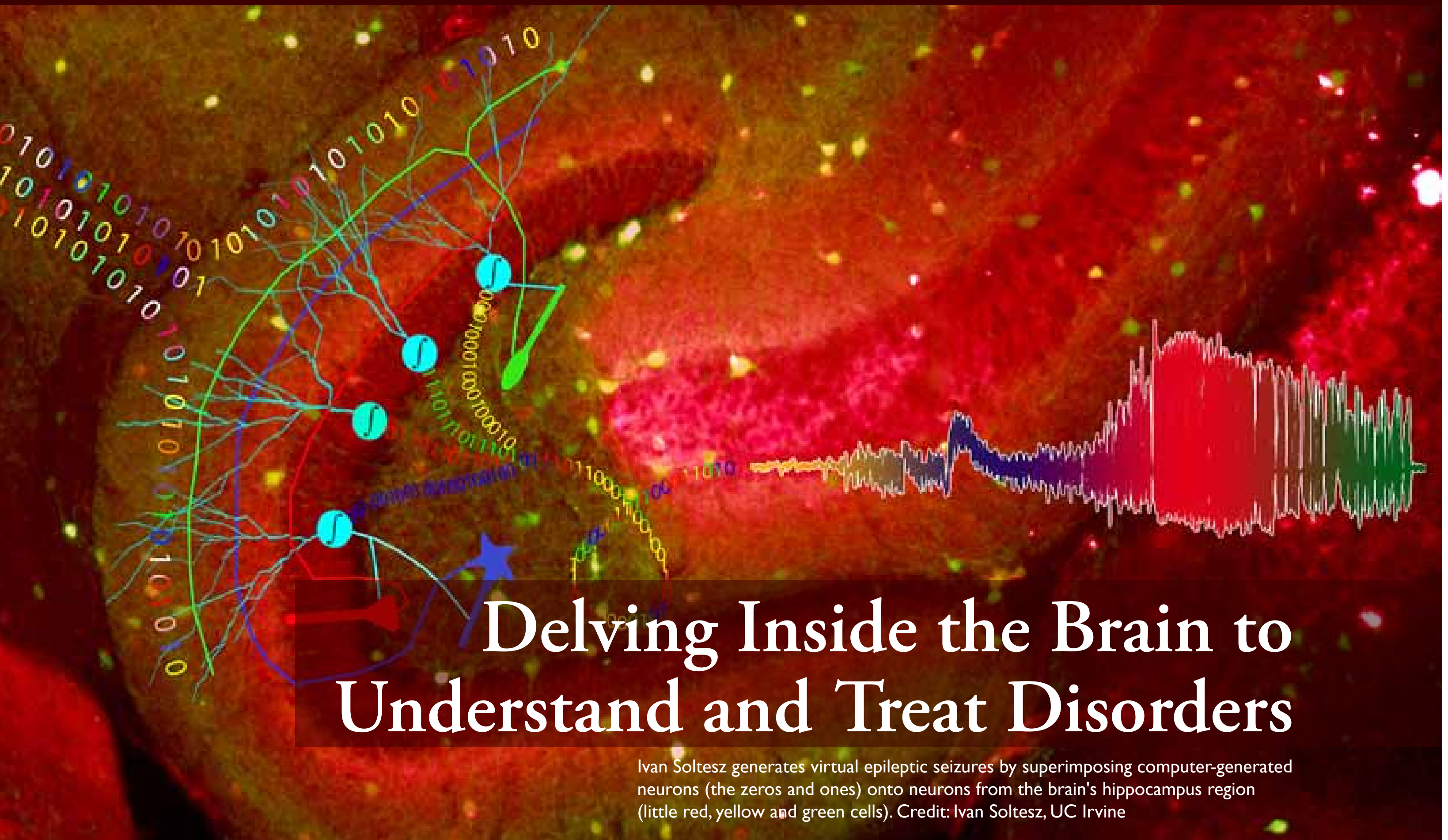
Last July, SDSC—in conjunction with the University of California's High-Performance AstroComputing Center (UC-HiPACC)—hosted a two-week long summer school designed to help the next generation of astronomers manage "Big Data" generated by astronomy's multi-faceted instruments and computer simulations.

As described by Joel R. Primack, director of UC-HiPACC—a consortium of nine UC campuses and three Department of Energy (DOE) laboratories with its base at UC Santa Cruz—the summer school was designed to "empower astronomers to compare massive observational data with massive theoretical outputs."

A key feature of this summer school was access by all students to powerful supercomputers, including SDSC's data-intensive supercomputer, *Gordon*, which contains several relevant astronomical data sets and simulations. The school program was open to graduate students and post-doctoral fellows.

"Astronomers will need to know how to leverage the capabilities of data-intensive supercomputers to analyze all this data efficiently while bringing these observations and simulations into a common framework," said Michael Norman, director of SDSC and a world-renowned astrophysicist.

HiPACC does not directly fund research or major hardware. Instead, it supports activities to facilitate and encourage excellence and collaboration in astronomy across the UC system.



Delving Inside the Brain to Understand and Treat Disorders

Ivan Soltesz generates virtual epileptic seizures by superimposing computer-generated neurons (the zeros and ones) onto neurons from the brain's hippocampus region (little red, yellow and green cells). Credit: Ivan Soltesz, UC Irvine

Computational studies have allowed scientists to probe the universe, to inspect violent weather phenomena and, closer to home, to examine that most-studied and least understood part of the body—the brain. Limited research can be done without interrupting or damaging its processes, leaving models and simulation as the route to delving into the brain to understand how it works, and what happens when it does not, as with epilepsy and Parkinson's disease.

Pinpointing Structure

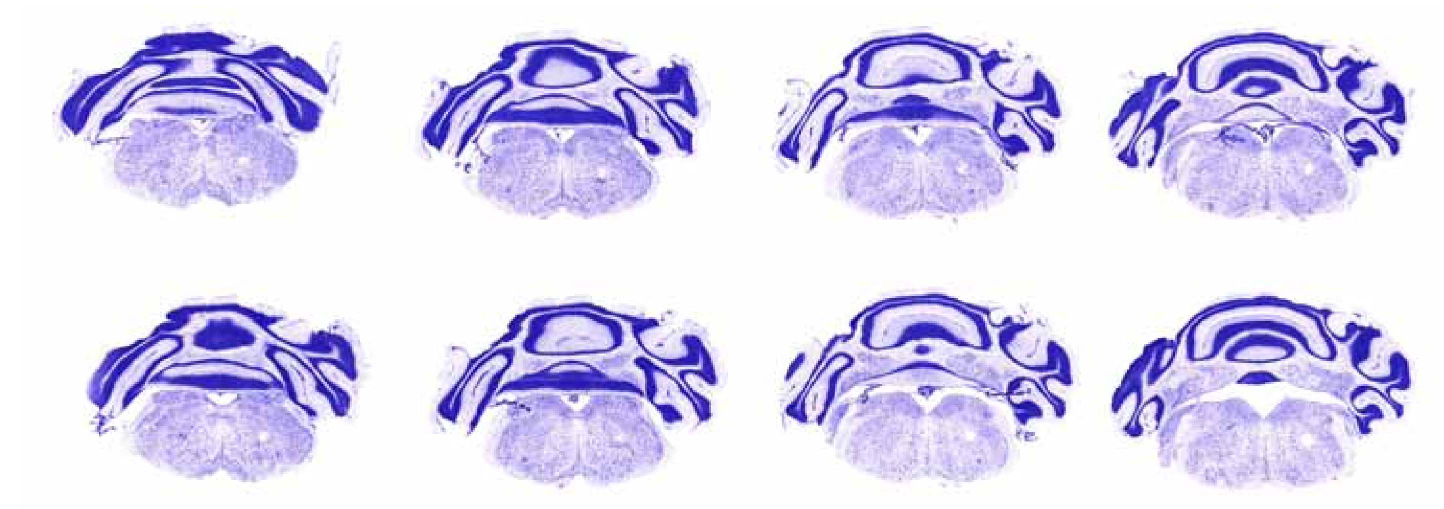
Not all brains are the same, and a tool that neuroanatomists have been lacking is an easy way to pinpoint a location in a particular brain and to compare it with the same location in other brains. Yoav Freund, professor of computer science and engineering at UC San Diego, has been developing a tool to analyze "stacks" of two-dimensional mouse brain images to create a reference that allows researchers to call up the same location from numerous individuals at once.

Each "stack" of images provided by David Kleinfeld, professor of physics and neurobiology at UCSD, contains between 30 to 80 slices of a single brain and consumes between 30 to 100 gigabytes of memory, generating a data-intensive problem. With the assistance of graduate student Yuncong Chen, Freund has been using SDSC's *Gordon* to computationally compare the difference between each consecutive image in a stack. This information will be captured in a database that can be used to align other image stacks. Adapting the work for parallelization will be a next step that allows new data to be analyzed and added to the reference much more quickly.

Modeling Epilepsy

Epilepsy affects more than 50 million people worldwide, and while there are treatments that alleviate some of its some symptoms, there is no cure. Understanding why epilepsy happens and what causes it will eventually lead to seizure-stopping techniques. Ivan Soltesz, Chancellor's Professor and chair of anatomy and neurobiology at UC Irvine, studies how brain cells communicate and how that communication changes in epileptics using both experimental and computational techniques.

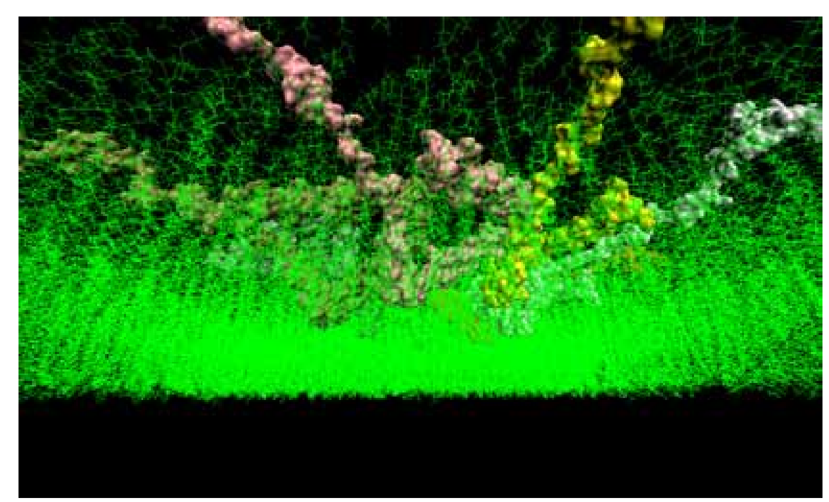
The hippocampus is a structure deep in the brain that is ordinarily involved in memory and spatial navigation but also happens to be a particularly vulnerable site for epileptic seizures.



(Above) Yoav Freund, professor of computer science and engineering at UC San Diego, has been developing a tool to analyze "stacks" of two-dimensional mouse brain images to create a reference that allows researchers to call up the same location from numerous individuals at once. Credit: Yoav Freund, Yuncong Chen, David Kleinfeld, UC San Diego

It consists of layers of neurons, tidily organized, that make it amenable as a model for neurophysiologists. Marianne Bezaire, a graduate student Soltesz' lab, studies the functions of interneurons in a portion of the hippocampus called the CA1 using a detailed, biologically realistic computer model of the rat brain.

To generate her model with a degree of accuracy that allows her to pinpoint how neurons interact involves simulating systems from 350 to 350,000 cells on SDSC's *Trestles* system. Knowing how the brain is connected, said Case, allows her to examine how and why the CA1 changes during epilepsy, with certain cells dying off or sprouting new connections. In particular, she hopes her model will answer whether those changes result in altered feed-forward inhibition—where interneurons are activated before the excitatory cells they inhibit.



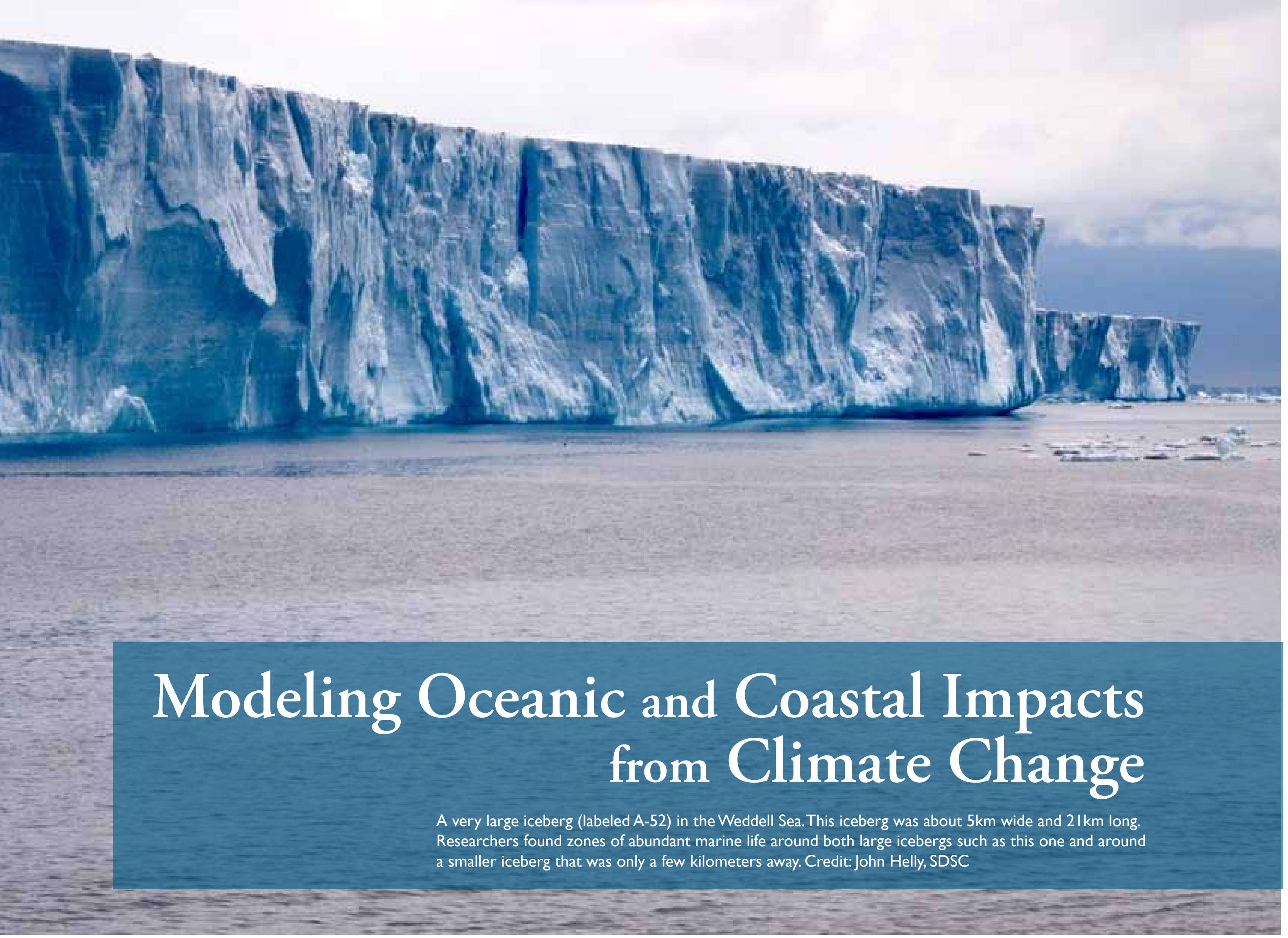
Simulations of Parkinson's Disease

For a long time the prevailing theory of Parkinson's disease was that an aggregation of alpha-synuclein proteins in the brain damaged neurons, reducing their function or killing them altogether. By using computing resources at SDSC and Argonne National Lab to study the steps leading to aggregation of those proteins Igor Tsigelny, SDSC researcher and senior program project scientist in neurosciences department at UC San Diego, reported that the damage may be caused earlier, before the proteins have a chance to aggregate to long fibrils.

According to his model, which was featured on the cover of the *FEBS Journal*, the protein alpha-synuclein penetrates the exterior membrane of a brain cell, and forms a ring structure that breaches the cell membrane. In the case of people with a familial version of Parkinson's, the protein is mutated and more severely punctures the neural membrane and forms a ring more quickly than the normal form. The modeling results are consistent with the electron microscopy images of neurons in Parkinson's disease patients; the damaged neurons are riddled with those ring structures.

Knowing the structure is a first step toward developing drugs to block the assembly of the ring structure. "Thanks to the power of supercomputing, we are making major progress in understanding the origins of Parkinson's disease and developing ways to treat it," said Tsigelny.

This molecular dynamics simulation shows one possible position of a ring of proteins penetrating a cell membrane. Such rings are thought to damage neurons, leading to Parkinson's and other degenerative brain diseases. Credit: Igor Tsigelny, UC San Diego



Modeling Oceanic and Coastal Impacts from Climate Change

A very large iceberg (labeled A-52) in the Weddell Sea. This iceberg was about 5km wide and 21km long. Researchers found zones of abundant marine life around both large icebergs such as this one and around a smaller iceberg that was only a few kilometers away. Credit: John Helly, SDSC

Researchers are concerned how the rise in greenhouse gases affects ocean currents, which is of great importance not just to the coastal nations of the world, but our entire planet. Covering more than 70 percent of the earth's surface, oceans are a vital component of our climate system, whose ever-changing composition and chemistry affects the atmosphere and our eroding polar ice caps.

Thanks in part to advances in visualization and computer modeling techniques, our understanding of oceanic circulation has improved during the past decade, offering scientists the ability to create and analyze highly detailed images that take into account a diverse range of influencing factors.

James McWilliams, Louis B. Slichter Professor of Earth Sciences at UCLA's Institute of Geophysics and Planetary Physics and Department of Atmospheric and Oceanic Sciences, is using SDSC's *Gordon* and *Trestles* systems to conduct computational studies of oceanic circulation that include water quality, biogeochemistry, sediment movement, and planktonic ecosystem population dynamics along the coastal regions of the Americas, Africa, and Asia.

One of McWilliams' goals is to devise realistic simulations to better understand how small-scale processes contribute to regional and climate equilibria and variability. By focusing on computational simulation and data analysis for theory and the validation of models, he and his group want to further improve the realism of these simulations by continually refining the models. That, in turn, may lead to new discoveries and explanations about how climate changes affect ocean currents.

California's Coastal Waters

SDSC resources are being used to specifically track and monitor changes along California's coastal and inland waters, which are crucial to ensuring a reliable, sustainable water supply for the state. In recent years, for example, observations in South San Francisco Bay, a shoal-channel estuary, have demonstrated the importance of small- and intermediate-scale motions to estuarine circulation and transport.

Mark Stacey, an environmental engineering professor at UC Berkeley, is using SDSC's *Trestles* supercomputer to perform a series of simulations to create a high-resolution, three-dimensional data set to better understand changes in the bay due to climate change and rising sea levels.

Stacey and his colleagues are also interested in finding ways to restore perimeter marshes and wetlands in the area. Their goal is to develop a predictive modeling tool that will have important implications for coastal and estuarine modeling throughout the world, while helping California's water management organizations explore future scenarios and make effective, environmentally sound decisions. (See more California-centric projects on pages 34 and 35.)

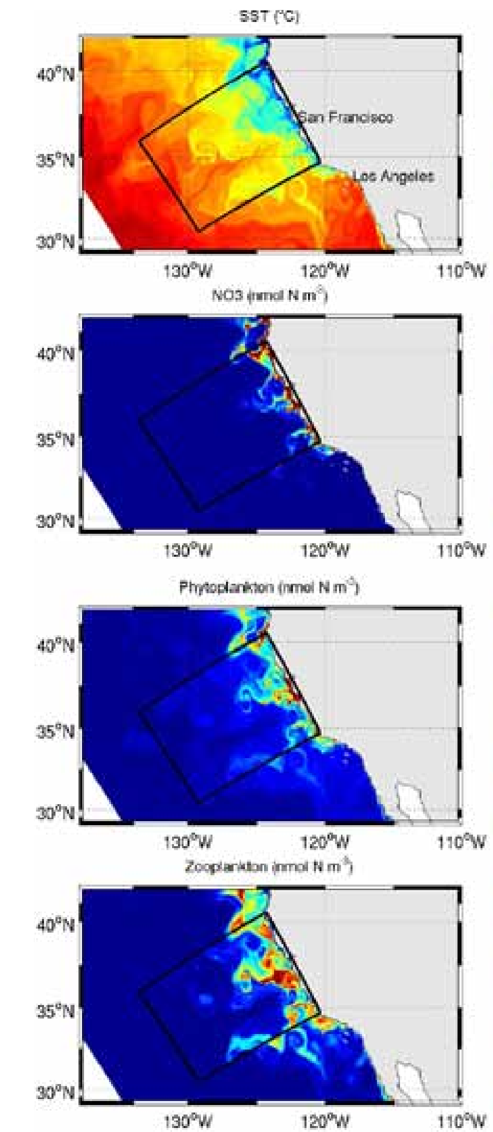
Another Kind of Cloud Computing

Clouds are now considered key to understanding climate change phenomena, such as rising temperatures and greenhouse gases. Clouds play a complex role in moderating incoming and outgoing solar radiation, as well as the transport of heat and water from the surface to the atmosphere and back. However, the inadequate representation of clouds in general circulation models (GCMs) has significantly limited scientists' ability to more accurately model historical and future climate studies.

The spatial resolution of traditional GCMs is typically on the order of about 100 kilometers (km), while cloud systems require spatial resolution on the order of one to four km. This puts the computing requirements for global cloud-resolving model runs at the upper end of supercomputing. Only the largest supercomputers can run such models, and only then for very short periods of time, not the hundreds or even tens of years needed to conduct climate studies.

John Helly, a researcher with SDSC and UC San Diego's Scripps Institution of Oceanography, is using SDSC's *Gordon* supercomputer, among others, to conduct a series of simulations and analyses using a new parameterization of cloud microphysics. Helly is running the simulations in the Multi-scale Modeling Framework (MMF), a relatively new approach to climate modeling that focuses on cloud-scaling processes.

The project is part of a larger, ongoing effort with the Center for Multi-scale Modeling of Atmospheric Processes (CMMAP), an NSF Science and Technology Center whose goal is to develop a new kind of global atmospheric model that represents the effects of clouds on weather and climate.



Instantaneous surface distributions of temperature (SST), the nutrient nitrate (NO3), phytoplankton abundance, and zooplankton abundance off central California. The evident eddy fluctuations regulate the biological productivity rate on top of the general coastal upwelling of subsurface nutrients into the euphotic zone. Credit: James McWilliams, UCLA

Manmade Pollutants May Be Widening the Earth's Tropical Belt

Black carbon aerosols are tiny particles of carbon produced from biomass burning and incomplete combustion of fossil fuels. Most of the world's black carbon production occurs in the Northern Hemisphere, with Southeast Asia being a major producer. The same is true of tropospheric ozone, a secondary pollutant that results when volatile organic compounds react with sunlight.

Climatologist Robert Allen, an assistant professor of earth sciences at UC Riverside, has been studying the effects of these pollutants on the earth's oceans and atmosphere, and the news is not good. Allen and his colleagues are first to report that they are most likely the main culprits in pushing the boundary of the tropics further poleward in that hemisphere.

In other words, the earth's tropical belt is getting wider, according to the study, published in the May 17, 2012 issue of *Nature*.

"If the tropics are moving poleward, then the subtropics will become even drier, and the southern portions of the United States may get drier if the storm systems move further north than they were 30 years ago," noted Allen. "If a poleward displacement of the mid-latitude storm tracks also occurs, this will shift mid-latitude precipitation poleward, impacting regional agriculture, economy, and society."

Allen is using SDSC's *Triton* compute cluster to further his research. He is investigating the role of multi-decades of ocean variability on tropical-edge displacements, as well as additional aerosol simulations using satellite-based estimates of aerosol radiative forcing. The latter shows larger solar absorption than model estimates, and therefore may result in a larger poleward displacement of the tropical belt, confirming Allen's earlier studies.

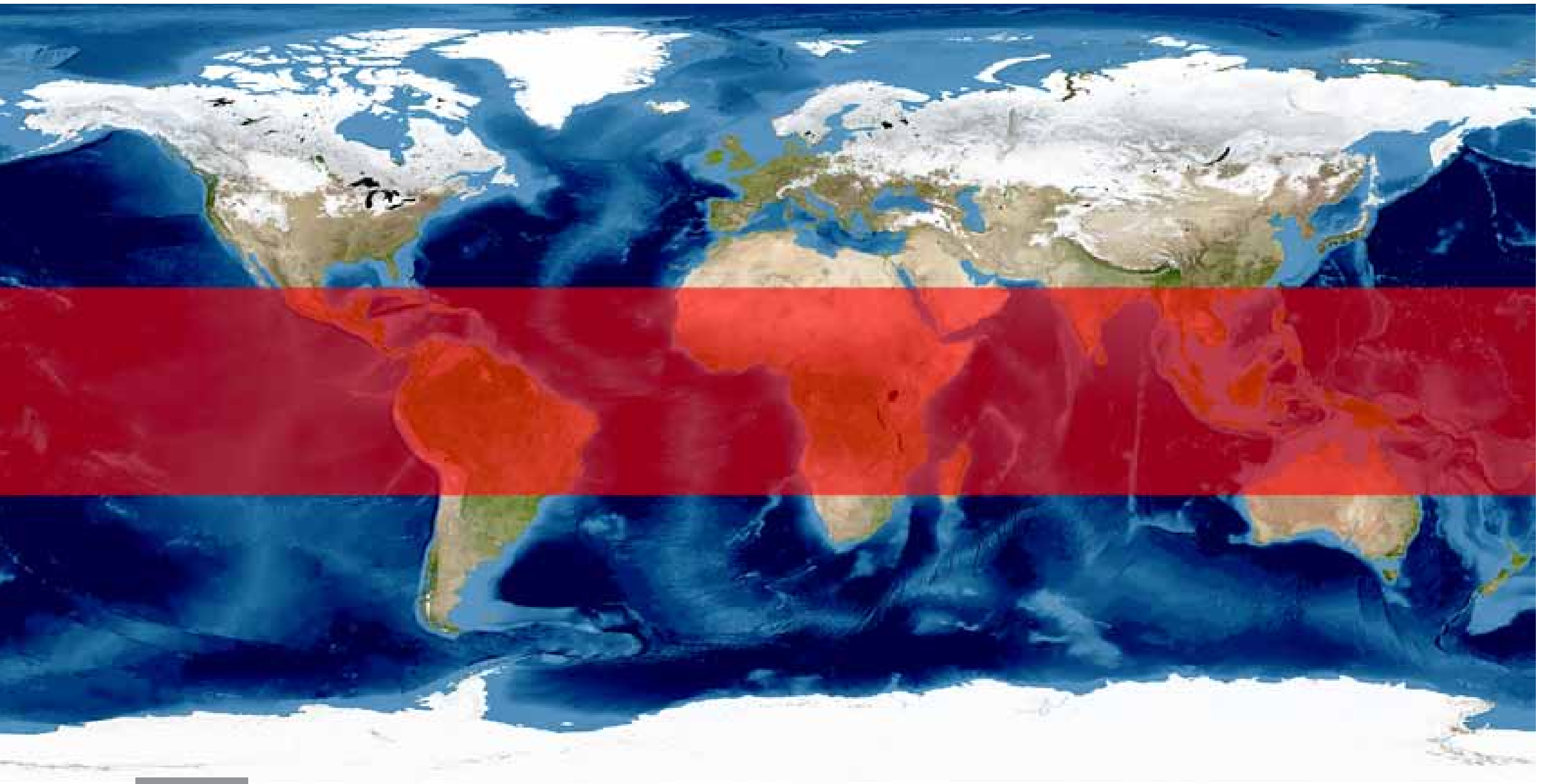
Our Shrinking Sea Ice

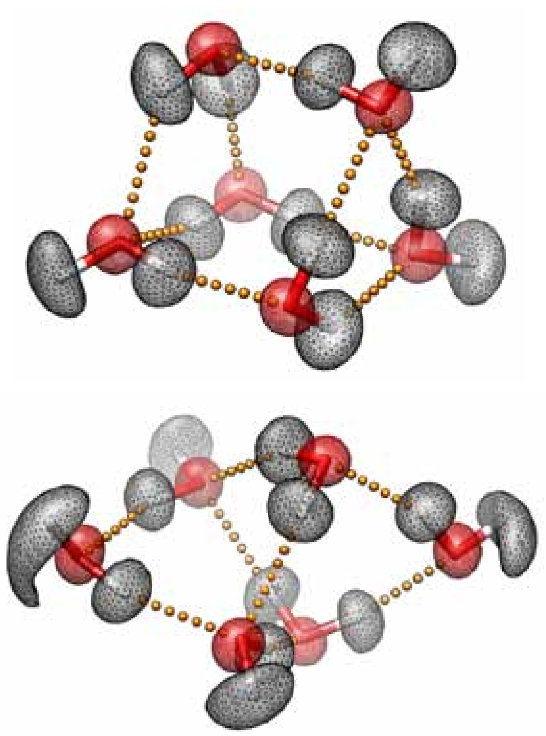
SDSC's *Triton* compute cluster is being used to develop new models to better understand the response of sea ice to significant changes in climate during recent decades, and how it interacts with other components of the climate system.

Ian Eisenman, an assistant professor at UC San Diego's Scripps Institution of Oceanography, is using *Triton* to explore simulations using a hierarchy of models with increasing levels of complexity and culminating in tests of the theories with satellite-derived observations. A principal tool will be an idealized atmosphere–sea ice global climate model, which helps bridge the gap between our understanding of fundamental processes and simulations using comprehensive climate models.

Using this and other models, Eisenman and colleagues are testing theories developed for Arctic sea ice loss in the broader context of sea ice changes in a continuum of climates, ranging from a completely ice-covered planet to a completely ice-free one. These theories will be used to interpret observed sea ice changes, and to evaluate the aspects of current projections that differ among the models, such as the rate of Arctic sea ice loss during the rest of this century.

Manmade pollutants may be driving the expansion of the Earth's tropical belt seen here in red. Credit: Robert Allen, UC Riverside. Image: NASA





Three-dimensional representations of the prism (top) and cage (bottom) structures of the water hexamer, the smallest drop of water. The mesh contours represent the actual quantum-mechanical densities of the oxygen (red) and hydrogen (white) atoms. The small orange spheres represent the hydrogen bonds between the six water molecules. Characterizing the hydrogen-bond topology of the water hexamer at the molecular level is key to understanding the unique and often surprising properties of liquid water, our life matrix. Credit: Volodymyr Babin and Francesco Paesani, UCSD

Researchers Reveal Behaviors of the Tiniest Water Droplets

A study by researchers at UC San Diego and Emory University has uncovered fundamental details about the hexamer structures that make up the tiniest droplets of water—the key component of life and one that scientists still don't fully understand.

The research, published in *The Journal of the American Chemical Society*, provides a new interpretation for experimental measurements, as well as a vital test for future studies of our most precious resource. Moreover, understanding the properties of water at the molecular level can ultimately have an impact on many areas of science, including the development of new drugs and advances in climate change research.

"Understanding the molecular properties of water's hydrogen bond network is vital to understanding everything else that happens in water," said Francesco Paesani, an assistant professor in UC San Diego's Department of Chemistry and Biochemistry.

Researchers used SDSC's data-intensive *Gordon* and *Triton* supercomputers to conduct the data-intensive simulations. "Our simulations took full advantage of *Gordon*, distributing the computations over thousands of processors," added researcher Volodymyr Babin, also with the university's Department of Chemistry and Biochemistry. "That kind of parallel efficiency would be hardly achievable on a commodity cluster."

Harnessing Light and Energy from the Sun

The sun provides 50,000 times more energy to Earth than humans consume, yet less than one percent of the energy we use is provided by solar power. The rising cost and consequences of oil and coal have generated a renewed interest in solar energy. A key to increasing the use of solar power is to ease the cost and difficulty of manufacturing solar panels through the use of new materials. Another is to harness the sun to drive an "artificial leaf" that efficiently generates clean fuel hydrogen in the same way that nature uses the sun to drive the chemical reactions to create sugar.

Solar Gold

Existing solar panels are manufactured with materials that are either expensive or not particularly efficient. A goal of scientists has been to find cheaper, nontoxic, yet efficient materials to use in solar panel manufacturing. One potential material that exists in abundance is iron pyrite, also known as fool's gold. Pyrite matches with the solar spectrum and absorbs energy well, making it an ideal candidate for photovoltaics. However, while those properties make it an attractive possibility, real world trials have shown pyrite breaks down during manufacturing and contains too many impurities on its surface to efficiently transmit energy.

Ruqian Wu, professor of physics and astronomy at UC Irvine, used the NSF's XSEDE (Extreme Science and Engineering Discovery Environment) and SDSC resources, including the *Dash* system, to model what conditions are required to grow high-quality pyrite films. Pyrite, consisting of iron and sulfur, has an irregularly charged surface that detracts from its photovoltaic qualities. Wu tested whether changing the concentration of sulfur might reduce the number of defects and improve the performance of pyrite-based devices, in particular the low open-circuit voltage.

Using computational simulations and lab experiments he found that pyrite with a higher concentration of positively-charged sulfur atoms on its surface to be the most stable and to have a reduced energy barrier, both positive attributes for photovoltaic use. Wu's work has implications not only for the use of pyrite but other materials as well, including copper indium gallium selenide and copper zinc tin sulfide.

Energy Plant

The premier example of a natural, cheap, renewable, durable photocell is already available in abundance in plants. They can transform solar energy, carbon dioxide, and water into chemical energy in the form of sugar and do it with high selectivity, albeit very low quantities, at room temperature with the additional benefit of producing oxygen as a byproduct. Taking a cue from plants may prove to be the path to solving energy and climate problems, according to Phillip Christopher, assistant professor of chemical and environmental engineering and materials science and engineering at UC Riverside.

Artificial photosynthesis requires two central parts: a collector to capture solar energy and catalysts that effectively channel the solar energy into the energy of chemical bonds. Using SDSC's *Trestles* system, Christopher has been investigating the mechanisms that control artificial photosynthesis on titanium dioxide nanostructures loaded with a metal catalyst such as copper, ruthenium, or cobalt. Simulating what is happening at a molecular level requires quantum chemical calculations to demonstrate how the catalyst manages to overcome activation barriers, and the paths energy take through the steps of artificial photosynthesis. The results, said Christopher, will help scientists to design more efficient photocatalytic materials and processes in order to produce fuels from entirely sustainable resources.

Pyrite is a commonly found material that might be used in solar panels if researchers can overcome its shortcomings. Ruqian Wu, at UC Irvine, models what conditions are required to grow high-quality pyrite films. Credit: Andrew Silver, United States Geological Survey

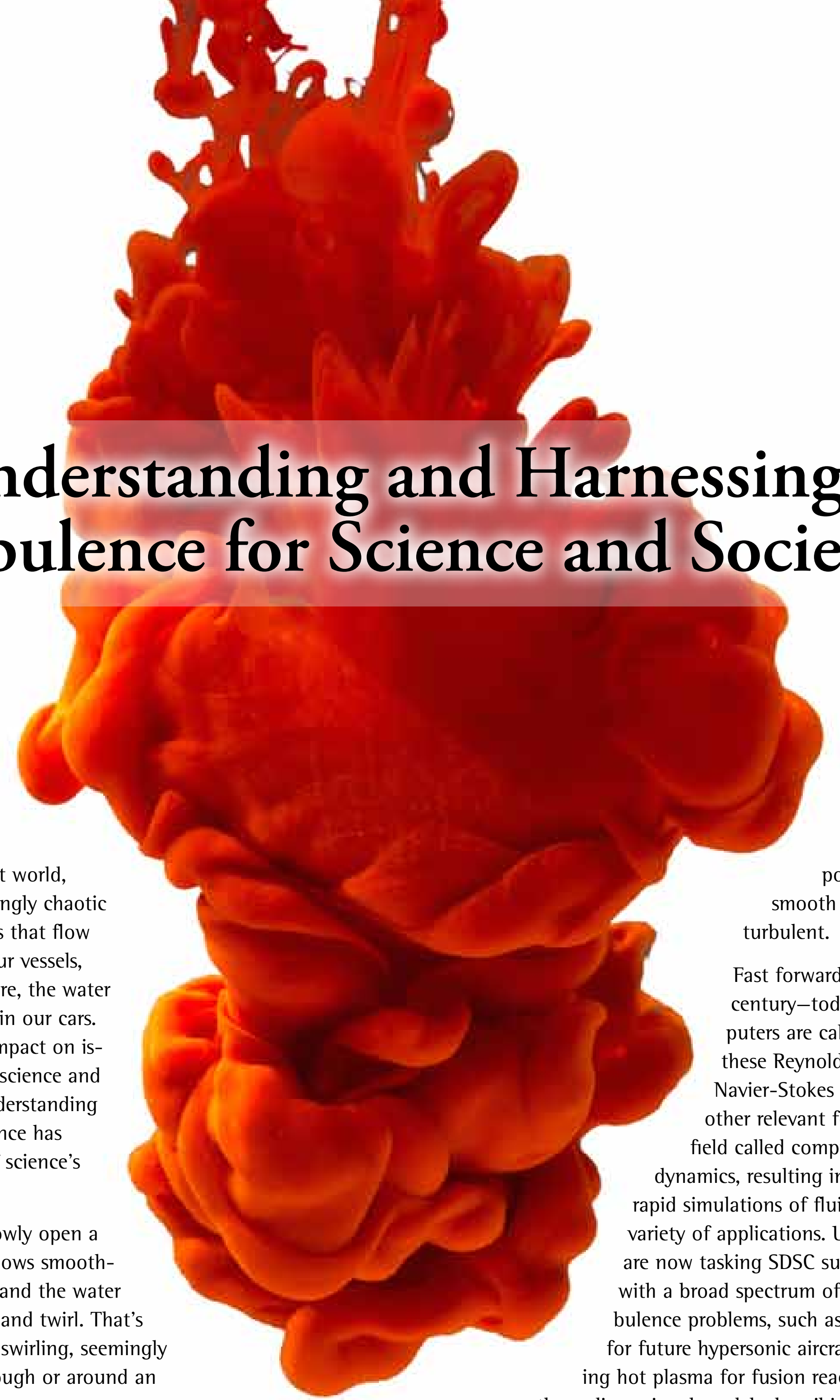
This protein trimer from the green sulfur bacteria is an example of a natural basic light harvesting complex, one that might provide clues how to harvest light for human use. Phillip Christopher, assistant professor of chemical and environmental engineering and materials science and engineering at UC Riverside, has been investigating the mechanisms that control artificial photosynthesis on titanium dioxide nanostructures loaded with a metal catalyst such as copper, ruthenium, or cobalt. Credit: RCSB Protein Data Bank (3ENI) visualized using Jmol

Forecasting Solar Power Production

Even with superior technology, the efficiency of tomorrow's solar panels will still depend on how they are installed and used. Solar panel installers take into account trees, chimneys, and poles for each individual installation, but the California Public Utilities Commission needs a regional model that would account for changes in elevation to predict solar output.

The Commission contacted Jan Kleissl, a professor of environmental engineering at the Jacobs School of Engineering at UC San Diego, and Juan Bosch, a postdoc, to create a solar output model including information about elevation. Taking detailed elevation data from NASA's Shuttle Radar Topography Mission, the researchers focused on regions with a concentration of solar power — the Bay Area and Southern California. They used SDSC's *Triton* to generate a more realistic model of the solar power output utilities and homeowners can typically expect to produce. This is an especially important tool for utilities, because it gives them a better idea of how much revenue they can actually generate, Kleissl said.

The longer it takes for the sun to rise above the local horizon in the morning and the earlier it sets in the evening, the more solar fuel is lost. Solar days are longest on top of tall mountains. They are shortest in steep valleys oriented north-south, where it can take more than an hour longer for the sun to appear in the east.



Understanding and Harnessing Turbulence for Science and Society

Ours is a turbulent world, filled with seemingly chaotic eddies and swirls that flow through the blood in our vessels, the air in our atmosphere, the water in our seas, or the fuel in our cars. Because of its critical impact on issues of importance for science and society, for decades understanding and harnessing turbulence has been considered one of science's "grand challenges."

What is turbulence? Slowly open a faucet, and the water flows smoothly. Open the tap wider, and the water suddenly begins to roil and twirl. That's turbulence—it's unruly, swirling, seemingly random movement through or around an object.

Understanding the characteristics of non-turbulent, smooth or laminar flow, is relatively simple. However, calculating turbulent events—the gusts of wind that turn turbines, water crashing over a dam, oil churning through pipelines—is frustratingly complex.

The first insights came more than 150 years ago, when French engineer Claude Navier provided the first equations governing the velocity and pressure of air, water, and other fluids around an object. Later, Irish mathematician George Stokes generalized these equations, giving rise to the field of fluid dynamics. An Irish-born physicist, Osborne Reynolds, subsequently defined the point at which smooth flow becomes turbulent.

Fast forward to the 21st century—today, supercomputers are called on to solve these Reynolds-averaged Navier-Stokes equations and other relevant formulas in a field called computational fluid dynamics, resulting in models and rapid simulations of fluid flows for a variety of applications. UC researchers are now tasking SDSC supercomputers with a broad spectrum of complex turbulence problems, such as engine design for future hypersonic aircraft, containing hot plasma for fusion reactions, and three-dimensional models describing the internal structure of the earth's protective magnetic shield.

Designing and Testing Hypersonic Aircraft

Consider for a moment the design of aircraft and jet engines. Prior to supercomputers, wind tunnels essentially provided the only means to test the aerodynamics of aircraft and jet engines, followed by the perils of manned flight testing. During the past few decades, computational fluid dynamics or aerodynamics joined this design and testing duo, owing to its ability to simulate and predict performance accurately, quickly, and cost-effectively.

From the start, computational aerodynamics was used to shed light on the basic structure of flows over an aircraft's fuselage, wings, flaps, and engines to sort through the most promising configurations for speed and energy efficiency, and to correct and improve resolution of wind tunnel data. As planes moved faster and faster, the problems became increasingly complex.

John Kim, who holds UCLA's Rockwell International Chair in Engineering, is one of the world's leaders in the field of computational fluid dynamics and aerodynamics, developing numerical simulations and large eddy simulations for the aerospace industry. Recently, Kim has been studying ways to better understand and predict the location of laminar to turbulent transition in boundary layers that cause extreme surface temperatures in hypersonic Mach 6 (approximately 4,000 mph) aircraft. When flow transitions from laminar to turbulent at these extreme speeds, heat rises near the boundary layers, also generating large shear stresses which, in turn, increase friction drag on the surface. To better understand what happens during this transition in hypersonic aircraft, Kim received an allocation on SDSC's *Trestles* supercomputer to simulate this phenomenon.

"Thermal protection and propulsion systems must assume turbulent flow over the entire body," said Kim. "In spite of its importance, the mechanisms leading to the transition of hypersonic boundary layers are still poorly understood. I'm hoping that our *Trestles* simulations will give us some new insights into this problem and how to accommodate this solution into new designs."

Since wind tunnels can be contaminated by the influence of the tunnel's walls and the structure that holds the model in place, computational fluid dynamics (CFD) is the only viable tool available to test turbulent features affecting the design of hypersonic aircraft (those that fly at up to 20 times the speed of sound), such as NASA's X-43A Hypersonic Experimental Vehicle pictured here. Credit: NASA

Realizing the Promise of Fusion Energy

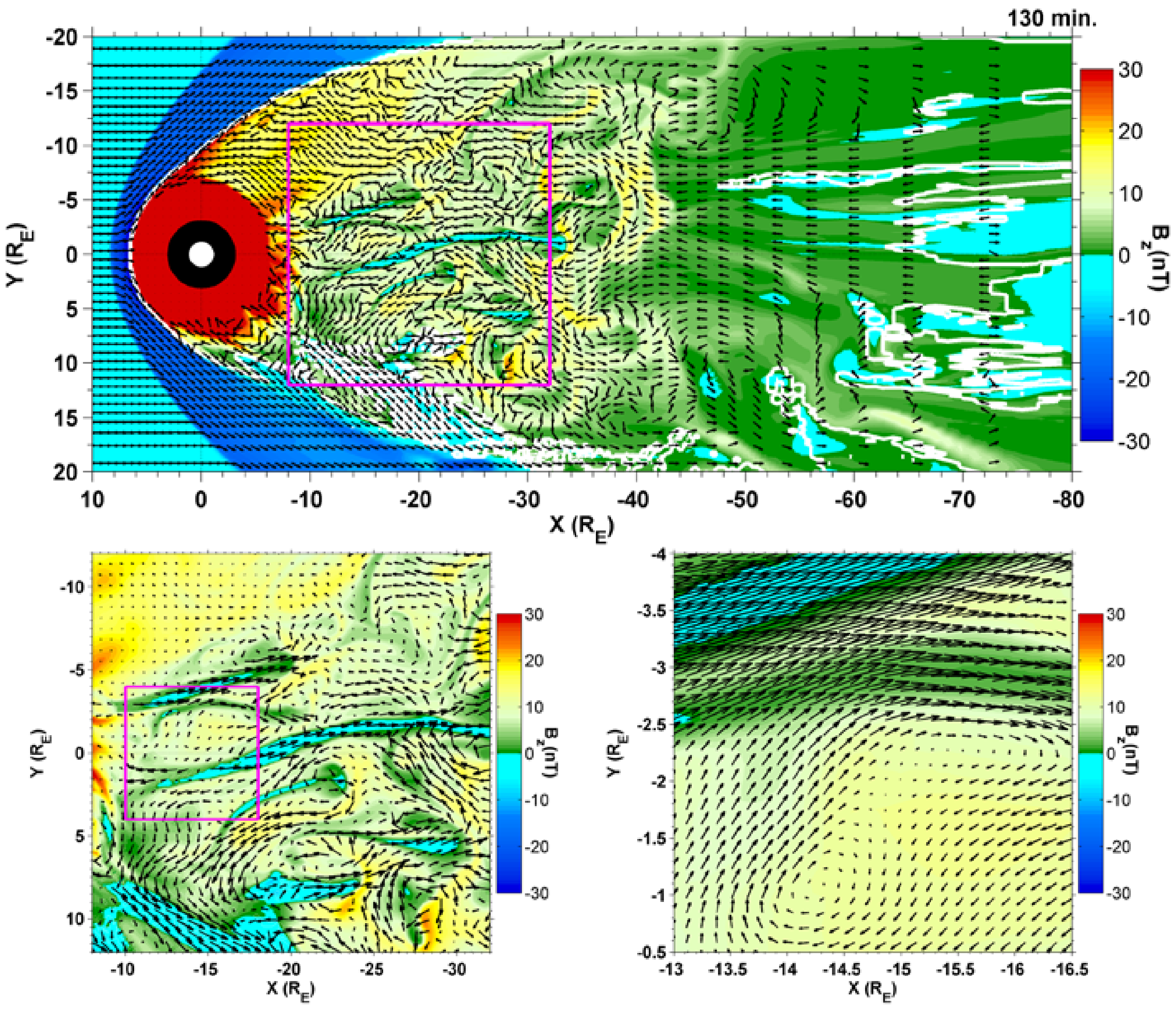
For more than a half century, scientists have talked about harnessing the power of the sun to provide a virtually unlimited supply of inexpensive, clean energy for the planet. Despite the promise and predictions, the first commercial power plant based on fusion energy remains elusive.

At issue is how to contain the power of a manmade star. Fusion occurs when hydrogen atoms fuse together under extreme heat and pressure to create denser helium atoms, releasing huge quantities of energy as a byproduct. Our sun turns 600 million tons of hydrogen into helium every second. But without the gravitational forces on the sun, earth scientists are building experimental containment vessels that house artificial magnetic fields designed to mimic star power. Here, a fusion reaction occurs when two types of hydrogen isotopes (deuterium and tritium) combine to form super-hot ionized gas or plasma that not only produces helium, but also neutrons—the source of a huge amount of energy.

"One of the primary challenges for developing economically viable nuclear fusion-based power plants is confining the ionized gas at sufficient density and temperatures in the center of the containment device to generate fusion reactions, and low enough values next to the vessel walls to not instantly melt them," said Christopher Holland, a researcher with UC San Diego's Center for Energy Research.

Part of the problem stems from a family of turbulent instabilities—generated by inherent pressure that moves from the hot center to the cool edge of the plasma—that reduces the number of fusion reactions and thus the energy generated from those reactions.

"This turbulent transport determines the level of confinement achieved, and thus directly impacts the necessary size and cost of a fusion reactor," said Holland. "It is therefore essential that we develop validated models of this transport in order to confidently design and optimize fusion-based power plants."



With the aid of magnetohydrodynamic (MHD) calculations, Mostafa El-Alaoui, a researcher with UCLA's Space Plasma Science at the Physics and Astronomy Department, used SDSC's *Gordon* supercomputer to model three-dimensional turbulent flows in the earth's magnetosphere to learn what drives these swirls and eddies, as well as their consequences. The results, seen in this simulation, show numerous distinct eddies inside the magnetotail, some nested within larger and smaller eddies. Credit: Mostafa El-Alaoui, UCLA

Toward that end, Holland and colleagues are using SDSC's *Triton* to simulate these specific instabilities and resulting turbulence that occur near the "wall" of a fusion confinement experiment. In this case, the "wall" is basically the first surface, or layer, between the plasma and the containment.

"Understanding the turbulence in this region is particularly important since it determines both the fluxes of particles and energy to the walls themselves (which determine the lifetime of those plasma-facing components), as well as setting the effective 'boundary condition' for the hot core region where the fusion reactions occur," Holland said.

Simulating the Earth's Magnetic Protective Shield

Our planet is surrounded by a bubble of magnetism, generated by flows in the molten iron in the Earth's core, which provides the first line of defense against the solar wind—streams of charged particles or plasma launched from the sun that travel across the solar system. When the solar wind presses against the outer layers of this magnetic bubble, it stretches out, creating a giant magnetotail as long as the moon's orbit, visually akin to a prairie wind sock hit by a gust of wind.

The alignment of the solar wind's interplanetary magnetic field (IMF) with the earth's magnetic field determines how many of the charged particles carried in the solar wind find their way into the planet's atmosphere and residents below.

Solar winds load the magnetosphere with energy, causing magnetic field lines to spontaneously break and reconnect with nearby field lines. When this occurs, the magnetosphere cleaves, and charged particles penetrate the atmosphere. These particles generate a spectacular light show in the northern hemisphere, known as aurorae. During intense "space weather" called magnetic storms, charged particles may wreak havoc on electrical power transmission, damage satellites, and interfere with communications. For astronauts, there is the additional hazard of particles penetrating their cells and damaging molecules within them.

More recently, satellite observations surprisingly discovered huge 40,000 kilometer-wide swirls of plasma along the boundary of the magnetosphere—the magnetopause—that could allow solar wind particles to enter, even when the magnetic fields of the earth and solar winds are aligned. In essence, the boundary region is stirred like ingredients in a blender and particles penetrate what is usually a barrier.

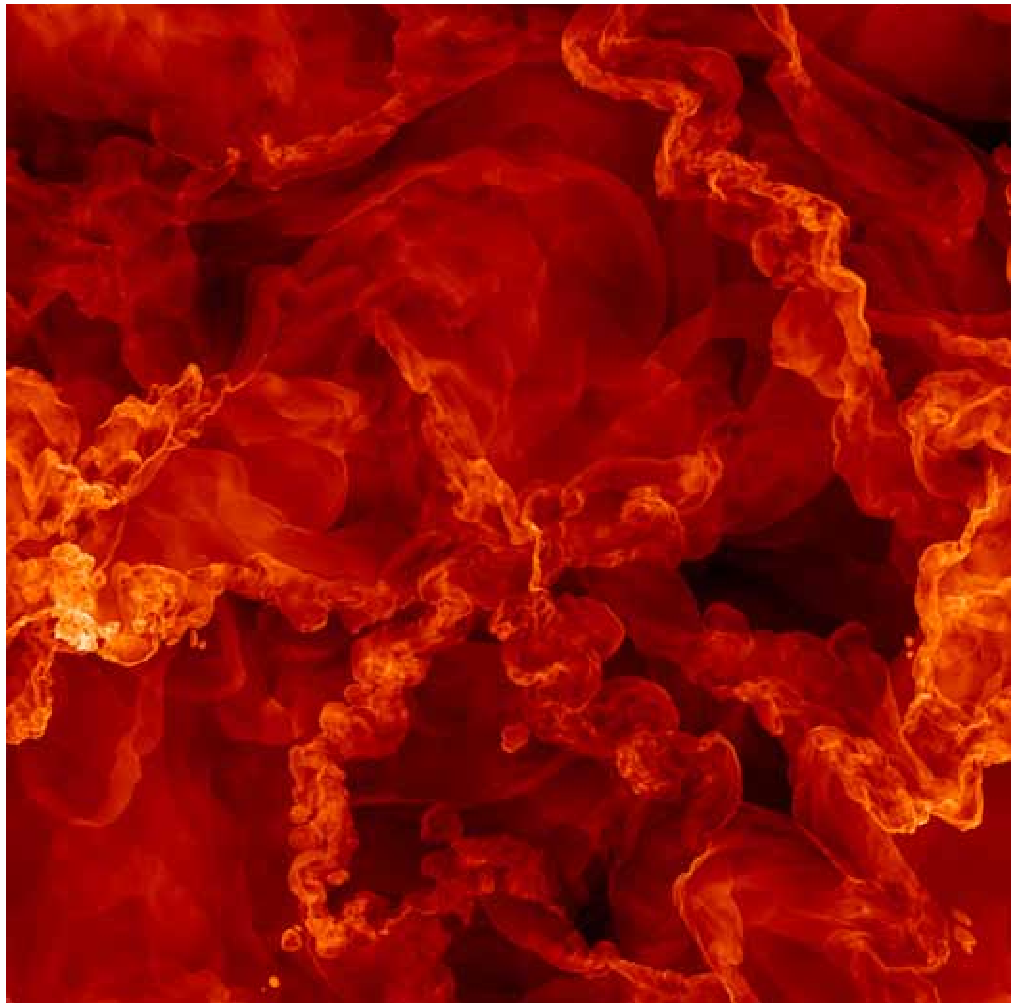
With so much at stake, scientists are calling on supercomputers to get a clearer picture of the internal workings of the magnetosphere, with the goal of identifying other areas of magnetohydrodynamic (MHD) turbulence in its plasma. MHD, also known as magnetic fluid dynamics, studies the dynamics of electrically conducting fluids, such as those found in plasma.

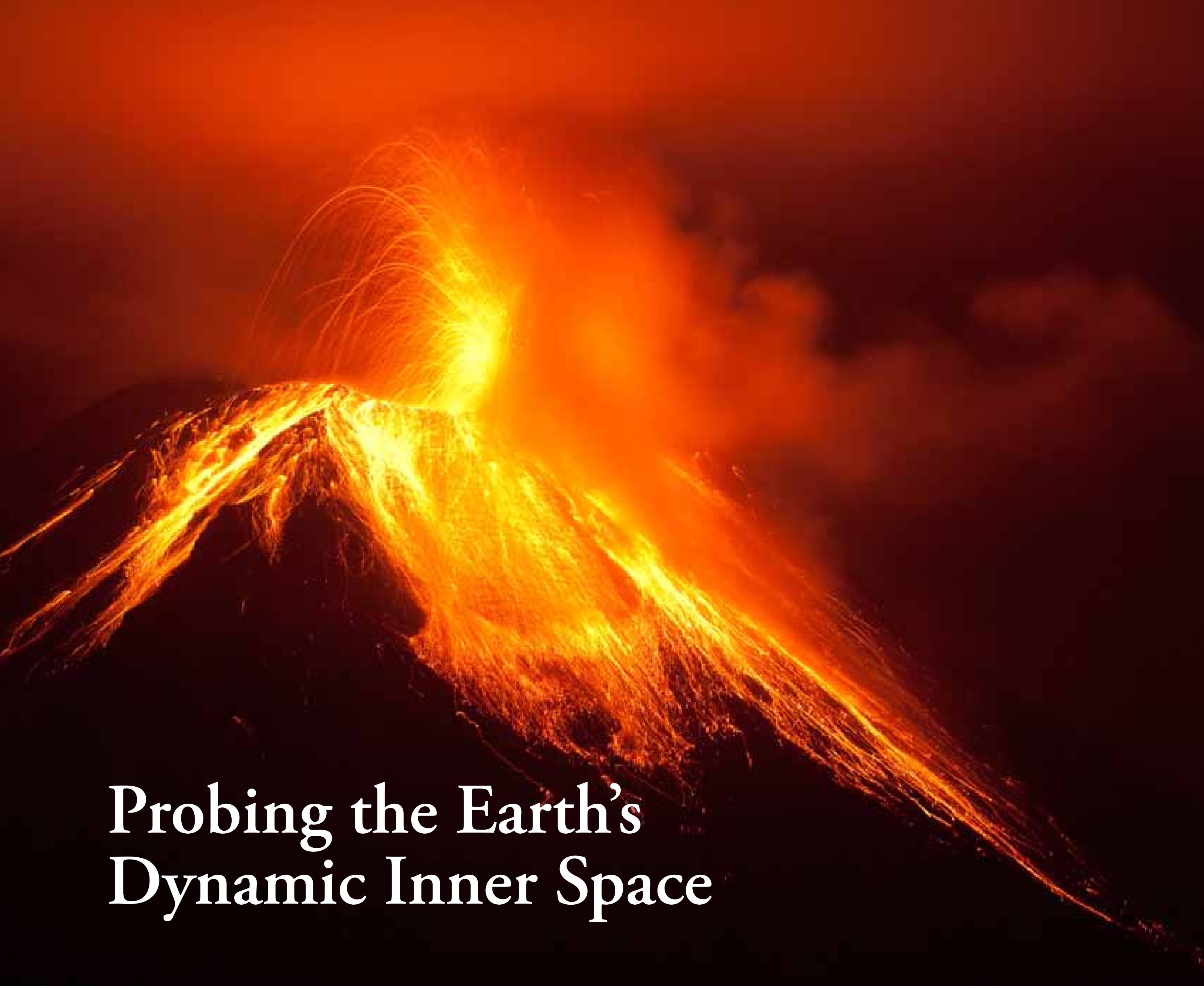
With the aid of MHD calculations, Mostafa El-Alaoui, a researcher with UCLA's Space Plasma Science at the Physics and Astronomy Department, is using SDSC's *Gordon* supercomputer to model three-dimensional turbulent flows in the earth's magnetosphere, and learn what drives these swirls and eddies as well as their consequences. The models are then compared to satellite observations.

The results thus far are showing numerous distinct eddies inside the magnetotail, some nested within larger and smaller eddies.

"Using *Gordon* at SDSC, we performed MHD simulations that realistically describe the magnetosphere and its turbulence," said El-Alaoui. "We verified that the spectrum of eddies had the correct properties by simulating observed events and comparing the results directly with spacecraft observations."

UC San Diego researchers have shown how fundamental laws of turbulent geophysical flows can also be extended to supersonic turbulence in the interstellar medium of galaxies. This image, stored and analyzed at SDSC, shows the density field from one snapshot of the simulation. The brightest regions in the image represent gas at the highest density, compressed by the action of a complex system of shocks in the turbulent flow. Dense filaments and cores, created in such a way by supersonic turbulent flows, are subject to massive gravitational collapse, and that leads to the birth of stars. Credit: Alexei Kritsuk, Michael Norman, Paolo Padoan, and Rick Wagner, UCSD





Probing the Earth's Dynamic Inner Space

Scientific observations of earthquakes go back only 100 years or so, meaning that researchers have a relatively short time span of records from which to glean clues about how such events actually occur—and how likely they are to occur again.

Advances in supercomputer capabilities—key ones being the ability to process extremely large data sets and the advent of new scientific visualization techniques—have allowed scientists to create some of the most detailed simulations of earthquakes and other seismic disruptions. These high-resolution models are being used to explore the earth's myriad subterranean forces at work, and to help better predict when, where, and how they might turn into major disruptions that could mean injuries, loss of life, and economic hardship.

Numerous projects underway throughout the UC system are relying on SDSC's supercomputer resources and expertise to further seismological science. Highlights include:

Investigating Major Earthquake Faults

Researchers are using the data-intensive muscle of SDSC's *Gordon* supercomputer to develop detailed, large-scale computer simulations of earthquake faults as part of a new $4.6 million grant from the National Science Foundation (NSF). The five-year project, led by UC Riverside, includes researchers from San Diego State University, the Southern California Earthquake Center (SCEC) at the University of Southern California, Brown University, and Columbia University, in addition to the U.S. Geological Survey (USGS).

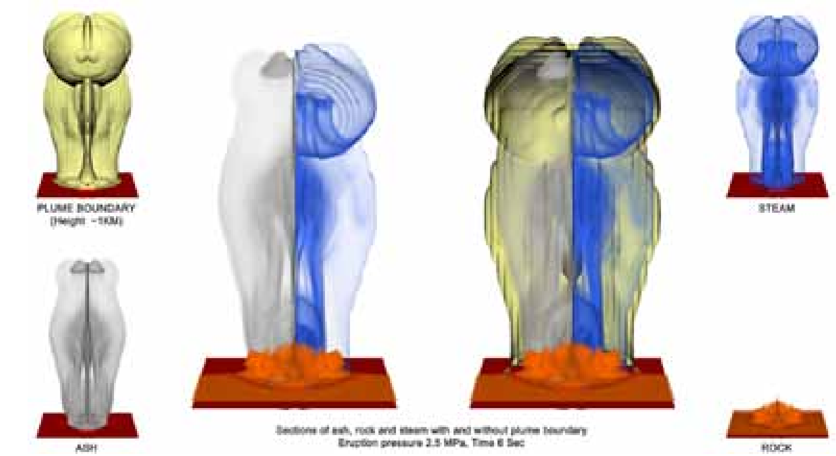
Scientists plan to develop and apply the most capable earthquake simulators to investigate these fault systems, focusing first on the North American plate boundary and the San Andreas system across Northern and Southern California.

"Our simulations of plate boundary fault systems will span more than 10,000 years of plate motion and consist of up to a million discrete earthquake events, giving us an abundant array of data to analyze," said James Dieterich, a distinguished professor of geophysics in UCR's Department of Earth Sciences, and principal investigator of the project.

These simulations will provide the means to integrate a wide range of observations from seismology and earthquake geology into a common framework. Added Dieterich: "They are computationally fast and efficient, and one of the project's goals is to improve our short- and long-term earthquake forecasting capabilities, which can save lives and prevent injuries."

The primary computational development for this effort is to enable an existing earthquake simulator developed at UCR to run efficiently on supercomputers with tens of thousands of cores, said Yifeng Cui, director of the High Performance GeoComputing Laboratory at SDSC. Cui recently participated in a project to create the most detailed simulation ever of a Magnitude 8.0 earthquake in California, whose related code will be used in the UCR project for detailed single event rupture calculations. (see *Collaborates to Save Lives and Property for Californians*, pp.34–35)

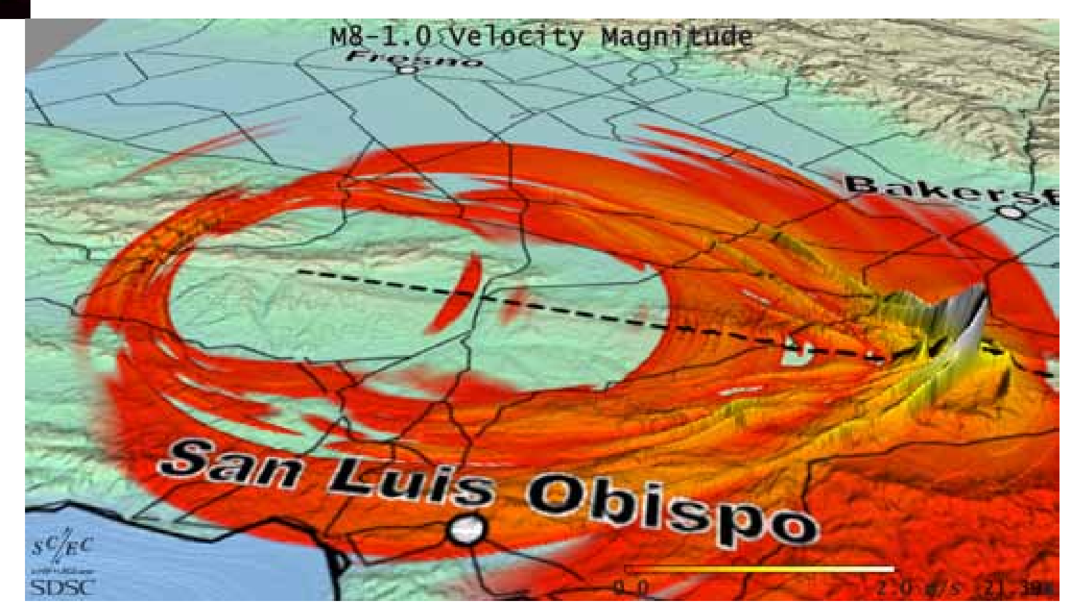
(Right) These visualizations show how eruptive conduits feeding volcanic jets and plumes are connected to the atmosphere through volcanic vents that can alter the dynamics and structure of these eruptions. The simulations were done using CFDLib software from the Theoretical Division of the Los Alamos National Laboratory. Credit: Amit Chourasia, SDSC



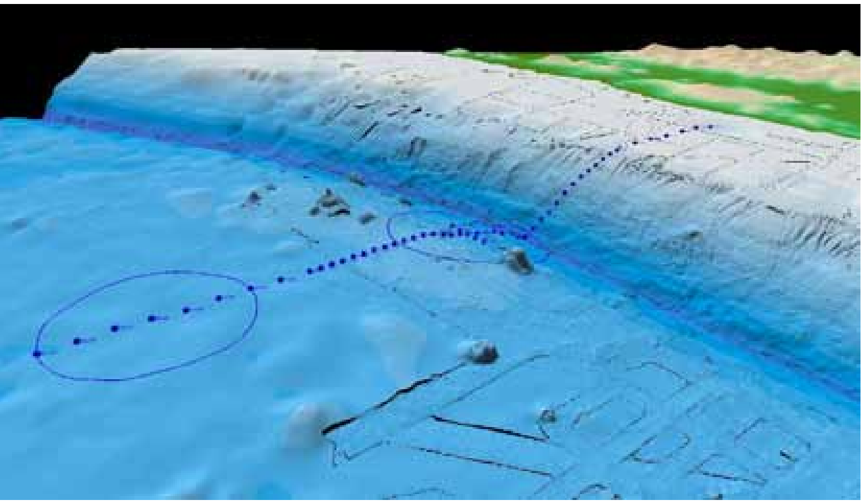
Modeling Volcanic Eruptions

The dangerous nature of explosive volcanic eruptions makes their behavior important to study, as researchers try to better understand them to protect the surrounding populace and infrastructure. However, it is this very behavior that also makes them so difficult to study.

Supercomputing resources are being increasingly called upon to run computational fluid dynamics (CFD) models capable of addressing the complex flow fields or volcanic activity, with large ranges in scales. Using a library of CFD codes developed in the Theoretical Fluid Dynamics and Solid Mechanics group at Los Alamos National Laboratory (LANL), Darcy Ogden, an assistant professor with UC San Diego's Scripps Institution of Oceanography (SIO) is using SDSC's *Trestles* supercomputer to conduct volcanic modeling, while relying on SDSC's Advanced User Support group for visualization, parallel scaling improvements, and adaptation of program codes for SDSC's *Gordon* system.



(Left) Snapshot of the ground motion after a simulated Magnitude 8.0 earthquake along the San Andreas Fault through some densely populated areas. Researchers at SDSC and San Diego State University created the largest-ever earthquake simulation in 2010. Credit: Amit Chourasia, SDSC



A 3D image of the Nicaraguan Trench. The light blue colors indicate a sea floor depth of 3500m; the dark blue trench goes to 6000m deep; and the white colors on the continental shelf are about 100m deep. Credit: Kerry Key, SIO/UCSD

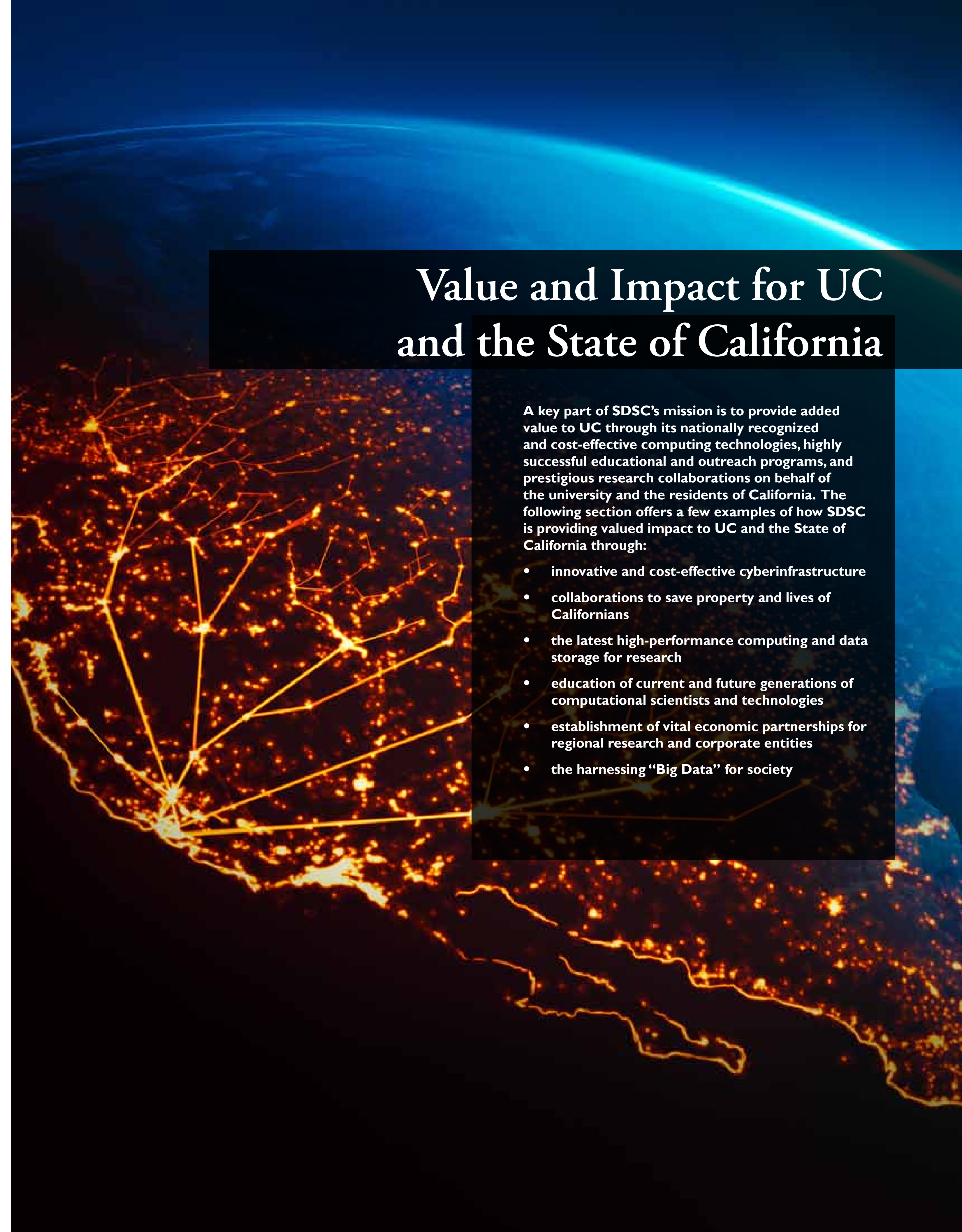
Mapping Seafloor Plate Tectonics

Another SIO project is being led by Kerry Key, an associate research geophysicist at the institution's Marine Electromagnetics (EM) Laboratory. Key is using SDSC's *Triton* compute cluster to create images of seafloor electrical conductivity from two geophysical data sets collected to map an offshore gas reservoir and to image plate tectonics at a subduction zone.

The first data was collected at the Scarborough gas field off Western Australia under industry sponsorship. *Triton* is being used to help map the distribution and extent of electrically resistive hydrocarbons trapped in conductive sediments of this hydrocarbon reservoir. Data was collected during a month-long survey that included towing an electric dipole transmitter close to the seabed across hundreds of kilometers to broadcast EM energy. An array of seafloor EM receivers recorded the strengths of these transmissions, which depend on the electrical conductivity of the surrounding geology and pore fluids.

The second data set is part of a NSF-funded project called SERPENT, for Serpentinite Extension and Regional Porosity Experiment across the Nicaraguan Trench. This was where a major tsunami earthquake struck in 1992—the first such event to be recorded by modern seismic broadband networks. *Triton* has been running large-scale, non-linear inversion algorithms that convert the observed EM recordings into images of electrical conductivity so researchers can develop more detailed models of the region to better understand subduction earthquakes and volcanism.

(Above) The R/V Roger Revelle research vessel. (Below) The Marine EM Lab specializes in pioneering new technologies for electromagnetics (EM) exploration in the oceans. The two main instrument systems developed at the Marine EM Lab are the SIO seafloor EM receiver and the Scripps Undersea EM Source Instrument (SUESI). SUESI is a deep-towed horizontal electric dipole transmitter used for the controlled-source EM (CSEM) method. The receivers are used for measuring the seafloor electromagnetic fields for both the CSEM and magnetotelluric (MT) method. Credit: Kerry Key, SIO/UC San Diego



Value and Impact for UC and the State of California

A key part of SDSC's mission is to provide added value to UC through its nationally recognized and cost-effective computing technologies, highly successful educational and outreach programs, and prestigious research collaborations on behalf of the university and the residents of California. The following section offers a few examples of how SDSC is providing valued impact to UC and the State of California through:

- **innovative and cost-effective cyberinfrastructure**
- **collaborations to save property and lives of Californians**
- **the latest high-performance computing and data storage for research**
- **education of current and future generations of computational scientists and technologies**
- **establishment of vital economic partnerships for regional research and corporate entities**
- **the harnessing "Big Data" for society**

Provides Innovative and Cost-Effective Cyberinfrastructure for UC

If data is the simplest unit that defines the Information Age, then cyberinfrastructure is the force that drives and converts data to solutions that meet society's needs. Like other infrastructure, such as the electric power grid or the nation's highways, cyberinfrastructure combines complex elements to create a dynamic system. Key cyberinfrastructure components include performance-oriented computer hardware, sophisticated software, data storage, and human expertise, working in collaboration across a wide communications grid.

SDSC is internationally recognized as a pioneer and leader in cyberinfrastructure, providing a unique competitive edge to UC San Diego and UC system-wide through cost-effective services. The Center's cyberinfrastructure—encompassing numerous UC research collaborations—includes projects ranging from research to understand the impact of climate change on California's coastline, to the storage and analysis of genomic data that one day may be used to personalize treatment for cancer.

Performance-Oriented Computer Hardware

In October 2009, SDSC launched the *Triton*, a data-intensive, high-performance computer featuring massive data analysis and preservation system. The resource—featuring large-memory nodes that provide some of the most extensive data-analysis power available at any research university—has to-date provided more than a million core-hours to more than 600 users across eight UC campuses. During the past year, *Triton* has been used for studies across a variety of disciplines such as the biological sciences, geosciences, engineering, data and knowledge systems, and computer science and engineering. Aside from research, *Triton* is used as a teaching tool for computer courses at UC San Diego and UC Santa Barbara.

SDSC also has made half of its 19,000-square-foot data center available as a recharge-based regional colocation facility for the UC system, now including more than 90 groups spanning eight UC campuses. This recharge structure not only leverages an enormous prior capital investment in the facility and power/cooling upgrades paid by SDSC and the National Science Foundation (NSF), it also leverages the critical mass of expertise/management and the core technology infrastructure at SDSC. According to estimates, co-locating computers and other equipment at SDSC now represents an annual UC-wide utility savings of about $822,000.

Sophisticated Software

During its more than quarter-century history, SDSC researchers have pioneered some of the most sophisticated and widely used software for high-performance computing, data management, and workflows anywhere.

The Rocks Cluster Toolkit, developed in 2000 by SDSC researchers as part of the NSF National Partnership for Advanced Computational Infrastructure (NPACI), today provides a blueprint for the construction and software configuration of computer clusters for about 1,400 academic, government, and commercial organizations worldwide. Rocks is a software package that gives the user of a cluster computer the ability to quickly and easily build the system software suite, install new software on the processors, and manage the system configuration.

The Storage Research Broker (SRB), released in 1997 by the Data Intensive Cyber Environments (DICE) research group at SDSC and collaborators, was designed to serve as "middleware" to hold together cache sites from NPACI. In 2004, the DICE group introduced iRODS (for I Rule Oriented Data Systems), providing a more adaptive middleware to meet the needs of the end-user community, kind of a "glass box" where users can see how the system works and can tweak the controls to meet their demands.

SDSC researchers, in collaboration with colleagues at UC Santa Barbara and UC Davis, also developed the open-source Kepler project, a user-friendly workflow tool that can organize and automate scientific tasks, helping scientists take full advantage of today's complex software and Web services. With Kepler, researchers from many disciplines are able to automate workflows without having to become expert programmers.

Data Storage

More recently, SDSC Cloud was added to the Center's cyberinfrastructure toolkit, providing what is believed to be the first large-scale academic deployment of cloud storage in the U.S. SDSC Cloud offers north/south geographical replication, system administration services and database hosting, and provides cost-effective solutions for long-term data-sharing needs now mandated by federal agencies. Today, SDSC Cloud and other resources host numerous high-impact data collections for science and society including: the Protein Data Bank (protein structures); the Library of Congress Chronopolis digital preservation system that includes the "web-at-risk" collections of the California Digital Library; Open Topography.org (LiDAR data); and the American Red Cross Safe and Well website, developed at SDSC in urgent response to the Hurricane Katrina disaster and the need to match missing people with families and friends.

In 2010, SDSC also launched—with the Office of California's CIO—the California Spatial Data Infrastructure Project, with 96 terabytes of dedicated storage including the California Coastal Atlas which, among other things, is being used to assess the impact of the sea level rise from climate change along the California coastline.

SDSC also is partnering with bioinformatics expert David Haussler at UC Santa Cruz to house the Cancer Genomics Hub (CGHub), a large-scale data repository and user portal for the National Cancer Institute's cancer genome research programs. CGHub is connected by high-performance national research networks to major centers nationwide who are participating in these projects, including UCSC. SDSC provides a secure environment to store patient records for the UC Riverside School of Medicine and UC Davis, and is fully compliant with regulations in the Health Insurance Portability and Accountability Act of 1996 (HIPAA). The Center also is hosting and managing collected data as part of the National Children's Study, the largest long-term examination of children's health and development ever conducted in the United States.

Human Expertise Working Across a Wide-Area Communications Grid

SDSC is the only supercomputer center in the western part of the nation participating in the NSF's Extreme Science and Engineering Discovery Environment (XSEDE) as a service provider for advanced cyberinfrastructure services for the U.S. open research community. Regionally, SDSC serves as a transit hub for internet activity through the San Diego Network Access Point (SD-NAP), directly connected to SDSC's data center. Established in February 1998, SD-NAP provides high-performance, configurable, and reliable connectivity to customers, in addition to "peering" with national and regional service providers—a voluntary interconnection of separate networks that permits the exchange of internet traffic among customers. SD-NAP customers include The Scripps Research Institute, Salk Institute, Sanford-Burnham Medical Research Institute, San Diego State University, San Diego County Office of Education, and the Corporation for Education Network Initiatives in California (CENIC).

ShaRCS Digested Digital Data for UC Research

Spurred by a system-wide survey showing a widespread need for high-performance research computing, about three years ago the University of California created a pilot program named Shared Research Cluster Services (ShaRCS) for its 10 campuses. The pronunciation of this word, "sharks", inspired the names of the northern cluster, *Mako*, housed at the Lawrence Berkeley National Laboratory (LBNL)—and the southern cluster, *Thresher*, located at SDSC.

The goal of the project was to further advance the university's academic mission by fostering collaborations, while enhancing its competitiveness through a centrally located, energy-efficient, and cost-competitive research computing resource. Initially, 24 pilot projects were chosen for allocations on the system, 13 using *Mako* at LBNL, and 11 on *Thresher* at SDSC. These projects ran the gamut of scientific disciplines, including oceanic simulation of surface waves and currents, UCLA and UCSB; climate modeling capacity, UC Berkeley; dynamics and allosteric regulation of enzyme complex, UC Riverside; physics-based protein structure prediction, UC San Francisco.

While the pilot program met its objectives, the current budget environment precluded a follow-on system-wide program although the pilot results are being incorporated into several campus-level systems.

Collaborates to Protect and Preserve Lives and Property for Californians

From its sandy coastline along the Pacific to its expansive deserts, snow-capped mountains and tempered climate, few areas in any part of the world can match California's diverse landscape and spectacular views. It's the reason many migrated to this state from other parts of the nation. But, as most are aware, the state's precious environment is under duress from manmade and natural forces that threaten both lives and property of its residents. SDSC is collaborating with University of California researchers on several projects to help predict and mitigate damage from potential ecological and geologic disasters.

Climate Change and California's Coastline

Climate studies show that California temperatures have warmed significantly during the 21st century, with marked increases in the frequency, magnitude, and duration of heat waves, along with rising sea level extremes. Moreover, hydrological models project increased flooding by the end of the century for both the Northern and Southern Sierra Nevada, caused by increases in the size and frequency of major storms, and more precipitation falling as rain due to continually melting snowpacks.

Projections also show that reductions in California's freshwater supply would require an increased reliance on groundwater supplies at unsustainable levels, unless significant management changes occur. John Helly, a researcher with SDSC and UC San Diego's Scripps Institution of Oceanography (SIO), is providing his computational expertise to several projects aimed at better understanding and characterizing the risks of sea-level rise to California resources.

One is the U.S. Navy's SPAWAR (Space and Naval Warfare Systems Command) project, aimed at developing a risk assessment method related to the impact of local mean sea level rise and associated phenomena on military installations in the southwestern U.S. Helly and colleagues are studying two California bases—Naval Base Coronado and Marine Corps Base Camp Pendelton—to develop an analysis framework for determining potential vulnerabilities that could affect any coastal military installation in the region. This same technology could be applied to civilian infrastructure.

Another is the CASCaDE II project, an extension of CASCaDE I, or Computational Assessments of Scenarios of Change for the Delta Ecosystem. While simulations and results from CASCaDE I focused on climate as the primary driver, CASCaDE II aims to create simulations that take into account many more related aspects—watershed flows, stream temperatures, downstream sea levels, sediment transport, salinity levels, and dynamics of myriad underwater populations.

CASCaDE II is using SDSC's *Gordon* supercomputer to perform highly detailed simulations of the San Francisco Bay-Sacramento River Delta—an area critical to ensuring a reliable and sustainable water supply for the Golden State. The goal is to provide more accurate analyses and effective responses to how this challenge will be met in this complex bay-delta system for both the near and longer term.

Being Better Prepared for "The Big One"

Seismologists have long been asking not if, but when 'The Big One' will strike southern California. Just how big will it be, and how will the amount of shaking vary throughout the region? And how will we be better prepared to cope?

Seismologists at SDSC, San Diego State University (SDSU), and the Southern California Earthquake Center (SCEC) at the University of Southern California (USC) recently created the largest-ever simulation of a Magnitude 8.0 (M8) earthquake, along primarily the southern section of the San Andreas Fault. About 25 million people reside in that area, which extends as far south as Yuma, Arizona, and Ensenada, Mexico, and runs up through southern California as far north as Fresno.

SDSC provided the high-performance computing and scientific visualization expertise for the simulation. The research was selected as a finalist for the Gordon Bell prize, awarded annually for outstanding achievement in high-performance computing applications, and the simulation won the people's choice OASCR awards at the 2011 SciDAC (Scientific Discovery through Advanced Computing Program) conference. 'OASCR' stands for the Office of Advanced Scientific Computing Research, and in the tradition of the Hollywood original, recognizes some of the best work in computer-generated visualizations.

The landmark M8 simulation is more than just a visualization—it represents a breakthrough in seismology both in terms of computational size and scalability. It also opens up new territory for earthquake science and engineering with the goal of reducing the potential for loss of life and property. While researchers note that this massive simulation is just one of many possible scenarios that could occur, the simulations confirm that high-rise buildings are more susceptible to the low-frequency, or a roller-coaster-like motion, while smaller structures typically suffer more damage from the higher-frequency shaking, which feels more like a series of sudden jolts.

Earthquake simulations such as this can be used to refine early warning systems and advanced disaster relief planning, as well as help engineers, emergency response teams, and geophysicists better understand seismic hazards, not just in California but around the world.

Assessing Environmental Impacts on California's Forests and Mountains

A wide range of global environmental issues, many of which directly affect California's citizens and surroundings, is being studied by UC Santa Barbara's Department of Ecology, Evolution, and Marine Biology (EEMB). Department Chair Cherie Briggs has been using SDSC's *Triton* cluster to investigate how Sudden Oak Death (SOD)—an introduced disease sweeping through California's oak woodlands and destroying several dominant tree species while altering the non-living environment for the animal community—is potentially affecting the risk of Lyme disease. Briggs' lab is developing computer models to quantify how the relative abundance of lizards, wood rats, and other vertebrate species influence the prevalence of this disease. The next step is to test the models by perturbing the system, to see if it responds as predicted by those models. That is where SOD comes in—the researchers are using the changes caused by SOD to forest communities as a large-scale perturbation.

Mountain yellow-legged frog. Credit: USGS

Briggs' lab also is investigating why a fungus is now killing certain frog populations in the high-elevation regions of the Sierra Nevada. Mountain yellow-legged frogs (*Rana muscosa*) were once extremely abundant in that region, but now they are close to extinction. Although a number of factors, including introduced trout and chemical contaminants, have contributed to their decline during the past century, the disease cytridiomycosis—caused by the chytrid fungus B. dendrobatidis—has resulted in the local extinction of a large number of populations in the last decade. In some regions of the Sierra, however, small infected populations of *R. muscosa* are apparently persisting with the fungus. Researchers are using a combination of field surveys, genetic and molecular techniques, and mathematical modeling on SDSC's *Triton* cluster to better understand why some populations are able to persist with this pathogen while others are nearing extinction.

Modeling California's Water Resources

Ilya Zazlavsky, director of Spatial Information Systems Laboratory at SDSC, with the Consortium for the Advancement of Hydrologic Science, Inc. (CUAHSI), is partnering with the UC Center for Hydrologic Modeling (UCCHM) to build a state-of-the-art, integrated computer model of California's water resources. The model would allow UCCHM, based at UC Irvine, to address pressing issues in response to climate change and a diminishing snowpack; how these resources will vary in response to climate oscillations such as El Niño; and how the frequency of hydrologic extremes such as flooding and drought will affect California. The goal is to inform and advise state, regional, national, and international leaders of projections and the future of water availability, while also providing valuable information to local, regional and state planning agencies with a stake in the water policy.

Builds the Latest High-Performance Computing and Data Storage for Research

For more than 25 years, the San Diego Supercomputer Center has led the way in cutting-edge, high-performance computing (HPC) systems for use by the scientific community at the local, state, and national levels. From the earliest CRAY computers to today's integrated, data-intensive systems, SDSC has focused on providing innovative architectures designed to keep pace with the changing needs of science and engineering. Today, about 1600 researchers including 150 prinicipal investigators across eight UC campuses are using SDSC compute and HPC storage resources for research, including its high-performance national systems, described here.

Gordon: Meeting the Demands of Data Intensive Computing

With exponential volumes of digital data being generated by large scale simulations and scientific instruments, the computing capability of traditional FLOPS-based (Floating Point Operations per Second) systems may no longer be sufficient for many research challenges. More than FLOPS, many researchers now require systems that can move huge amounts of data from disk to the processor at rates and in volumes that are an order of magnitude or more than the capabilities of most HPC resources available today.

Gordon, the result of a five-year, $20 million National Science Foundation (NSF) award, is SDSC's newest HPC resource – and the first system built specifically for the challenges of data-intensive computing. Debuting in early 2012 as one of the 50 fastest supercomputers in the world, *Gordon* is the first such system to use massive amounts of flash-based memory, more common in smaller devices such as cellphones and laptops, but with much greater performance and durability.

While computing speed is an important metric, so is capability. Think of *Gordon* as the world's largest thumb drive, but with the capacity to ingest about 220 movies per second from Netflix, or to consume the entire catalog of about 100,000 Netflix movies – while still having room for another 200,000 titles.

Along with its large memory "supernodes", *Gordon's* speed and large memory capabilities make it an ideal platform for tackling data-intensive problems in areas such as genomics, computational chemistry, structural mechanics, image processing, geophysics, and data mining applications.

Trestles: High-Productivity Workhorse

Since entering production in early 2011, SDSC's *Trestles* has attracted researchers from diverse disciplines who need access to a fully supported supercomputer – but with shorter turn-around times than has been typical for most systems. Featuring flash-based memory like SDSC's larger *Gordon* system, *Trestles* was specifically designed to enable researchers with modest-scale computational needs to be as scientifically productive as possible. In response to requests by researchers for more flexible access modes, *Trestles* offers users long run times (up to two weeks), as well as access to pre-emptive, on-demand queues for applications that require urgent access because of unpredictable natural or manmade events that may have a societal impact.

Trestles, the result of a $2.8 million NSF grant, is also recognized as the leading science gateway platform in the NSF's Extreme Science and Engineering Discovery Environment (XSEDE) program. *Trestles* users span a wide range of domains, including phylogenetic research, computational chemistry, material sciences, geographic information analysis, high-impact storm event prediction, biophysics, astronomy, cosmology, and gravitational physics.

High-Performance, Scalable Storage

Both *Gordon* and *Trestles* are connected to SDSC's innovative *Data Oasis* parallel file storage system, which ranks as one the fastest parallel file systems in the academic community. *Data Oasis* contains four petabytes of capacity and sustained speeds of 100 gigabytes per second (GB/s) to handle just about any data-intensive research challenge.

This means researchers can retrieve or store 64 terabytes of data – the equivalent of *Gordon's* entire random access memory (RAM) – in only about 10 minutes, significantly reducing research times needed for retrieving, analyzing, storing, or sharing extremely large datasets. Just one terabyte of data equals all the information printed on paper made from 50,000 trees.

"We view *Data Oasis* as a solution for coping with the data deluge going on in the scientific community, by providing a high-performance, scalable storage system that many of today's researchers need," said SDSC Director Michael Norman. "We are entering the era of data-intensive computing, and that means positioning SDSC as a leading resource in the management of Big Data challenges that are pervasive throughout genomics, biological and environmental research, astrophysics, Internet research, and business informatics, just to name a few."

Educates Current and Future Generations of Computational Scientists and Technologists

As an integral member of a university community, SDSC's philosophy includes the empowerment of science and engineering researchers in the present, along with the education and training of the next generation of scientists, technologists, and engineers for the future.

Historically, the Center has embraced this challenge, providing education and outreach programs designed to inform and offer hands-on training to regional teachers and students about the latest technological tools and trends for the classroom and future careers, in addition to training for computational skills needed by university students.

Among other things, SDSC's education, outreach, and training program seeks to identify knowledge, skills, and trends for current and future marketplaces, to best deliver relevant, targeted, and effective educational programs to meet those needs.

Teaching Technology to the Teachers

During the past year, SDSC and San Diego State University received a three-year grant from the National Science Foundation (NSF) to jointly expand computer sciences curriculum among San Diego high schools, community colleges, and universities.

The project, called "Computing Principles for All Students' Success', or ComPASS, is designed to teach both content and pedagogy around computer science principles to teachers in the San Diego region.

"This project strategically targets the critical elements necessary for offering stimulating and engaging college-preparatory computer science courses to all students in high school, when they are exploring directions and possibilities for their own future," said Diane Baxter, director of education at SDSC and UCSD Principal Investigator (PI) for the ComPASS project.

"We educate K-14 teachers so they understand both the underlying principles of computing (hence, how those principles can help us approach problem-solving), and how scientists and engineers use today's computational tools as a major supporting leg of exploration and discovery," she added.

Baxter, UC San Diego colleague and co-PI Beth Simon, and SDSU colleagues are teaching current (in-service) and future (pre-service) secondary school teachers through a newly piloted training program for teaching Computer Science Principles (CSP). Current high school and community college teachers are offered courses through summer workshops and year-round professional development, and future teachers are offered a course blending CSP content and pedagogy in their senior undergraduate year. About 15 regional high schools and six community colleges are participating in the program.

This new teacher training initiative complements SDSC's highly successful TeacherTECH program which helps educators bring the latest science discovery technology tools and technology-enabled science concepts into the K-14 curriculum. Selected workshops from its widely acclaimed TeacherTECH Science Series are broadcast locally on UCSD-TV and nationally on UCTV, the satellite channel for UC. Created in 2001, TeacherTECH has to-date attracted more than 8,000 teachers from more than 250 area schools, reaching more than 140,000 students annually.

Summertime Sessions at SDSC for the Next Generation

Summertime at SDSC is anything but a vacation from learning for students from regional middle and high schools. A record 340 students from San Diego and Los Angeles, plus parts of northern California, Nevada, and Arizona, signed up for a wide range of science and technology workshops as part of SDSC's 2012 StudentTECH Summer Program. Students enrolled in SDSC's 17 courses that ranged from basic computer programming and learning how to make one's computer more secure to creating animations, video games, and developing websites. Other workshops explored the world's oceans and extreme weather events through actual research-generated geospatial datasets.

"The StudentTECH Summer Program has come a long way since we started it in 2006, when we offered only one workshop and had 14 students," said Ange Mason, SDSC's education program manager. "This remarkable growth underscores the heightened interest among middle and high school students in computer science and technology, and this program offers a valuable, fun experience for them as they look ahead to continuing their education."

SDSC also offered unpaid summer internships to high school students interested in computational science and research. The seven-week program, called Research Experience for High School Students (REHS), pairs SDSC mentors with students, who work as part of a team of researchers. During the past summer, REHS interns were offered opportunities to explore a search engine for biological data; an easy-to-use scripting tool for use in earthquake visualizations; privacy consent permission/instruction challenges and issues for sharing data such as medical records and bio-samples; further development of a nationally funded immunology and protein structure database; and more.

Undergraduates also filled SDSC's halls and offices during the past summer under a new internship program designed to give these students paid, hands-on experiences using *Gordon*, the Center's unique data-intensive supercomputer. Offered to a select group of students who have completed coursework in

electrical engineering, computer science, or a related field, the 10-week internship focused on application performance analysis and high-performance computing systems deployment. Selected students had an opportunity to work closely with SDSC staff with a range of technical backgrounds (computer science, physics, chemistry, and engineering).

Graduate students, too, are part of SDSC's community. They are vibrant and vital members of SDSC staff research teams through collaborative projects with UCSD faculty members Steven Swanson, Tajana Rosing, Yoav Freund, and others.

"Big Data" Skills for Today's Workforce

The world is seemingly being overwhelmed by a data tsunami. But many researchers now believe that answers to key questions across a wide spectrum of scientific and engineering disciplines will be discovered somewhere in this enormous volume of data.

Clearly, no one can slow or stop the deluge ... nor should they. Instead, SDSC is embracing the challenge of harnessing "Big Data" for science and society. During the past year, the Center launched the Predictive Analytics Center of Excellence (PACE), aimed at leveraging SDSC's data-intensive expertise and resources to help create the next generation of data scientists interested in Big Data and its related field of predictive analytics.

In October, PACE held its first industry-academia event, a "data mining boot camp", providing entry-level and mid-career professionals in business and academia with improved tactics needed to design, build, verify, and test predictive data models. This quarterly two-day hands-on course, organized by Natasha Balac, director of PACE, complements workshops and tech talks open to the public in the growing discipline of data mining. In another initiative, SDSC's Center for Large Scale Data Systems (CLDS), headed by Chaitan Baru, is developing an executive education program in Big Data and Cloud Computing in partnership with UCSD's Rady School of Management. In the coming year, SDSC expects to refine and expand its role in this critical area of workforce education.

Partners with Industry and External Research Institutes for Regional Economic Impact

The capabilities and expertise established by SDSC over the past decade have intersected with a broader trend in information technology known as "Big Data." The increasing capabilities of sensor and storage systems mean that organizations can collect vast amounts of digital data, sift through all this data with sophisticated computer programs, and then derive previously undetectable relationships and trends for business, science, and even political campaigns … as illustrated in the recent presidential election.

The escalating interest in Big Data has many organizations looking for technological partners to provide computing and storage capabilities, and to conduct state-of-the-art research and innovation in this field. Toward that end, SDSC is developing external partnerships and collaborations with corporations and external research institutes in the greater San Diego region, creating technology transfer, employment, and other economic development opportunities. Some of these relationships include:

Genetics and Bioinformatics

The advent of next generation DNA "sequencing" instruments has resulted in an avalanche of data from sequencing (decoding) the genetic material of humans, other plants and animals, and microbes. Knowledge is gained from this sequence data using computer tools and technologies in the field of "bioinformatics." SDSC has partnered with research institutions such as The Scripps Research Institute and corporations such as Sequenom to provide computing and storage systems, in addition to bioinformatics expertise, to support genetics research and development.

In collaboration with the Scripps Translational Science Institute, a research partnership of The Scripps Research Institute and Scripps Health, SDSC has supported studies to collect and analyze the genetic material in cancerous tumors. This work is contributing to the field of "translational science" which involves using results from the scientific research domain, e.g., measuring and analyzing the genetic material from cancerous tissue, to inform practical clinical treatments, e.g., administering a specific anti-cancer drug based on the genetic "signature" of a tumor.

SDSC is working with other research institutes and companies on the Torrey Pines "Mesa," such as the Sanford Consortium for Regenerative Medicine, the La Jolla Institute for Allergy and Immunology, and the Salk Institute, to provide high-performance computing and storage support for wide-ranging studies and projects using next-generation DNA and RNA sequencing technologies. The demand for data storage, computational capabilities, and bioinformatics support is expected to grow as next-generation sequencing continues to be adopted as a critical component of genetic research and medicine. SDSC anticipates playing a key role as a local repository for such capabilities and expertise in San Diego's biotechnology community.

Energy Production and Efficiency

UC San Diego researchers in engineering, physical sciences, and other disciplines are pioneering new technologies for "green" and renewable energy. One such effort has been the deployment of a campus green energy "micro grid" consisting of solar panel arrays, waste gas fuel cells, and other technologies providing a "working laboratory" for understanding the performance, behavior, and management of green energy systems. In keeping with the Big Data trend, the campus micro grid has been "instrumented" with sensors that continuously collect many elements of data such as power output, power consumption, solar intensity, etc. SDSC has been contributing to this effort by employing sophisticated "predictive analytics" and data mining techniques to analyze the micro grid data and develop insights about the performance and optimization of renewable energy resources operating within an electrical grid. The analytic techniques developed by SDSC may one day be commercialized to support monitoring and control of renewable energy grids throughout the region, and beyond.

Modeling the Weather

Longtime residents of San Diego are familiar with the dangers of wildfires in the region, especially the devastating fires in 2003 and 2007. As a substantial stakeholder in the region's infrastructure, San Diego Gas & Electric (SDG&E) is investing in tools and technologies for early warning of dangerous wildfire conditions. Having advance notice of critical conditions for the potential ignition of a wildfire, particularly if the conditions can be highly localized, would permit SDG&E to put its resources on alert and to monitor and deploy resources to locales where they can be most effective. To this end, SDG&E has partnered with the National Weather Service and the University of California, Los Angeles, to develop a high-resolution weather forecasting model that can help predict dangerous conditions, such as Santa Ana winds, on a highly localized basis within San Diego County. Such a model, intended to be run multiple times per day, 365 days a year, requires significant high-performance computing resources to generate results within the required timeframes. SDSC has assisted SDG&E by providing computational support on its *Triton* high-performance cluster computing system. Using *Triton*, SDG&E and its partners have been able to test their model and evaluate their high-performance computing needs prior to purchasing a dedicated system to support operational use of the weather model.

"BIGDATA100" to Benchmark "Big Data"

The recent explosion in "Big Data" has spurred interest among academia and the burgeoning multi-billion dollar Big Data industry in developing benchmarks to help quantify hardware and software performance on Big Data tasks and applications. SDSC researchers are coordinating and providing intellectual leadership toward the creation of a 'BigData100' list, the first global ranking of its kind of Big Data systems, blending benchmarking approaches from high-performance computing, transaction processing, and database query processing.

The BigData100 list will rank systems according to their performance as well as price/performance. Cost is a key factor for Big Data applications, including energy costs, total cost of acquisition (TCA), and the total cost of ownership (TCO). The new benchmarking approach and ranking would be complementary to other such lists in high-performance computing community, including the Top500 and Graph500 rankings.

An initial board of directors has been formed to steer this activity, coordinated by Chaitan Baru, director of the Center for Large-scale Data Systems research (CLDS)—a new industry-sponsored center of excellence created within SDSC to develop concepts, frameworks, analytical approaches, and systems solutions to address technical as well as technology management challenges facing information-intensive organizations in the era of Big Data. The board includes Miland Bhandarkar, EMC/Greenplum; Dhruba Borthakur, Facebook; Eyal Gutkind, Mellanox; Jian Li, IBM; Raghunath Nambiar, Cisco; Ken Osterberg, Seagate; Scott Pearson, Brocade; Meikel Poess, Oracle; Tilman Rabi, University of Toronto; Richard Treadway, NetApp; and Jerry Zhao, Google.

"Big Data is now integral to almost every aspect of science and society and, indeed, to the functioning of the global economy," said Baru, SDSC's associate director of Data Initiatives. "The classical definition of Big Data is in terms of volume, velocity, and variety of data. In our benchmarking efforts, we are also interested in identifying a canonical set of analytics pipelines that can capture the essential computational characteristics of a wide range of Big Data applications. This is an exciting area of study with important ramifications for developing objective measures and profiles of system performance."

CLDS held the First Workshop on Big Data Benchmarking, co-sponsored by the National Science Foundation and industry sponsors, on May 7-8, 2012 at the Brocade Executive Briefing Center, San Jose. A report from this workshop was presented at the Fourth International Conference on Performance Evaluation Benchmarking, co-located with the 38th International Conference of Very Large Data Bases, July 2012, Istanbul, Turkey. The second workshop was held December 17-18 in Pune, India-.

Works with Research and Community Partners to Harness "Big Data" for Society

Because SDSC is part of a broader community—a research university based in San Diego which serves local, state, and national constituencies—its leadership and staff have placed a premium on how its technological advances and expertise could best serve society-at-large. At times, the Center's expertise has been called on to help state officials understand and prepare for potential natural and manmade disasters; at other times, SDSC researchers have been enlisted to use its resources during crises, in *ad hoc* collaborations with volunteer and government organizations, to save lives and property.

Take, for example, the High Performance Wireless Research and Education Network (HPWREN), led by Principal Investigator Hans-Werner Braun at SDSC and Co-Principal Investigator Frank Vernon at the Scripps Institution of Oceanography, in collaboration with scientists at San Diego State. As part of its daily mission, HPWREN moves large quantities of visual data in real time over wireless networks for a variety of research projects, from astronomical observations at the Palomar Observatory to animal movements on ecological reserves, in addition to setting up educational networks for the Native American tribal reservations in Southern California and beyond.

HPWREN now works with CalFire, providing wireless capabilities to remote regions across San Diego County and beyond. Credit: HPWREN

During the summer of 2006, HPWREN researchers were recruited by another organization: the California Department of Forestry and Fire Protection (CDF). Then, thousands of acres of San Diego County were being threatened or already ablaze by a massive wildfire known as the Horse Fire. CDF called on HPWREN to establish a critical communications lifeline for firefighters battling the fire in the county's eastern region, in the Cleveland National Forest. HPWREN researchers set up hardware at key points to allow firefighters in remote locations to communicate via a wireless link from the Horse Fire incident post to the Internet, providing video data that was needed to assign resources to dangerous hot spots. Thanks to those efforts, many homes

SDSC, in collaboration with the American Red Cross and Microsoft, developed data and technology for the "Safe and Well" website, used to reconnect missing friends and relatives during times of disaster such as hurricanes and tornadoes. The "Safe and Well" website is hosted at SDSC. Credit: American Red Cross

and potentially lives were saved. Since then, CalFire has made HPWREN part of its arsenal to fight wildfires, providing wireless capabilities to remote regions across San Diego County and elsewhere.

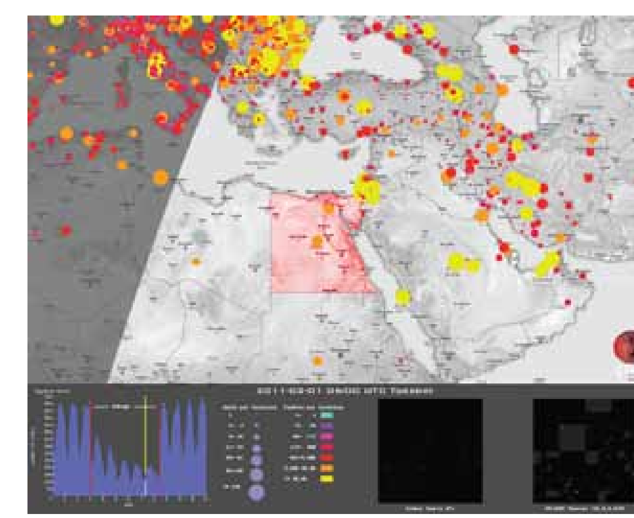
SDSC's researchers have also volunteered to use their expertise in times of national disaster. During the summer of 2005, for example, Hurricane Katrina crashed into the Gulf Coast of the nation, creating one of the greatest natural disasters in the country's history. In its aftermath homes were torn down and families were scattered far and wide. Many ended up in shelters, athletic arenas, or homes hundreds and thousands of miles away. That flight started a great separation of friends and families. Missing persons lists sprang up, seemingly every place that housed evacuees. SDSC couldn't do much to stem the flood waters, but data scientists led by SDSC's Chaitan Baru did something about the deluge of missing persons. In collaboration with the American Red Cross and later Microsoft, the team amalgamated names from more than 30 independent lists into a single, easy-to-use, master list. The result was a "missing and found" superlist of those displaced by the hurricane, processing more than 250,000 names from multiple sources for public website dissemination during the emergency. Today, that data base has evolved into the "Safe and Well" website, sponsored by the American Red Cross and hosted at SDSC.

SDSC also is collaborating with several local, state, and national organizations to help Southern California understand and prepare for other natural and manmade disasters, including earthquakes and the impact of climate change along the California coastline. SDSC and the Southern California Earthquake Center (SCEC) have formed a close partnership that has resulted in significant programs in earthquake research. SDSC introduced SCEC earthquake scientists to National Science Foundation (NSF) supercomputing resources, and have offered its expertise to provide the largest and most detailed simulations ever of an 8.0-magnitude "wall-to-wall" earthquake on a 230-kilometer stretch of the San Andreas Fault. SDSC's Amit Chourasia created visualizations of the "Big One," helping scientists gain new insights into how such a massive quake might shake Southern California and affect tall buildings in the Los Angeles area, where peak predicted motion would reach more than 3.5 meters per second.

In 2010, SDSC researchers launched, in partnership with the office of California's Chief Information Officer CIO, the California Spatial Data Infrastructure Project. Among other things, the project provides 96 terabytes of dedicated storage including the California Coastal Atlas which is being used to assess the impact of sea level rise from climate change along the California coastline.

Monitoring Internet Censorship through the Haze of "Malware" Pollution

To help explain how governments disrupt the internet, witnessed during the uprisings in Egypt and Libya in early 2011, a team of scientists—led by k.c. claffy, founder of the Cooperative Association for Internet Data Analysis (CAIDA) at SDSC—conducted an analysis based largely on the drop in a specific subset of observable Internet traffic that is a residual product of ubiquitous traffic pollution, sometimes referred to as "malware." The analysis represented the first published research to demonstrate how malware-generated traffic pollution can be used to analyze Internet censorship and other macroscopic network outages from natural disasters, including earthquakes and hurricanes. The researchers believe this novel methodology could be adopted on a wider scale to support an automated early warning system to help detect similar Internet disruptions in the future.

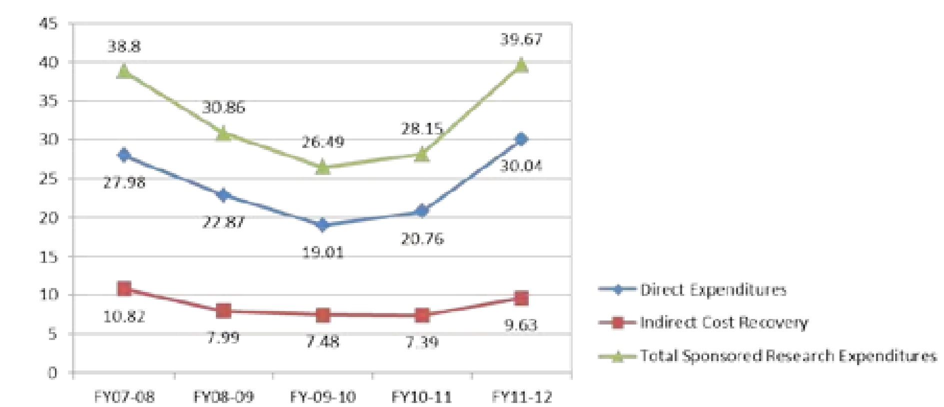


Facts and Figures

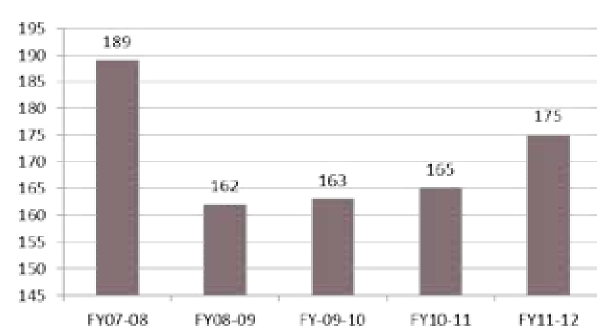
Proposal Success Rate FY11 - FY12

FY11-12	PROPOSALS SUBMITTED	PROPOSALS FUNDED	SUCCESS RATE
DOE	6	2	33%
INDUSTRY	6	1	17%
NIH	11	6	55%
NSF	74	28	38%
OTHER FEDERAL	17	13	76%
STATE	2	1	50%
TOTAL	116	51	44%

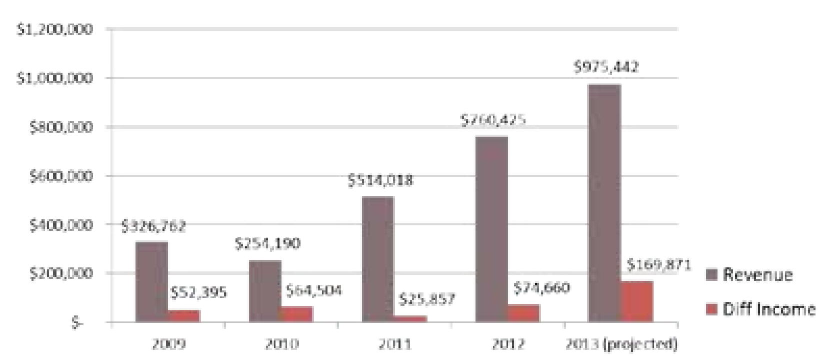
Sponsored Research Expenditures ($M) FY08 - FY12



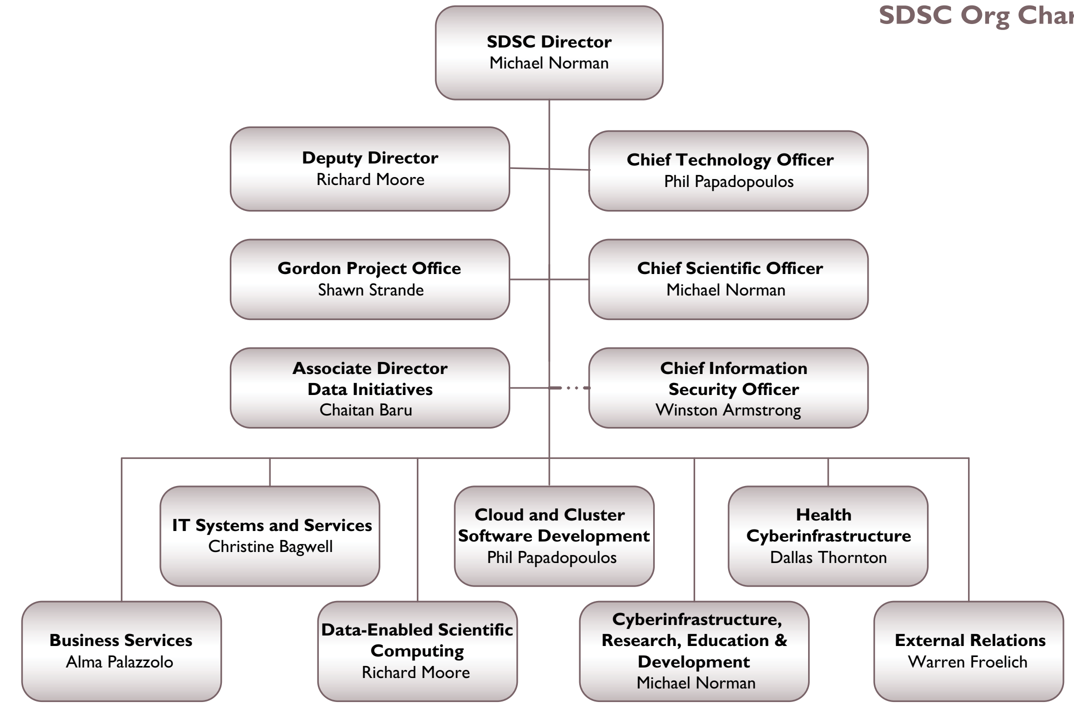
Sponsored Research Awards FY08 - FY12



Industry Revenue Data FY09 - FY13



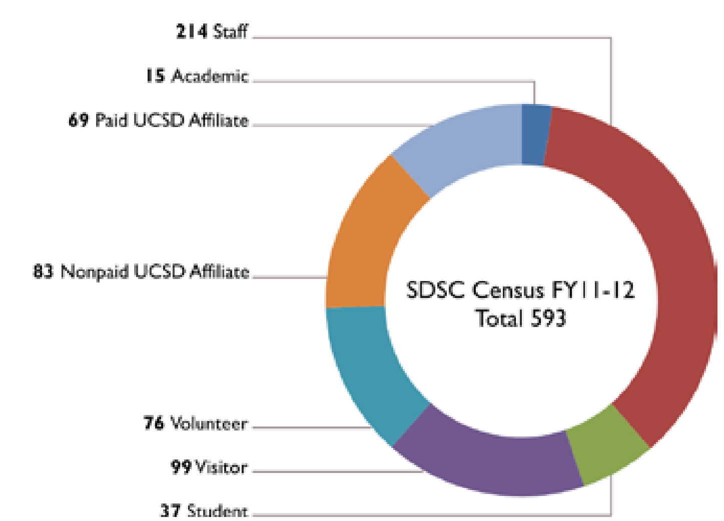
SDSC Org Chart



An Impressive Record for Attracting Federal Grants

Apart from the extraordinary research impact of SDSC collaborations and partnerships, a quick look at the fiscal impact of these collaborations is impressive. During its 27-year history, SDSC revenues have exceeded $1 billion, a level of sustained funding matched by few research units in the country. During the six-year period from July 2005 to June 2011, SDSC received a total of 310 awards for more than 50 Principal Investigators (PIs) yielding an aggregate $250 million in extramural funding. In perhaps the most competitive landscape for federal funding in the last two decades, SDSC's success rate on federal proposals is currently 44%, compared to a national average of roughly 15% for computer science and engineering proposals at the National Science Foundation. SDSC's outstanding reputation offers a competitive advantage and leverage for those seeking government grants—small and large—for their research.

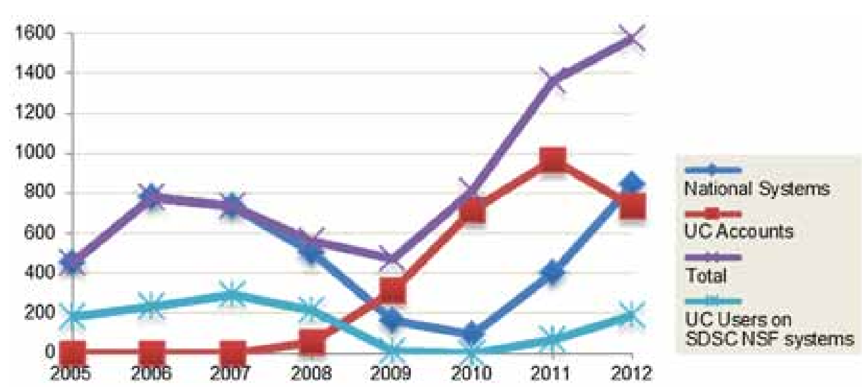
SDSC Census FY11 - FY12



Facts and Figures

Researchers Across UC Using SDSC's HPC Resources

	NSF Systems (2005-2008) Averaged Info		NSF Systems (2011-2012) Aggregated info		UC Systems (2011-2012) Aggregated info	
	# users	# core-hours	# users	# core-hours	# users	# core-hours
UC Berkeley	39	2,210,362	24	2,488,066	29	5,845
UC Davis	9	289,222	11	1,605,303	5	1,357
UC Irvine	6	107,819	8	1,362,747	18	1,724,978
UC Los Angeles	21	809,193	63	7,523,267	34	4,815,590
UC Merced	1	3,820	0	0	22	2,382
UC Riverside	8	57,578	4	473,004	15	74,481
UC San Diego	87	2,102,274	112	8,937,471	1285	4,850,809
UC San Francisco	2	7,903	10	719,571	1	0
UC Santa Barbara	59	4,777,669	11	476,225	185	4,868,175
UC Santa Cruz	1	34,536	1	6	12	0
TOTALS	231	10,400,376	244	23,585,660	1606	16,343,617



UC Principal Investigators Using SDSC HPC Resources

NAME	CAMPUS	FIELD OF STUDY
Agrawal, Divyakant	Santa Barbara	Computer Science
Ahlehagh, Hast	San Diego	Video Technology
Allen, Robert	Riverside	Climatology
Allison, Karmel	San Diego	Molecular Biosciences
Amaro, Rommie	San Diego	Chemistry
Archuleta, Ralph	Santa Barbara	Seismology
Arrar, Mehrnoosh	San Diego	Physical Chemistry
Ascenzi, Maria-Grazia	Los Angeles	Mechanics and Materials
Ashour-Abdalla, Maha	Los Angeles	Magnetospheric Physics
Asta, Mark	Davis	Materials Research
Aue, Donald	Santa Barbara	Organic and Macromolecular Chemistry
Bandrowski, Anita	San Diego	Biophysics
Baron, Riccardo	San Diego	Chemistry
Benner, Chris	San Diego	Cellular & Molecular Medicine
Boechi, Leonardo	San Diego	Biochem/Molecular Structure/Function
Boeriu, Staphan	Santa Barbara	Scientific Computing
Briggs, Cherie	Santa Barbara	Ecology, Evolution, and Marine Biology
Bullock, James	Irvine	Extragalactic Astronomy/Cosmology
Canalizo, Gabriela	Riverside	Extragalactic Astronomy/Cosmology
Carrington, Laura	San Diego	Performance Evaluation Benchmarking
Case, Marianne	Irvine	Neuroscience Biology
Chandrasekaran, Shivkumar	Santa Barbara	Electrical & Computer Engineering
Chi, Neil	San Diego	Cardiology
Christopher, Phillip	Riverside	Chemical and Reaction Processes
Colcord, Ben	Irvine	Mechanical & Aerospace Engineering
Coppinger, Judith	San Diego	Pathology

UC Principal Investigators Using SDSC HPC Resources (continued)

NAME	CAMPUS	FIELD OF STUDY
Daniels, John	Davis	Political Science
Davila, Lilian	Merced	Nanomaterials
Eisenman, Ian	San Diego	Climate Change
El Abbadi, Amr	Santa Barbara	Computer Science
El-Alaoui, Mostafa	Los Angeles	Magnetospheric Physics
Faucher-Giguere, Claude-Andre	Berkeley	Extragalactic Astronomy/Cosmology
Fowler, James	San Diego	Political Science
Freund, Yoav	San Diego	Neuroscience Biology
Gavilan, German	Merced	Engineering
Gilbert, John	Santa Barbara	Computer Science
Gilson, Michael	San Diego	Biochem/Molecular Structure/Function
Gygi, Francois	Davis	Advanced Scientific Computing
Helly, John	San Diego	Atmospheric Sciences
Holland, Chris	San Diego	Physics/Fusion Research
Houk, Kendall	Los Angeles	Chemistry
Huerta, Ramon	San Diego	Algorithm Development
Ismail-Beig, Sohrab	Berkeley	Materials Research
Ivanov, Ivaylo	San Diego	Biophysics
Jackson, Bernard	San Diego	Solar Terrestrial Research
Jiang, Steve	San Diego	Biophysics
Kang, Myungshim	Riverside	Materials Research
Kaplinghat, Manoj	Irvine	Physics
Key, Kerry	San Diego	Climate Dynamics
Kim, John	Los Angeles	Fluid, Particulate, Hydraulic Systems
Kimble, Katie	Santa Barbara	Economics
Klemke, Richard	San Diego	Pathology
Kocherzher, Aleksey	Berkeley	Physical Chemistry
Kritzuk, Alexei	San Diego	Astronomical Sciences
Lake, David	San Diego	Political Science
Li, Weizhong	San Diego	Genomics/Metagenomics
Majumdar, Amit	San Diego	Neuroscience Biology
Marsden, Alison	San Diego	Fluid, Particular, Hydraulic Systems
Martin, David	Riverside	Organic and Macromolecular Chemistry
McCammon, James Andrew	San Diego	Molecular Biophysics
McQuinn, Matthew	Berkeley	Extragalactic Astronomy/Cosmology
Mcwilliams, James	Los Angeles	Physical Oceanography
Meiburg, Eckart	Santa Barbara	Mechanical Engineering
Meng, Shirley	San Diego	Nanoengineering
Metiu, Horia	Santa Barbara	Chemistry and Biochemistry
Miller, Mark	San Diego	Genomics
Mofrad, Mohammad	Berkeley	Biophysics
Nichols, Sarah	San Diego	Biochem/Molecular Structure/Function
Norman, Michael	San Diego	Astronomical Sciences
Offner, Stella	Berkeley	Astronomical Sciences
Ogden, Darcy	San Diego	Volcanology and Mantle Geochemistry
Paesani, Francesco	San Diego	Chemistry
Paolo, Fernando	San Diego	Climate Change
Paxton, Bill	Santa Barbara	Institute for Theoretical Physics
Parrish, Ian	Berkeley	Astronomical Sciences
Peroomian, Vahe	Berkeley	Magnetospheric Physics
Pinzke, Anders	Santa Barbara	Physics
Primack, Joel	Santa Cruz	Extragalactic Astronomy/Cosmology
Quach, Diana	San Diego	Bioengineering
Quataert, Eliot	Berkeley	Astronomical Sciences
Ren, Bing	San Diego	Cellular & Molecular Medicine
Riguelme, Mario	Berkeley	Astronomical Sciences
Sarovar, Mohan	Berkeley	Quantum Biophysics
Sausman, Noriko	Riverside	Bioengineering
Saul, Lawrence	San Diego	Machine Learning
Shi, Chuntai	Riverside	Physics
Singh, Tajendra	Los Angeles	Training
Snavely, Allan	San Diego	Advanced Scientific Computing
Spaldin, Nicola	Santa Barbara	Materials Science
Stacey, Mark	Berkeley	Ocean Sciences
Stickley, Nathanial	Riverside	Astronomy
Sun, Yixiao	San Diego	Economics
Szydagis, Matthew	Davis	Atomic, Molecular, and Optical Physics
Tantillo, Dean	Davis	Organic and Macromolecular Chemistry
Taufer, Michela	San Diego	Chemistry
Theofanous, Theo	Santa Barbara	Chemical Engineering
Timmerman, Allen	San Diego	Economics
Turner, Kimberly	Santa Barbara	Mechanics of Microscale Systems
Weakliem, Paul	Santa Barbara	California Nanosystems Institute
Whaley, Birgitta	Berkeley	Atomic, Molecular, and Optical Physics
White, Douglas	Irvine	Anthropology
Yan, Xifeng	Santa Barbara	Algorithm Development
Yang, Tao	Santa Barbara	Computer Science
Yeo, Eugene	San Diego	Genomics, Neuological Diseases
Yu, Benjamin	San Diego	Dermatology
Zhang, Kun	San Diego	Bioengineering

SDSC Computational Scientists

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

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

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

Andreas Goetz, Ph.D.
Quantum chemistry
ADF developer
GPU accelerated computing

Robert Harkness, Ph.D.
Computational physics/astrophysics and cosmology
Petascale applications development
Large-scale data management
Intel MIC/Phi applications development

Amit Majumdar, Ph.D.
Director, Scientific Computing Applications group, SDSC
Associate Professor, Department of Radiation Medicine and Applied Sciences, UCSD
Algorithm development
Code optimization
Code profiling/tuning
Mathematical library implementation
Nuclear engineering

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

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

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

Mahidar Tatineni, Ph.D.
User Support Group Lead, SDSC
Research programmer analyst
Optimization and parallization for HPC systems
Aerospace engineering

Igor Tsigelny, Ph.D.
Research scientist, SDSC
Research scientist, Department of Neurosciences, UCSD
Project scientist, chemistry and biochemistry, UCSD
Molecular modeling/molecular dynamics
Neuroscience

Ross Walker, Ph.D.
Director, Walker Molecular Dynamics Lab
Molecular dynamics
Quantum chemistry
GPU accelerated computing

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

SDSC Data Scientists

Ilkay Altintas, Ph.D.
Director, Scientific Workflow Automation Technologies Lab, SDSC
Deputy Coordinator of Research, SDSC
Scientific workflows
Kepler Scientific Workflow System
Distributed computing
Bioinformatics
Data querying

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

Natasha Balac, Ph.D.
Director, Predictive Analytics Center of Excellence, SDSC
Director of Data Application and Service, SDSC
Data mining and analysis
Machine learning
Scientific data management
Data-intensive computing

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

Amit Chourasia, M.S.
Senior Visualization Scientist, SDSC
Lead, Visualization Services Group
Principal Investigator, SEEDME.org
Scientific Visualization
Computer graphics
Animation

kc claffy, Ph.D.
Director and Principal Investigator, The Cooperative Association for Internet Data Analysis, SDSC
Adjunct Professor, Computer Science and Engineering, UCSD
Internet data—collection, analysis, visualization
Internet infrastructure—development of tools and analysis methodologies for scalable global Internet

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

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

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

Julia V. Ponomarenko, PhD.
Senior Research Scientist, SDSC
Computational biology
Bioinformatics
Immune epitope database
Protein Data Bank

Karen Stocks, Ph.D.
Specialist, SDSC
Ocean and biodiversity informatics
Metadata
SeamountsOnline

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

SDSC

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

www.sdsc.edu