

San Diego Supercomputer Center

MAGAZINE OF CYBERINFRASTRUCTURE & COMPUTATIONAL SCIENCE

ENVISION

VOL. 22, NO. 1

SPRING 2006

On the Path to **Petascale** *Cyberinfrastructure*

A Star is Born

*SDSC Simulations Help Rule Out
Alternative Theory of Star Formation*

SDSC Grows

*Center Breaks Ground on
Building Expansion*

SDSC

ON THE COVER

Simulations of star formation by astrophysicist Richard Klein and colleagues using SDSC's DataStar supercomputer, reported in *Nature* and described on page 6 of this issue, shed new light on how stars are born. These simulations complement growing observational data, such as this infrared image of the "Mountains of Creation" in the W5 star formation region in Cassiopeia from NASA's Spitzer Space Telescope. Cover image courtesy NASA/JPL-Caltech/ L. Allen (Harvard-Smithsonian CfA). Cover design, Ben Tolo.

THE SAN DIEGO SUPERCOMPUTER CENTER

For more than two decades, the San Diego Supercomputer Center (SDSC) has enabled educational advancement and research discovery through cutting-edge computational and information technologies. SDSC is a full-service science and technology center, and provides an indispensable information technology resource to both academic and business institutions throughout the nation. SDSC enables more than 50 projects spanning astronomy to medicine to digital data management. SDSC is an organized research unit of the University of California, San Diego and one of the National Science Foundation's national cyberinfrastructure centers.

SDSC INFORMATION

Fran Berman, *Director*

Vijay Samalam, *Executive 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

Greg Lund, *Director of Communications*

greg@sdsc.edu

858-534-8314

ENVISION

ENVISION magazine is published by SDSC and presents leading-edge research in cyberinfrastructure and computational science.

ISSN 1521-5334

EDITOR: Paul Tooby

CONTRIBUTORS: Dave Dooling,

Lynne Friedmann, Paul Tooby,
and Ashley Wood

DESIGN: Beyond The Imagination Graphics

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

© 2006 The Regents of the University of California

4

SDSC Grows

On the path to advanced data cyberinfrastructure, SDSC breaks ground on an 80,000 square foot expansion of its facility on the campus of UC San Diego.

FROM THE DIRECTOR

On the Path to a Petascale and Other Frontier Goals

Working closely with users, SDSC is helping open the path to petascale computing, data, networking, and integrating software that will enable researchers to answer frontier questions



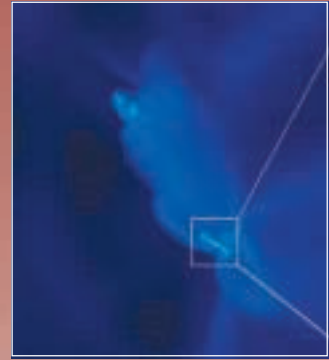
2

FEATURES

6

A Star is Born

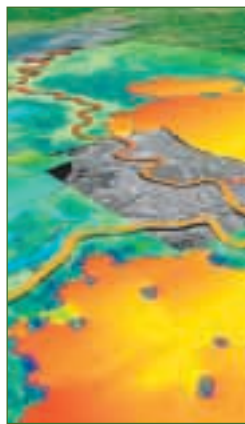
As reported in *Nature*, SDSC's DataStar supercomputer helps astrophysicist Richard Klein and colleagues gain new understanding of how stars form and rule out an alternative theory of star formation



10

SDSC to the Data Storage Rescue

Pioneering SDSC DataCentral program helps researchers hang onto valuable Hurricane Katrina data, showing the value of the first formal NSF allocations process for data storage



14

Virtual Propellers, Plumes, and Scramjets

SDSC helps user Krishnan Mahesh cross the threshold to simulations of unprecedented realism for real-world engineering flows, opening the way to faster, cheaper, better designs



18

Probing the Ultimate Building Blocks of Matter

Physicists in the MILC collaboration use SDSC's Blue Gene supercomputer to simulate the exotic world of quarks and gluons

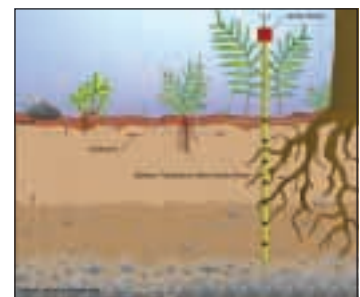
SDSC NEWS



20

THE BACK COVER

**Building Ecological
Cyberinfrastructure for NEON**





From the Director

On the Path to a **Petascale** *and Other* **Frontier Goals**

“For the research and education community, the ability to use frontier resources productively is often at least as important as their capacity, and considerable human, software, and hardware infrastructure will be required to get the most out of them.”

Petascale designates a scale of 10^{15} , and there is currently a great deal of excitement about the path to petaFLOPS HPC architectures. Achieving a petaFLOP (10^{15} floating point operations per second) constitutes a “frontier goal” in computing. In the next few years, we are likely to see the first frontier PetaFLOPS machine on the Top500 list, completing the path to a petaFLOP and commencing the path to an **exaFLOP** (10^{18} FLOPS). SDSC is working towards this goal and we endeavor that each HPC architecture we procure for our compute-intensive and data-intensive users provides an order of magnitude jump in capability and capacity.

The path to a petascale presents interesting challenges and goals beyond computing. In the data world, a number of communities already store more than a petabyte of data, including high-energy physics projects such as BaBar and Belle. When the LHC at CERN becomes operational in 2007, it will have an initial annual data output of 2 petabytes, ultimately 10 petabytes, per year. At SDSC, our tape archives currently have a capacity of 6 petabytes, and next year will have a 25 petabyte capacity (roughly 2,500 times the digital plain text equivalent of the printed collection of the Library of Congress).

The next frontier milestone in the data world is **exabyte scale collections** (1,000 petabytes or 1,000,000 terabytes). We are also moving towards the goal of **integration of knowledge management** with systems for information and data management. Such systems will provide expanded functionality, including automated execution of curation and metadata

extraction policies, dynamic checking of consistency constraints and assertions made about a collection, and automated execution of preservation policies for migration and integrity checking.

Achieving frontier milestones will have direct benefit to the most resource-hungry users and longer term benefit to many others. For the research and education community, the ability to *use* frontier resources productively is often at least as important as their capacity, and considerable human, software, and hardware infrastructure will be required to get the most out of them. As users typically need some combination of compute, data, networking, visualization, and other resources, **integrating software systems** are the critical “glue” required to enable successful applications. The combination of integrating software, frontier, and other resources constitutes critical “*Cyberinfrastructure*” required to support today’s broad spectrum of science and engineering applications.

Interestingly, the analogue “petascale” frontier goal for software is not so easily articulated. It is clear that over time, software systems are expected to be more robust, more functional, more secure, and more usable, but it is difficult to define exactly what that means in an easy-to-evaluate and objective way. It is relatively straightforward to decide if an architecture can perform at the petaFLOP level on the LinPACK benchmarks (earning an impressive place on the Top500 List), but much more difficult to determine if software is highly robust and usable.

Pioneer goals are important for the research and education community, as they define a bar to reach, and a point on the horizon to advance to. Equally important are objective ways of measuring whether we’ve gotten there. The Cyberinfrastructure community is in discussion about what “successful Cyberinfrastructure” means in a concrete way, i.e. how will we know when we reach the bar? What is the best balance of frontier and smaller-scale resources? How

will we measure (and achieve) both broad and deep impact?

As our *EnVision* readers ponder these important questions, we invite you to read in this issue about community projects that have reached important frontier domain goals: as reported in *Nature*, astrophysicist Richard Klein’s calculations have shed important new light on star formation, ruling out a widely-held alternative formation theory. Krishnan Mahesh has run the most-realistic simulations ever of real-world engineering flows such as propellers and scramjets, opening the way to faster, cheaper, and better designs. Bob Sugar and the MILC particle physics team, long among the largest users of national computational resources, have recently successfully predicted the decays of unstable D mesons, in part using simulations, providing new ideas about what may lie beyond the Standard Model of particle physics.

In addition, we have an article on SDSC’s pioneering data repository, now available to the national community,

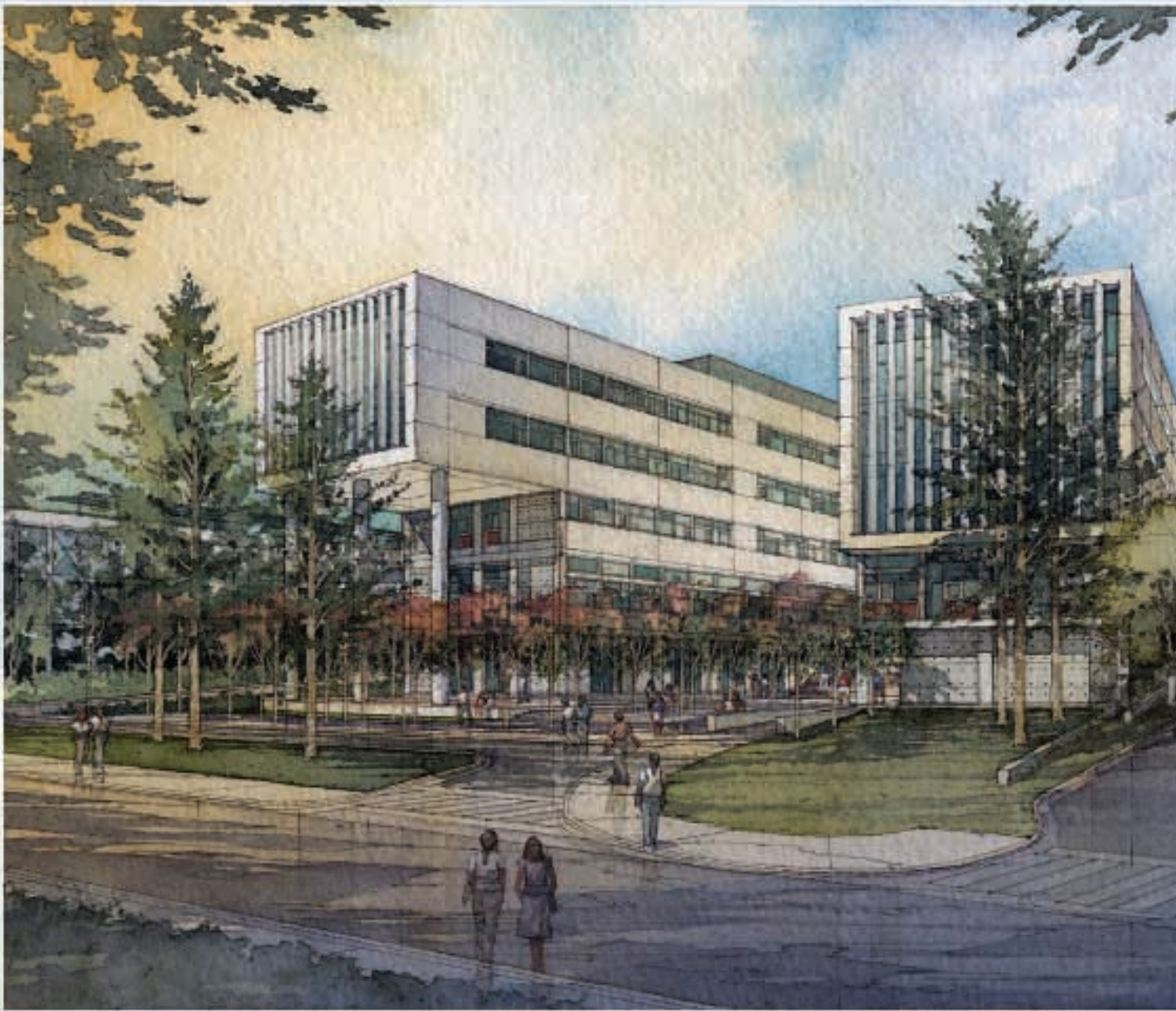
DataCentral. (For information about an allocation for your data, go to datacentral.sdsc.edu). DataCentral staff were recently able to help colleagues at LSU participating in SCOOP, an innovative coastal ocean program running sophisticated simulations of hurricane storm surge, with overflow data storage needs. The SDSC/LSU collaboration allowed the researchers to retain and share valuable data from Hurricane Katrina, along with other data in the SCOOP collection.

Finally, a bit of bookkeeping: SDSC publishes *EnVision* twice yearly, but for our readers who would like to keep up more frequently SDSC publishes a monthly newsletter, the *SDSC Nuggets*. To receive the *Nuggets*, send email to nuggets-admin@sdsc.edu, and we’ll include you on the emailing list.

We hope you enjoy this edition of *EnVision*. Let us know if SDSC can help in your path to your own frontier goals.



SDSC is working closely with users to open the path towards petascale computing, data, networking, and integrating software. The next frontier milestone on the horizon for data is exabyte scale collections. Ben Tolo.



SDSC Grows

Center Breaks Ground on Building Expansion

On June 8, 2006, the San Diego Supercomputer Center (SDSC) broke ground on an 80,000 square foot expansion of its building at the University of California, San Diego campus. The additional space will essentially double the size of the Center's facility to nearly 160,000 gross square feet.

"SDSC is an integral part of the UCSD community," said UC San Diego Chancellor Marye Anne Fox at the groundbreaking ceremony. "The center's quest for innovation in high-performance computing and data technologies is tremendous—this expansion will further advance SDSC's leadership role in these areas."

"The new building extension will facilitate SDSC's role as a national data repository and a national Cyberinfrastructure Center," remarked SDSC Director and HPC Endowed Professor Fran Berman. "What makes this building extension so important to UCSD and the nation is the human, software, and hardware resources it will house, and the discoveries it will enable."

The larger facility, scheduled for completion in June of 2008, will also promote synergy by reuniting all of SDSC's 400 staff under one roof. Previously, SDSC staff have been spread across several different locations on the UCSD campus.

Designed by the firm of EHDD (Esherick Homsey Dodge and Davis), the

SDSC expansion earned the 2005 Best Practices Award for innovative heating, ventilation, and air conditioning design from the Higher Education Energy Partnership. Using LEED (Leadership in Energy and Environmental Design) sustainability standards, the integrated design approach will adapt to local micro-climate and provide substantial energy savings.

"SDSC hosts an amazing assembly of organizations, programs, projects and, above all, people that have led the way in creating and applying cyberinfrastructure to research and education," said Dan Atkins, Director of the National Science Foundation's Office of Cyberinfrastructure, who spoke via videotape at the groundbreaking ceremonies. "Our nation is fortunate that this Center continues to serve and catalyze our research communities."

In addition to offices and meeting rooms, the new expansion will incorporate a 250-seat auditorium that can be reconfigured into smaller spaces, research neighborhoods, a high-tech conference room with sophisticated teleconferencing gear and technology tools, and an advanced visualization lab. The expansion will add nearly 5,000 square feet of machine room space, bringing total machine room area to approximately 18,000 square feet.



SDSC Expansion Highlights

Building expansion area	80,000 gross square feet
Bringing total building area to	160,000 gross square feet
Additional machine room space	5,000 square feet
Scheduled completion	June, 2008
Advanced visualization lab	
250 seat reconfigurable auditorium	
State-of-the-art teleconferencing	
Research neighborhoods	

► Far left and above: Artist's views of the 80,000 square foot expansion of SDSC's facilities. Al Forster.

Left: (left to right) UCSD Jacobs School of Engineering Dean Frieder Seible, UCSD Senior Vice Chancellor Marsha Chandler, UCSD Chancellor Marye Anne Fox, SDSC Director Fran Berman, and UCSD Vice Chancellor Richard Attiyeh break ground on the SDSC building expansion. Photo Alan Decker.

by Paul Tooby

A Star is born

*Astrophysicists Use SDSC's
DataStar Supercomputer to Rule
Out an Alternative Theory of
How Stars Form*

In the ongoing evolution of the Universe, stars continually form, develop, and die in regions of giant interstellar clouds of gas and dust. Such zones have been abundantly photographed in brilliant color by the Hubble and Spitzer space telescopes, but astrophysicists are still working to unravel the puzzle of just how new stars are born within them. What are the intricate steps that lead enough matter in these turbulent clouds to coalesce, until a star suddenly ignites with nuclear fusion and shines a new beacon out into the Universe?

In studying this mystery, astrophysicists have developed two competing theories of how stars form, with intense debate swirling around them for most of a decade. And because star formation is central to many other phenomena in astrophysics such as how galaxies form, much depends on the outcome of studies of these two competing models.

But which theory is right? Now, astrophysicists Mark Krumholz of Princeton University, Christopher McKee of the University of California Berkeley (UCB), and Richard Klein of UCB and Lawrence Livermore National Laboratory (LLNL) have reported in the journal *Nature* on research using large-scale simulations that produced decisive evidence supporting the “direct gravitational collapse and fragmentation theory” and ruling out the

other theory, known as the “competitive accretion model.”

By running carefully designed simulations on supercomputers, especially DataStar at the San Diego Supercomputer Center (SDSC) at UC San Diego, the scientists have taken into account the intertwined physical processes of gravitational attraction within a turbulent gaseous core collapsing to form a star, and have shown that the competitive accretion model fails to explain a number of key observations that astronomers have gathered from star-forming regions.

A TALE OF TWO THEORIES

Both theories predict that within giant interstellar clouds, typically 100 light years across and containing 100,000 times the mass of our Sun, clumps develop with cores that are the seeds of future stars. In the direct gravitational collapse model, a core coalesces under the influence of gravity to form a protostar. What is key is that in this theory the cores are large and, although they may fragment into smaller pieces to form binary or multiple star systems (figure 1), at birth the new stars contain nearly all the mass they will ever have.

In contrast, the competitive accretion model, developed mainly by British and European astronomers, predicts that as cores



Simulations of star formation complement the growing body of observational data, such as this infrared image of the “Mountains of Creation” in the W5 star formation region in Cassiopeia from NASA’s Spitzer Space Telescope, helping astrophysicists unravel the mysteries of how stars form. NASA/JPL-Caltech/ L. Allen (Harvard-Smithsonian CfA).

PROJECT PARTICIPANTS

RICHARD KLEIN, LLNL/UCB
CHRISTOPHER MCKEE, UCB
MARK KRUMHOLZ, PRINCETON U.
ROBERT CROCKETT, PRINCETON U.
ROBERT FISHER, U. CHICAGO
PAK SHING LI, UCB
STELLA OFFNER, UCB

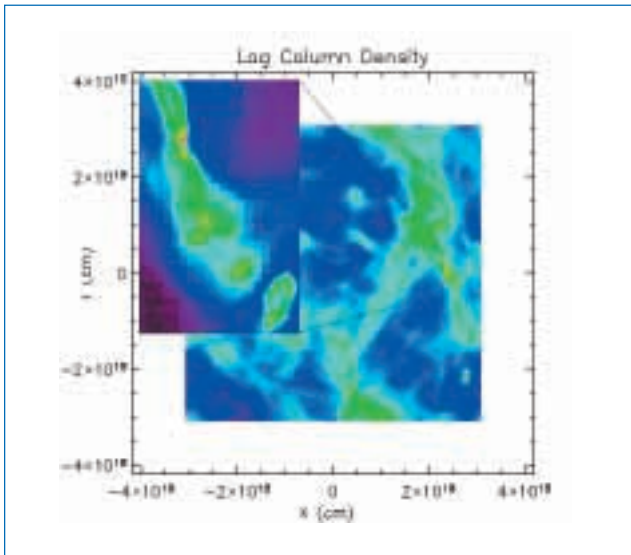


Fig. 1. Simulation of protostellar core formation along a filamentary structure in a supersonic turbulent molecular cloud. Inset shows the formation of several cores which will lead to multiple star systems, a feature frequently observed in nature. The height of the magnified region is 50,000 astronomical units (AU), with maximum resolution of six AUs (one AU is the distance from the Sun to Earth). Stella Offner, UCB.

collapse under their own gravity, they compete for gas in the surrounding clump, often gaining 10 to 100 times their original mass.

“In the competitive accretion theory, the cores are small seeds that keep growing to become stars; in our picture, the larger cores themselves turn into the stars,” explained McKee. “The observations to date, which focus primarily on regions of low-mass star formation, like the Sun, are consistent with our model and inconsistent with theirs.”

The research announced in *Nature* is expected to have a major impact on theories of star formation. “Competitive accretion has been the big theory of star formation in Europe, and we think our findings demonstrate that it has decisive weaknesses,” added Klein, a pioneering computational astrophysicist and a major user of SDSC resources since the mid-nineties.

“Star formation is a very rich problem, involving questions such as how typical stars like the Sun formed, why a very large number of stars are in binary star systems, and also how high-mass stars form that are 10 to 100 times the mass of the Sun,” said McKee. “Massive stars are important because, when they explode

in a supernova, they produce most of the heavy elements we see in the material around us.” These elements, above the weight of iron, such as zinc, copper, silver, and many others, are essential for life as we know it.

To achieve a comprehensive theory, astronomers need a model that can explain the formation of high-mass as well as low-mass stars. The competitive accretion model emerged in the late 1990s in response to problems with the earlier form of the direct gravitational collapse model, which seemed to have trouble explaining how high-mass stars form.

While theorists, among them McKee, Klein, and Krumholz, have continued to improve the ability of the gravitational collapse theory to explain the formation of large stars (figure 2), the competitive accretion theory has increasingly been found to diverge from observations.

THE ROLE OF TURBULENCE

To increase the realism of their direct gravitational collapse model, the researchers have taken pains to include what is learned from the growing body of observations of star-forming regions.



Fig. 2. In addition to simulating stars in the size range of the Sun, the astrophysicists simulate the formation of massive stars. This sophisticated adaptive mesh refinement simulation—including the full effects of radiation, turbulence, and self-gravity—follows the collapse of a turbulent gas cloud 100 times the mass of the Sun and the formation of a massive binary star system. Left panel shows entire cloud, and center and right panels zoom in by factors of 8 on the more massive star (white circle in right panel) and the formation of an accompanying accretion disk. Mark Krumholz, Princeton U.

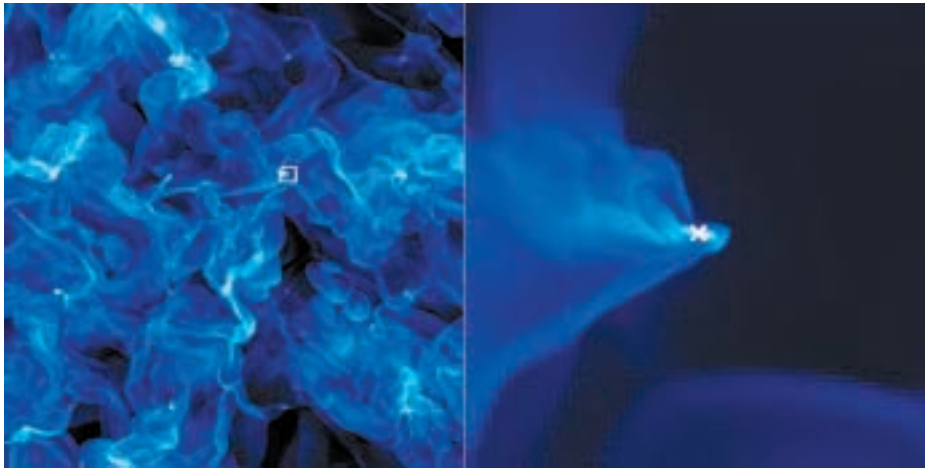


Fig. 3. Simulation of stars forming by gaining mass from a turbulent gas cloud. Left panel shows a slice of the entire cloud, with newly-formed stars in it to measure the average accretion rate. Right panel zooms in on one of the newly-formed protostars (x in right panel), surrounded by a turbulent wake. A key finding of the simulations is that the turbulent wake prevents the star from gaining further mass, supporting the direct gravitational collapse model. Mark Krumholz, Princeton U.

These observations show that turbulence characterizes environments of star formation, and the direct gravitational collapse model shows the effects of this turbulence on the rate at which a star accretes matter as it moves through a gas cloud, see figure 3. The scientists found that this is a key mechanism—the turbulence prevents protostars from being able to gather significant additional mass, as required by the competitive accretion theory.

“In our theory, once a core collapses and fragments, that star basically has all the mass it is ever going to have,” said Klein. “If it was born in a low-mass core, it will end up being a low-mass star. If it’s born in a high mass core, it may become a high-mass star.”

By including the added complexity of turbulence, the researchers’ model agrees with what is observed in nature. “Turbulence opposes gravity; without it, a molecular cloud would collapse far more rapidly than observed,” Klein said. “Both theories assume turbulence is there. The key is that there are processes going on as stars begin to form that keep turbulence alive and prevent it from decaying. Because the competitive accretion model doesn’t include this in the calculations, their model doesn’t agree with what is observed in real star-forming regions.”

A STAR IS BORN—VIRTUALLY

Astrophysicists’ understanding of the complexities of star formation continues to grow. This is reflected in models that are now far too complicated to be solved in simple, analytical ways. Computational astrophysics, a growing branch of the science, makes use of supercomputers to calculate numerical solutions for the elaborate, coupled equations.

Essentially, the scientists build a virtual molecular cloud in the computer, and under known physical processes let it evolve to form a core, and then follow the core as it collapses under gravity and fragments into protostars. And with the magic of supercomputers

they can run their model fast-forward, compressing 100,000 years of star evolution into weeks or months of simulations.

As pioneering computational astrophysicists, for over a decade Klein and his colleagues have developed highly sophisticated methods of modeling astrophysical processes. Their current code is called Orion, and it includes the strongly coupled physical processes that are known to take place. These include the mutual gravitational attraction of all the matter, the interacting magnetic fields and turbulent hydrodynamic flows known as magnetohydrodynamics, and the effects of the radiation—light given off by the stars—as well as the effects of feedback from the protostellar winds that develop from the release of energy, both gravitational and nuclear fusion, in the newly formed stars.

“Even with supercomputers, we still face major computational challenges in simulating star formation,” said Klein. One requirement is that the researchers need to follow the evolution of the collapsing core long enough for the protostars to gain sufficient mass from the cores to be clearly formed.

“We have to run a very long time, and this long collapse covers a very large dynamic range, from the largest scales of the initial clump down to the finest scales of the protostar—a factor of 15,000 to 30,000 times smaller,” said Klein. This can only be accomplished to the accuracy of the simulations with a very sophisticated adaptive mesh refinement code, which the researchers have pioneered in astrophysical simulations, and which lets the code automatically zoom in to track the key features of the huge initial clump as it collapses into a volume thousands of times smaller in the protostar.

For just one simulation with a given set of initial conditions, the calculations can take in excess of 100,000 processor hours. The power of a parallel supercomputer can be seen from the fact that a 100,000 hour allocation can be run on 1,000 processors in 100

hours, while running on a single-processor computer would take more than 11 years!

The more different kinds of physical processes the scientists include and the more closely they are coupled, the more intense communication the code requires among processors, so that the Orion code scales well up to 128 or 256 processors with full multi-physics (hydrodynamics, radiation, and magnetic fields), and can scale well to 512 processors or above for pure hydrodynamic turbulent simulations.

A simulation such as that shown in figure 4 takes some 800 hours of machine time on DataStar, more than one month if run continuously. It can take months to complete a single simulation, and the researchers must be very careful in choosing the simulation parameters. In addition to the computing challenges, the researchers must also deal with huge data sets. The large-scale simulations produce a full cube of simulation data—3-D results on full grids for many different variables. And even saving the data for a single run only for select points in time, say every 50 timesteps, for a run of several weeks or months produces several terabytes of data (a terabyte is 1,000 gigabytes, or about 500 million pages of standard printed text).

As a leader in the development of data cyberinfrastructure, SDSC's experience and high-end storage and archiving resources have the capacity to manage and archive this terabyte-sized data collection. "The data collection is too large to move around easily, so we analyze some of it remotely at SDSC, and pull down smaller pieces to process locally at Berkeley," explained Klein.

As data cyberinfrastructure develops, sharing data among researchers is becoming an indispensable component of scientific progress, and this research is no exception. "To extend our research, we're sharing our data with collaborators," said Klein. "For example, we're working with a group in Germany who are

looking at radiative transfer of our data cubes so we can make detailed comparisons with observations."

The challenges of validating complex codes and computational data sets remain more significant than for observed data sets. "But we're looking forward to the day when data sets such as ours will be included in emerging Virtual Observatories, to complement observational data," said Klein. "In addition to being a resource for the wider astronomical community, this will also provide valuable new insights that can improve our models."

The more they learn about star formation, the more complex the scientists find it is. "This drives us to keep adding new components to our code to more realistically model nature," said Klein. "But more complex codes require greater computational resources, and so we can only extend our models hand-in-hand with continued growth in supercomputers toward petascale computing, making it possible for us to address this grand-challenge problem in numerical simulation."

The research was supported by the National Aeronautics and Space Administration, the National Science Foundation, and the Department of Energy. Acknowledgement: This article includes material from a press release by Robert Sanders, UC Berkeley.

- Paul Tooby is a senior science writer at SDSC and editor of EnVision Magazine.

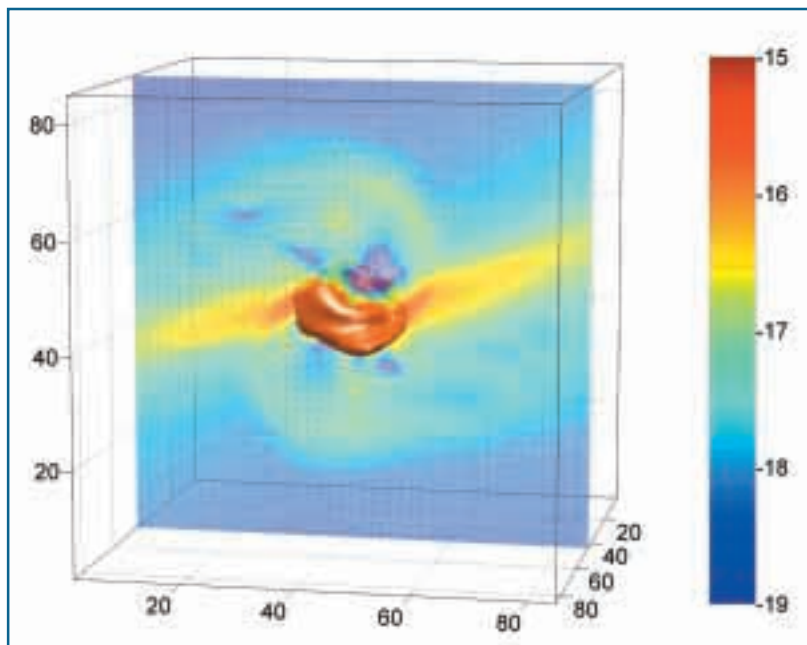
RELATED LINK

astron.berkeley.edu/faculty_pgs/klein.html

REFERENCE

Nature, 438, 332–334, November 17, 2005.

Fig. 4. Capturing the complexities of star formation, this simulation shows the core outflow in a simulation of fragmenting filaments in a turbulent cloud. The red surface of constant disk density shows the spiral structure of a twisted accretion disk surrounding and obscuring the central gravitationally collapsing core. Vectors reveal complex material flow around the core. Note the two bubble-like gas structures resulting from interaction between the outflow and accreting material. Pak Shing Li, UCB.



SDSC TO THE DATA STORAGE RESCUE

Pioneering DataCentral Program Helps Researchers Hang onto Valuable Hurricane Data

When Hurricane Katrina struck the Gulf Coast it produced widespread devastation. At the same time it presented an unprecedented opportunity for researchers at Louisiana State University (LSU) and regional partner institutions in an innovative coastal ocean program to better understand and predict future hurricanes by collecting observational data and running sophisticated simulations of storm surge and wave models for the crisis, building the project's collective knowledge base.

As they worked around the clock the Katrina data poured in. So much data, in fact, that eventually it threatened to overwhelm the team's onsite storage capacity at LSU. Even with backup tapes at LSU and online copies at Texas A&M University, the major practical difficulties of recovering the data meant that the researchers faced hard choices in creating space at LSU for new data—including the dismaying prospect of having to delete important data collected in the period leading up to the hurricane.

"We didn't want to throw away data we'd produced, but we were fast running out of onsite storage space," said Jon MacLaren, project manager for the SURA Coastal Ocean Observing and Prediction Program, known as "SCOOP" (see sidebar).

Fortunately, MacLaren and his colleagues heard about DataCentral, a unique new data storage service available through the San Diego Supercomputer Center (SDSC) at UC San Diego. An expedited request was approved in October, 2005 and SCOOP was able to upload one terabyte (1,000 gigabytes) of critical data to DataCentral using SDSC's Storage Resource Broker (SRB) data grid software.

"DataCentral came along at just the right time," said MacLaren. "At the time we got the data allocation, the Katrina data sets were the next thing we would have had to delete."

As projects like SCOOP endeavor to move cutting-edge research capabilities into the operational arena where they can benefit society through improved predictions of natural disasters, DataCentral's ease of access and data storage capability beyond that of the researchers' home environments is proving to be a crucial component of the emerging national cyberinfrastructure.

PIONEERING DATA SERVICES

Today's scientific and engineering research generates data at astounding rates. The product of millions of hours of experimental observations or computation time, data generation has reached the point where its storage and management have become a significant challenge. Left unaddressed, this growing problem can impede collaborative research efforts and hinder new discoveries, and in the worst case result in the loss of irreplaceable intellectual capital.

In response to this need, the National Science Foundation asked SDSC to spearhead a new data allocation effort. The result: In August 2005, SDSC launched DataCentral, a unique, first-of-its-kind program that offers data storage free of charge, through a competitive request and allocation process, to the US research community as a national repository for large community data collections and databases. DataCentral currently offers a large portion of SDSC's more than seven petabytes of online disk and archival tape capacity (one petabyte is one million gigabytes, about 100,000 times the storage of a typical 70 gigabyte PC disk).

Among the nation's supercomputer centers, SDSC is a recognized leader in creating data cyberinfrastructure for scientists and engineers, and when initial word of DataCentral began to circulate—through e-mail, the Web, presentations at conferences, and word of mouth—the result was a flood of applications, demonstrating that DataCentral is addressing an urgent need in the scientific community.

"We're receiving positive feedback from our new data users on the services and support DataCentral provides them," said Natasha Balac, Data Services Group Lead. "And this is leading to additional users from referrals by their colleagues."

DataCentral staff assist users with all aspects of data management, from data migration and upload, database selection, schema design, and query tuning, to data analysis and mining, portal creation, and collection publication. "DataCentral is very customer driven," said Balac. "As we strive to meet the needs of the diverse projects that come to us, we're learning more about community needs that we are applying to extend DataCentral's capabilities."

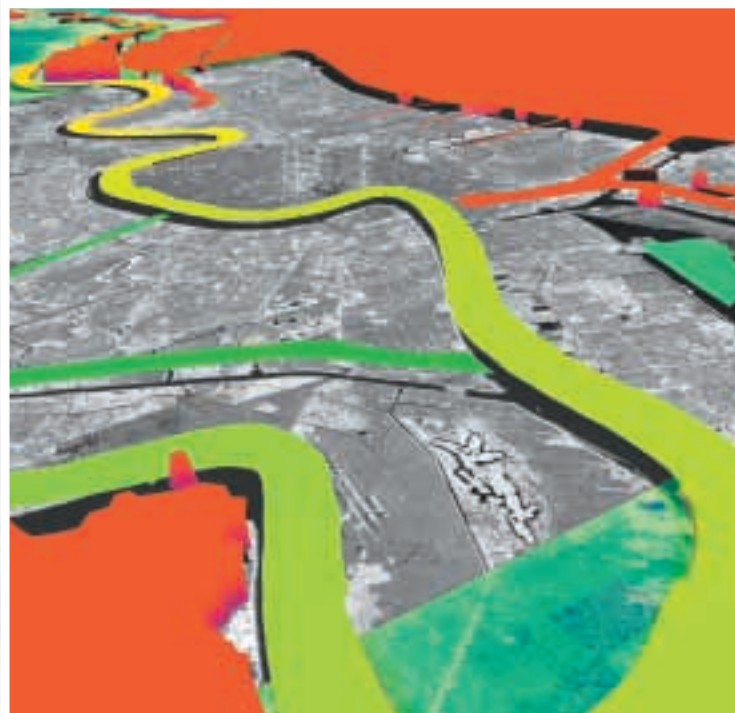
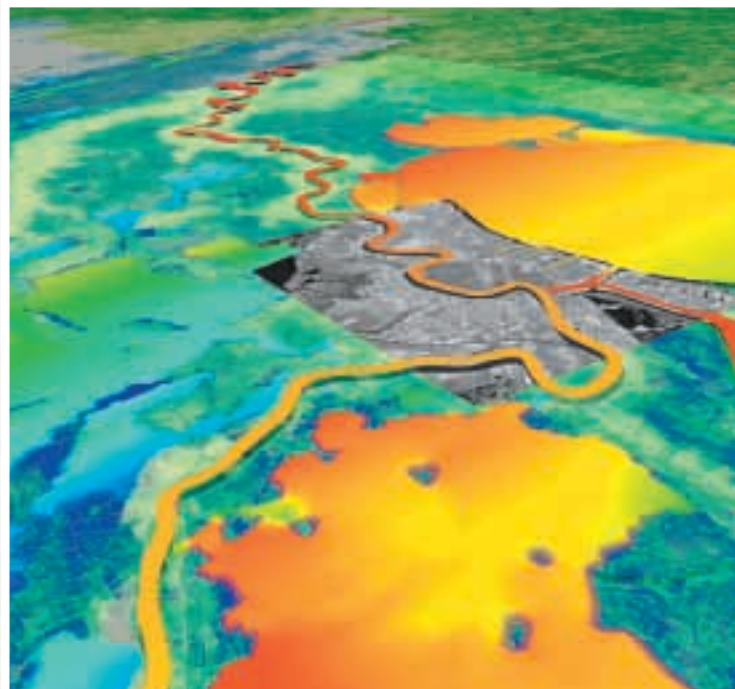
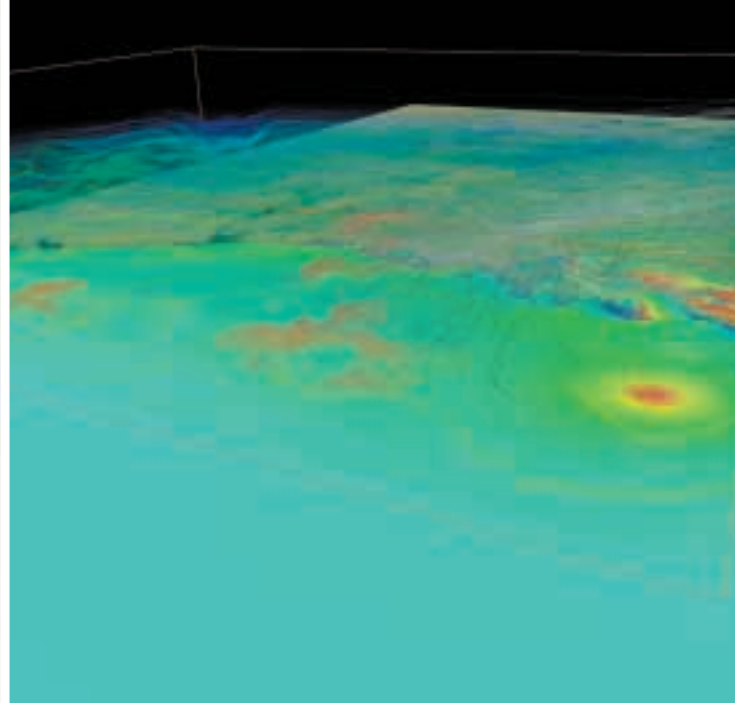
by Lynne Friedmann

www.sdsc.edu

Top: Simulations predicting the approach of Katrina to New Orleans. The wind fields of Katrina are shown as white/grey ribbons, clearly showing the hurricane vortex. The yellow to red coloring beneath the eye of the hurricane shows the storm surge moving across the gulf, pushed by the hurricane's wind. W. Bengert and S. Venkataraman, CCT/LSU.

Middle: The storm surge is shown by the yellow and red areas, overlaid on top of a map of New Orleans. Red is the highest water level. At a later stage of the simulation, the surge can be seen on the right coming in from the Mississippi River Gulf Outlet towards the Industrial Canal. In the lower part of the frame, St. Bernard Parish is flooded. W. Bengert and S. Venkataraman, CCT/LSU.

Bottom: The storm surge has reached Lake Ponchartrain, resulting in the water level rising quickly. W. Bengert and S. Venkataraman, CCT/LSU.



DESIGNING DATA ALLOCATIONS

While DataCentral storage capabilities are large, they are a finite resource. Therefore, the data allocation process is being refined at SDSC and integrated into the same national process that allocates time on NSF-supported supercomputers across the country.

In the national allocation process, an applicant must be a researcher or educator at a US institution, a principal investigator, or a postdoctoral researcher. Initially, SDSC will focus on allocating space for data collections that are publicly available and broadly used in significant collaborations. While data storage requests may be combined with computing time requests, they may also be stand-alone. A panel of experts convenes every quarter to review the requests.

For more than 20 years SDSC has made available allocated supercomputer resources. While computational scientists and engineers are well-acquainted with how to apply for computing time allocations in this system, allocating data storage presents some unique challenges.

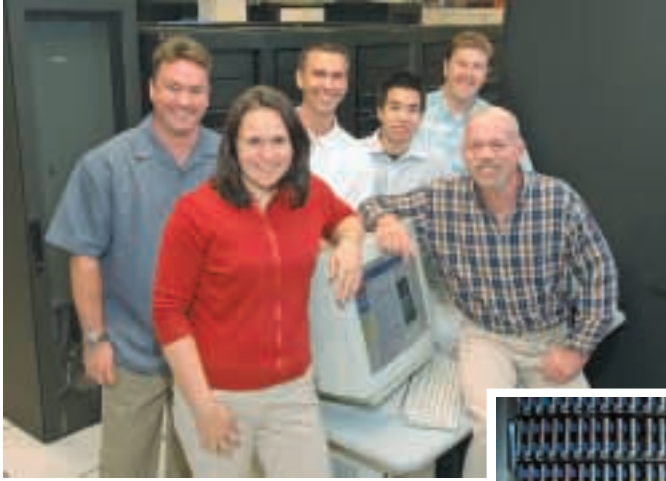
“The real trick is defining data storage in a standard way so that proposers and reviewers can compare resources and requests against one another,” said David Hart, SDSC Allocations Coordinator. “This involves not only creating useful definitions and procedures but also educating both the requesters and the reviewers.”

A key challenge for data storage is that allocations last for a limited time, typically one year, while the useful lifetime of a data collection can range from months to decades or even longer. In addition, reviewers need to weigh requests for such things as more-expensive continuous online access versus more economical lower-use archival storage.

Built into the process is an annual review of allocations that will enable DataCentral to utilize resources in the best possible way. “For example, based on usage and demand, SDSC can recommend that users would make more efficient use of the resources by migrating all or some of their data from online disk to tape storage,” said Hart.

PROJECT PARTICIPANTS

GABRIELLE ALLEN, JON MACLAREN, AND GREGORY STONE, CSI/LSU
NATASHA BALAC, DAVID HART, BRAD KROEGER, STEVE MEIER,
KEN NUNES, ROMAN OLSCHANOWSKY, AND TONY VU, SDSC/UCSD



SDSC DataCentral staff
Clockwise from left, Roman Olschanowsky, Ken Nunes, Tony Vu, Steven Meier, Brad Kroeger, and Natasha Balac, standing in front of a SATA disk array, part of SDSC's 1.4 petabytes of disk. Photo Alan Decker.

This tape silo is part of SDSC's High Performance Storage System, HPSS. With a current capacity of 6 petabytes, next year the archive will expand to a capacity of 25 petabytes (equivalent to 2,500 times the amount of the printed materials in the Library of Congress).



ADVANCED DATA MANAGEMENT TOOLS

In addition to hosting dynamic databases and long-term archiving, DataCentral fosters sharing and collaboration with these highly valuable data sets by providing expertise, infrastructure, and tools such as the Storage Resource Broker (SRB), the powerful data grid technology developed at SDSC.

"The SRB was developed because there are so many different operating systems and storage systems that do not talk to each other," said Roman Olschanowsky, SDSC Data Collections and SRB Specialist. "The ground-breaking software design makes it possible to easily access files from any one of hundreds of different types of operating and storage systems."

The SRB lets users access data files held on many distributed systems as if part of a single file system. It provides a uniform interface to access diverse types of data stored across local and wide-area networks, and maintains extensive user-controlled metadata about each stored object.

This approach offers many advantages over traditional file systems. What appears to the software user as a single collection is actually a virtual collection consisting of digital entities scattered across widespread, heterogeneous storage resources, including file systems, archives, and databases. The SRB makes all these differences transparent by negotiating protocols and access permissions with each of the multiple sites.

As scientific disciplines become more integrated, data sharing is becoming an indispensable capability. "Where the SRB is really finding a growing niche these days is in wide-area networks," said

Olschanowsky. "At DataCentral we see growing demand for the capability to store, integrate, and share data that is spread across multiple collaborating institutions." The SRB contributes in a significant way to collaborative science because it can present the same file system and file hierarchy to all the sites and finely tune data sharing and access according to the needs of individual researchers and groups.

"Until recently, the science of ocean observing in shallow water took place mostly in isolated efforts," said Gregory Stone of LSU's Coastal Studies Institute. "Now, SCOOP is building an integrated system. With SDSC's Storage Resource Broker software we can access these data sets in DataCentral through the computer Grid from anywhere in the world." This will help SCOOP researchers make their data available to the wider coastal modeling community. The data sets for the 2005 hurricane season are particularly valuable, and the SCOOP project collection in DataCentral covers Katrina, Rita, and Wilma—three of the strongest category five hurricanes on record. The SCOOP project is also planning to adapt its existing tools to interact directly with the SRB-based storage.

GROWING COLLECTIONS IN DATACENTRAL

With storage facilities offering more than seven petabytes (1.4 petabytes of online disk and 6 petabytes of archival tape storage), SDSC's DataCentral currently hosts more than 80 public data collections, including many not publicly available until now. In

addition to many large collections already at SDSC, in the short time since its inception SDSC DataCentral has already allocated more than 750 terabytes of additional storage, some 180 terabytes on disk and 575 terabytes on archival tape. These collections represent years of work by collaborating teams of researchers, and contain such information as sensor and simulation data, visualizations, and digital libraries of discipline reference collections.

Among these are databases and collections in a wide range of fields from astronomy, molecular biology, and computer networking, to earthquake engineering, ecology, geosciences, particle physics, and many more.

One of the most highly used community databases hosted at SDSC is the Protein Data Bank (PDB), a worldwide repository for the processing and distribution of 3-D biological macromolecular structure data. The PDB supports an enormous number of accesses, requires 24/7 reliability, and is central to progress in many biological sciences.

In a multi-faceted collaboration with the Southern California Earthquake Center (SCEC), SDSC researchers are helping run large-scale earthquake simulations and develop the SCEC digital library for standard seismology digital reference sets. SCEC collections at SDSC exceed 100 terabytes of data (100,000 gigabytes), covering a range of earthquake phenomena.

The need to preserve and access data extends beyond science and engineering, and SDSC is also offering these capabilities to new communities by hosting data collections from such areas as the Library of Congress and the arts community. An example is data from HASTAC ("Haystack"), a nationwide multidisciplinary consortium of humanists, artists, scientists, social scientists, and engineers who are fostering new forms of collaboration across institutions, disciplines, and communities as they promote creative uses of advanced technologies in their fields.

Building on more than a decade of research, SDSC has special expertise in addressing the challenges of long-term preservation of digital collections, some of which must be preserved for 100 years or more. Examples of this unique intellectual capital—the basis of so many advances in scientific understanding and quality of life—range from scientific collections of climate data, ecological data, and life sciences reference collections to historical archives such as presidential websites and the Shoah holocaust survivor testimony, which have essentially unlimited useful lifetimes.

To extend this vital preservation capability toward a large-scale production service, SDSC is continuing to work in collaboration with the National Science Foundation, the National Archives and Records Administration, and others to develop technologies, usage models, and support models for preserving the nation's most valuable digital collections for future generations.

- Lynne Friedmann is a freelance science writer living in Solana Beach, California.

The SCOOP Program

The Southeastern Universities Research Association (SURA) in partnership with Louisiana State University and a number of participating institutions near the Gulf and Atlantic coasts are collaborating in the SURA Coastal Ocean Observing and Prediction program, also known as "SCOOP."

Working toward an integrated and coordinated coastal ocean observing and prediction system, SCOOP is leveraging emerging regional efforts and cutting-edge Grid technologies to provide a prototype of a distributed national laboratory for coastal research and operations. The program focuses on numerical modeling, real-time data exchange, and 24/7 operational prediction and visualization for storm surge, wind waves, and surface currents, with special attention on predicting and visualizing phenomena that cause damage and flooding of coastal regions during severe storms and hurricanes.

SCOOP's partners include LSU, University of Alabama at Huntsville, Texas A&M University, Gulf of Maine Ocean Observing System (GoMOOS), University of Florida, University of North Carolina at Chapel Hill, National Oceanographic and Atmospheric Administration (NOAA), University of Maryland, University of Miami, and Virginia Institute of Marine Science. The program is funded jointly by the National Oceanic and Atmospheric Administration (NOAA) and the Office of Naval Research (ONR).

RELATED LINK

SURA Coastal Ocean Observing and Prediction (SCOOP)
scoop.sura.org/

Virtual Propellers Plumes *and* Scramjets

SDSC Users Simulate Real-world Engineering Flows with Unprecedented Realism

With today's advanced technologies, it might seem that engineers design ship propellers, jet turbines, and scramjet engines using precise engineering methods that always yield perfectly efficient designs. But in fact the real-world turbulent flows that occur in these devices are so complex that engineers are still forced to depend heavily on trial and error experience and approximate parameters in tasks from designing jet engines to predicting how far pollutants will spread from industrial smokestacks.

Now, Krishnan Mahesh, associate professor in the Aerospace Engineering and

Mechanics Department at the University of Minnesota, and his group are applying the power of supercomputers at the San Diego Supercomputer Center (SDSC) at UC San Diego to conduct a wide range of simulations of unprecedented realism. "Traditionally, such high-fidelity simulation methods have been restricted only to fairly simple geometries," said Mahesh. "But massive parallel computing platforms like SDSC's DataStar have now made it possible to simulate complex engineering flows that would have been inconceivable a decade ago."

What is novel about Mahesh's work is that his group has developed numerical

methods and turbulence models that are flexible enough to handle complicated real-world engineering geometries, without compromising the accuracy needed to reliably simulate turbulence.

By pushing forward the state of the art, the researchers' simulations open the door to understanding the intricate physics of turbulent flows in a range of engineering applications. This knowledge can help design better devices, as well as guiding researchers in devising new experiments that work in tandem with simulations to yield new insights into these important engineering problems.

EMERGENCY PROPELLER CRASHBACK

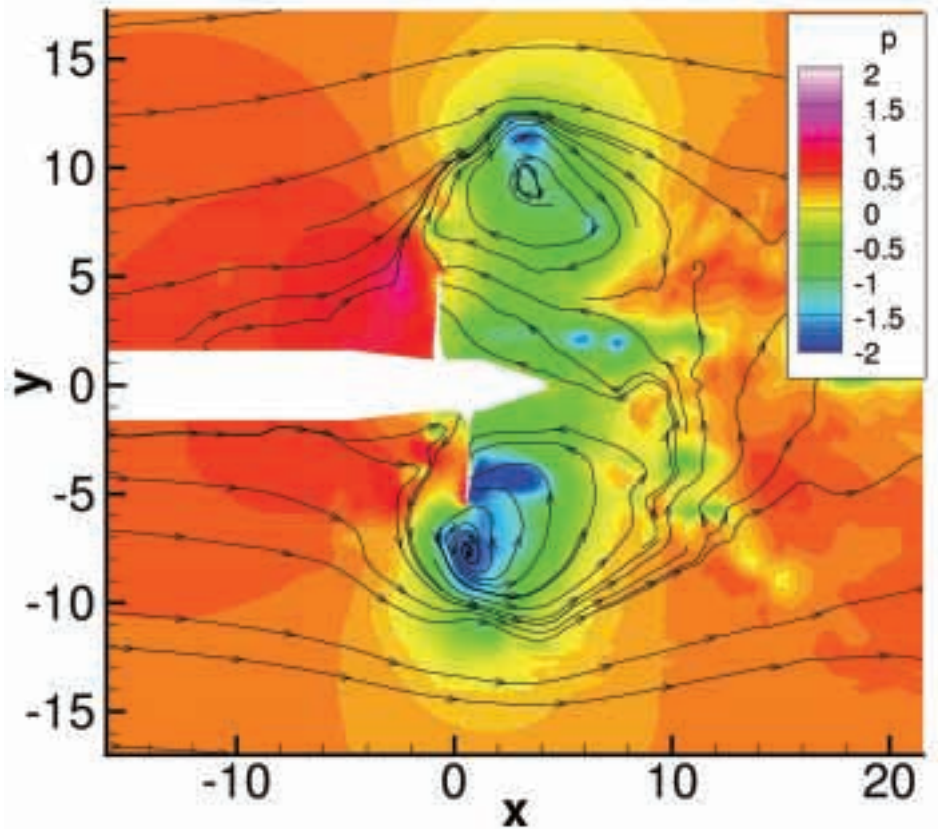
Imagine a large ship steaming along at full speed. Suddenly, there's an emergency and the ship needs to stop or turn, and the captain throws the propeller into full reverse. This dramatic propeller reversal produces large fluctuating forces on the blades that can sometimes impair the ability to steer the vessel or even break a propeller blade.

by Paul Tooby



Hypersonic scramjets show promise of providing cost-effective access to space. SDSC supercomputers powered simulations by Mahesh and his group using novel algorithms that compute accurate simulations of compressible turbulence, which can help design better scramjet engines. NASA.

Simulation of flow around a propeller in sudden reversal, known as “crashback.” Flow is left to right and shows streamlines and pressure contours in the cross-section. Note that the propeller pushes some fluid upstream, and behind the propeller is the fluctuating “ring vortex” that causes varying loads on the propeller. Vysohlied and Mahesh.



While current engineering methods are adequate to simulate a propeller operating under normal design conditions, they are unable to model the unsteadiness in blade forces that is encountered in experimental and real-world emergency crashback maneuvers.

To create a simulation approach that can capture the full complexity of propeller crashback and help engineers reliably predict required blade strength, Mahesh and graduate student Martin Vysohlied, supported by the Office of Naval Research, are performing Large-Eddy Simulations of the 3-D flow around the complicated geometry of a propeller as it is put into reverse rotation.

In a significant achievement, the researchers’ simulations are able to reproduce the low-frequency unsteadiness in forces on the propeller blade that is observed in experiments. In addition, the researchers produced simulated torque and thrust coefficients that were significantly closer to experimental values than those produced by previous methods.

This project demonstrates the potential of the researchers’ advanced simulation

methodology to predict how a propeller will perform in extreme off-design conditions such as crashback. This success has encouraged the researchers to extend this work to investigating another hard-to-predict problem—cavitation—the disruptive bubbles that can form and collapse on propeller blades under extreme conditions, leading to blade surface erosion and other problems.

SIMULATING SCRAMJETS

As the U.S. Space Shuttle fleet ages and the nation searches for cost-effective and reliable access to space, one promising technology is known as the scramjet, short for supersonic combustion ramjet. Unlike a traditional rocket, which must carry its own oxygen for combustion, a scramjet scoops air from the atmosphere, making the vehicle lighter and enabling it to carry heavier payloads. Scramjets are expected to be able to fly ten or more times the speed of sound at altitudes above 100,000 feet on the way to space, and NASA successfully tested one in 2004, the first time an air-breathing engine powered a vehicle at hypersonic speeds.

The design and testing of scramjet engines is costly, and engineers would very much like to have a reliable simulation capability that can let them build and test “virtual” scramjet engines on a supercomputer. But simulating such exotic engines is challenging, involving compressible turbulence in super- and hypersonic combustion and the interaction of shock waves with turbulent boundary layers—problems that have defied traditional simulation capabilities.

To give engineers tools to model these extreme flow regimes, Mahesh and graduate students Yucheng Hou and Jeffrey Doom have developed a novel computational algorithm that is proving able to handle the special challenges of these flows. This work is performed as part of the Air Force Office of Scientific Research-supported Center for Hypersonics at the University of Minnesota.

PROJECT PARTICIPANTS

KRISHNAN MAHESH, JEFFREY DOOM, YUCHENG HOU, SUMAN MUPPIDI, PRADEEP BABU, AND MARTIN VYSOHLIED, U. MINNESOTA



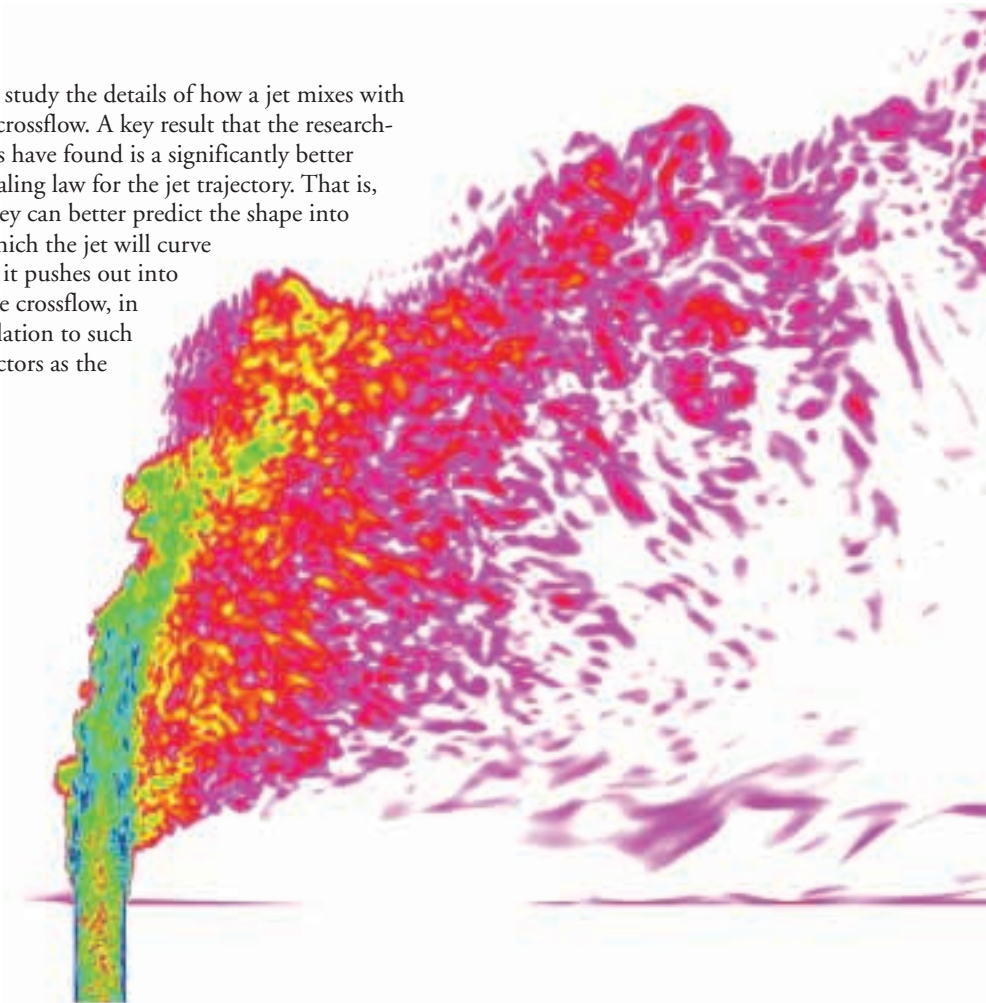
This plume rising from a smokestack on a windy day is an example of a turbulent jet in a crossflow. Simulations on DataStar by Krishnan Mahesh and his group have unprecedented realism, which can help better predict the spread of pollutants downwind and lead to improvements in health and the environment. US Fish & Wildlife Service.

VIRTUAL JET IN A CROSSFLOW

Another important simulation, which Mahesh has run for more than 45,000 processor hours on SDSC's DataStar, models a jet of fluid that shoots out into a crossflow. A plume of smoke rising from a smokestack on a windy day is one example of this situation. In this NSF-supported research, Mahesh's group wants to understand how these jets behave in order to help engineers reliably predict results such as how far a pollutant will spread downwind from a smokestack, and other applications involving jets in crossflows such as designing improved combustors for gas turbines and more efficient fuel injectors for internal combustion engines.

Graduate student Suman Muppidi has used exact solutions of the flow equations, without approximations—a method known as Direct Numerical Simulations—

to study the details of how a jet mixes with a crossflow. A key result that the researchers have found is a significantly better scaling law for the jet trajectory. That is, they can better predict the shape into which the jet will curve as it pushes out into the crossflow, in relation to such factors as the



An exact simulation, without approximations, of a turbulent jet using Direct Numerical Simulations on SDSC DataStar supercomputer. Muppidi and Mahesh.

jet's speed and size and the speed of the crossflow. This capability can help better answer key questions such as how far a contaminant will spread from a smoke-stack, or how well fuel will mix with air in an engine.

ADVANCING SIMULATION SCIENCE

“SDSC supercomputers and staff have been central to our group's ability to achieve these important new research results,” said Mahesh. In the last year, his prolific group has used nearly 90,000 processor hours on DataStar, running on up to 1,024 processors; nearly 30,000 processor hours on the TeraGrid IBM P690 large-memory nodes at SDSC, which have up to 256 gigabytes of memory per node; more than 100,000 hours on the distributed TeraGrid facility; and more than 200,000 hours on up to 512 processors of SDSC's Blue Gene machine. SDSC is initiating a new Strategic Applications Collaboration (SAC) to help the researchers improve the performance of their codes, including input/output and communications among parallel processors.

“The fast I/O capability of DataStar is a very important feature in efficient execution of our large-scale simulations,” said Mahesh. “In addition, our simulations produce large data sets, and so we archive the results in the multi-petabyte archiving facility that SDSC provides.” Because today's simulations often produce massive data sets, they can exceed the capacity of researchers' home facilities. In building data cyberinfrastructure, SDSC is taking care to construct a growing end-to-end capability, including data management and archiving for terabyte-sized data collections and beyond.

In addition to the above projects, Mahesh's group and their collaborators at Stanford University have used SDSC resources to conduct a first-time simulation of the turbulent flow inside a commercial gas-turbine combustor, in which they developed an innovative and powerful simulation methodology.

- Paul Tooby is a senior science writer at SDSC and editor of EnVision Magazine.

RELATED LINK

Krishnan Mahesh, Computational Fluids Laboratory
<http://www.aem.umn.edu/people/faculty/mahesh/>

www.sdsc.edu

Data Cyberinfrastructure for Computational Fluid Dynamics

Flowing fluids play a myriad of roles in the natural and human-made worlds. To better predict phenomena as diverse as the Earth's weather, fishery dynamics, the human circulatory system, or the behavior of the Sun, scientists must understand the complex, turbulent flows at the heart of these systems. As engineers strive to design improved devices from aircraft wings, internal combustion engines, and heart stents to pollution controls, they depend on being able to predict the associated fluid flows.

To supplement experimental and theoretical approaches, researchers are increasingly turning to computational fluid dynamics. But in this approach, where computers create “virtual” fluid flows to complement, guide, or even replace experiments, researchers have long been limited by the size of computers to modeling simplified geometries, and then having to extrapolate their approximate results to more complicated real-world engineering applications.

With supercomputers as powerful as SDSC's 15.6 teraflops DataStar, Mahesh and his group are reaching an exciting threshold—being able to simulate the major characteristics of such complex real-world flows as propeller crashback, scramjet design, and a jet in a crossflow. This promises to give engineers powerful new tools to build and test “virtual” devices, reducing the trial and error of traditional prototyping and leading to better designs.

To fulfill the promise of these more-realistic simulations, however, researchers must meet numerous challenges. A central problem in modeling turbulent flows, which have important interactions taking place across the full range of scales—from the smallest eddies up to the size of the entire flow—is that the simulation must capture all of these scales. But computer simulations work by dividing the modeled region into cells or boxes. The smaller the boxes, the higher the resolution and the more detail is captured for greater realism. At the same time, increasing the number of boxes also increases the computational demands of the model. Only with supercomputers as large as DataStar or larger can researchers simulate problems that capture the major features of the real world.

In the march toward petascale computing (supercomputers capable of 10^{15} or more calculations per second—more than 100,000 times the power of a typical PC), such real-world engineering applications will continue to make full use of the largest systems available.

Mahesh's group combines the power of ever-larger supercomputers to yield more realistic simulations with ingenious advances in numerical algorithms for computational methods on unstructured grids, including Large-Eddy Simulations (LES) and Direct Numerical Simulations (DNS). Instead of trying to model each tiny eddy in the fluid down to the smallest scales, which is not practical, the LES technique includes their influence by statistically summarizing how a group of small eddies can affect the larger-scale motion. DNS finds exact solutions of the flow equations, without approximations, giving a fully accurate simulation of the flow.

In harnessing the power of DataStar, Mahesh's simulations are noteworthy as among the first to cross the threshold of providing useful results in a range of complicated engineering problems.

SDSC's modern data cyberinfrastructure, which delivers fluid dynamics simulations of unprecedented resolution, with faster input and output, bigger data collections, and better visualizations, is opening the way to a new era of faster, cheaper, and better engineering solutions.



Probing the Ultimate Building Blocks of Matter

Physicists Use SDSC's Blue Gene Supercomputer to Simulate the Exotic World of Quarks and Gluons

by Dave Dooling

Over the years supercomputers have generated ever-more sophisticated simulations of the Universe growing outward from the Big Bang. But events on the largest scales ultimately rely on our understanding of what happens inside the tiny neutrons and protons that make up galaxies and planets, in events lasting much less time than it takes light to travel across an atomic nucleus.

Welcome to lattice quantum chromodynamics (QCD), in which scientists model how particles and forces that have never been seen interact to shape the Universe as we know it. "Chromo" is the fanciful name that describes the "color charge" by which the strong force binds quarks into protons, neutrons, and a host of other subatomic particles.

Using the new Blue Gene supercomputer at the San Diego Supercomputer Center (SDSC) at UC San Diego, the "MILC Collaboration," a group of physicists at the University of Arizona, University of California Santa Barbara, University of the Pacific, Indiana University, University of Utah, Washington University, and the American Physical Society are working to make better predictions of just what happens inside those particles.

"We're looking at incredibly small stuff," said physicist Douglas Toussaint of the University of Arizona. "There are frontiers at both ends of the space and time scales."

INSIGHTS FROM HIGH TEMPERATURES

QCD is one of two major portions of the Standard Model of high energy physics, but the model as we know it now has limits. "We want to test the Standard Model to find its range of validity and see where it breaks down," said physicist Robert Sugar of UC Santa Barbara. "That will determine the model's limits and, we hope, give hints of new physics that lies beyond our current theories."

"Forces in QCD are so strong that quarks and gluons are seen only indirectly," Sugar explained. "You can't get a quark out of an atom at low energies. But at high temperatures, things are very different—the strong force becomes weak, and we expect to see a

quark-gluon 'soup.' Experimentalists believe they have seen signs of this soup in collisions of heavy nuclei at accelerators such as the Relativistic Heavy Ion Collider at Brookhaven National Laboratory."

To complement what is learned in experiments, the physicists are using supercomputer simulations of QCD, a quantum field theory. One way of formulating this is through what are known as Feynman path integrals, which are a sum over all possible classical paths of the system, each path weighted by a prescription determined by physicist Richard Feynman nearly 60 years ago.

"Because of the uncertainty principle, the vacuum, or lowest energy state of QCD is not empty. Rather, it's a roiling sea of quarks and gluons being created and destroyed," Sugar said. Monte Carlo simulations select configurations according to the probability with which they occur in the path integral, enabling the physicists to accurately evaluate the sum over paths from only a few hundred of them.

Even then, evaluating the Feynman path integrals is no simple task. A 4-D lattice encompasses the three continuous dimensions of space and a fourth dimension, that of time. Since computers cannot handle an infinite continuum, a lattice is imposed on space and time to make the computational problem finite.

BIG CHALLENGES FROM TINY SCALES

To progress in their research, the physicists' strive to use ever-finer lattice spacing. "The purpose of moving to smaller lattice spacing is to make the calculations more accurate," Sugar explained. "If we average over smaller and smaller blocks, we come closer and closer to the continuum limit." This helps the researchers more realistically capture the physics taking place.

PROJECT PARTICIPANTS

CHRISTOPHER AUBIN AND CLAUDE BERNARD, WASHINGTON U., TOMMY BURCH, U. REGENSBURG, CARLETON DE TAR, U. UTAH, STEVEN GOTTLIEB, INDIANA U., ERIC GREGORY, U. ARIZONA, URS HELLER, FLORIDA STATE U., JAMES HETRICK, U. OF THE PACIFIC, JAMES OSBORN, U. UTAH, ROBERT SUGAR, UCSB, AND DOUGLAS TOUSSAINT, U. ARIZONA.

The MILC collaboration has developed an improved lattice discretization that significantly reduces “lattice artifacts,” which they have been running on SDSC’s Blue Gene, increasing the accuracy of the simulations and providing more useful predictions.

“The importance of the MILC collaboration’s fundamental physics research is indicated by the size of their computing allocations, which total millions of hours across the major NSF supercomputer centers, including one million hours on Blue Gene,” said Amitava Majumdar, group leader for the Scientific Computing Applications Group at SDSC.

One million processor hours on Blue Gene equals almost three weeks of continual use. In practice, users don’t tie up the entire machine for one continuous run, and Toussaint typically ran on 1,024 processors, half of Blue Gene’s resources, for 12 hours at a time, then checked the results and later ran another block.

SDSC’s remarkable new Blue Gene fits in a single rack with a footprint of only one square meter, housing 1,024 compute nodes and 128 input/output nodes. Each node consists of two IBM PowerPC 440 processors running at 700 MHz and sharing 512 megabytes of memory. This gives Blue Gene a total of 2,048 compute processors and aggregate memory of 512 gigabytes, with a peak speed of 5.7 teraflops (trillion floating-point operations per second). “It’s a remarkable packaging job to fit more than 2,000 processors in a single rack,” said Wayne Pfeiffer, a distinguished scientist at SDSC. “It’s a very compact, power-efficient machine.”

TRACKING THE ELUSIVE QUARK

The bulk of the computer resources used by the physicists in the MILC collaboration go into generating gauge configurations, or sample paths, in the Feynman path integrals. “We integrate over a set of gauge configurations and store the results on disk, and then average over the samples,” Toussaint said.

The researchers are using Blue Gene to expand their collection of gauge configurations, and each of several hundred paths spanning just 3.5 fermis (or femtometers, a million-billionth of a

meter) produces about 4.3 gigabytes of data. The SDSC Storage Resource Broker (SRB)—part of the Center’s end-to-end data cyberinfrastructure—helps the researchers preserve and access their collection of configurations, maintaining this data on call as a valuable tool used by MILC researchers and many other lattice gauge theorists to study a wide range of physical phenomena.

For example, the MILC group, in collaboration with physicists at the Fermi National Accelerator Laboratory, recently applied their configurations to study the decays of unstable D mesons. The predictions were subsequently confirmed by experiments at Fermilab and Cornell University. These results are important because the same methods can be used to calculate decay properties of the B meson, which are very difficult to determine experimentally and are critical in efforts to make precise tests of the Standard Model in the search for new physics that may lie beyond it.

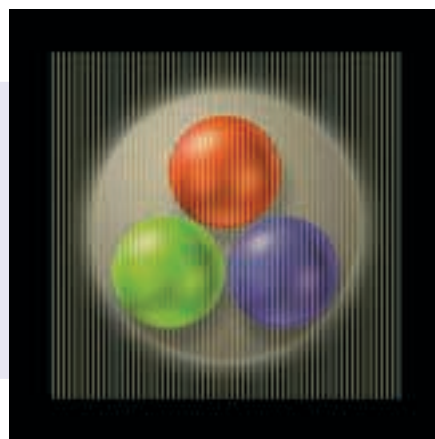
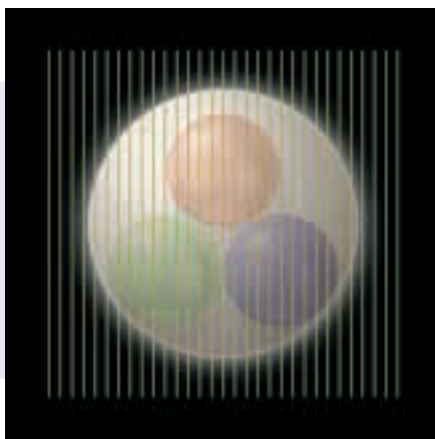
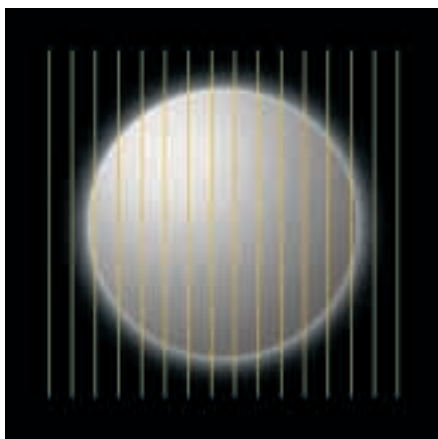
The MILC researchers and other groups have also used the configuration collection to determine such fundamental quantities as the constant that determines the strength of the strong interactions, the masses of quarks, the decay properties of pi and K mesons, and the masses of a host of strongly interacting particles.

“What makes this exciting to members of our field is that the objective is very ambitious,” Sugar said. “We are making progress in understanding the fundamental forces of nature and the fundamental building blocks of nature.”

- Dave Dooling is a freelance science writer living in New Mexico.

RELATED LINK

MIMD Lattice Computation (MILC) Collaboration
<http://physics.indiana.edu/~sg/milc.html>



To better understand Quantum Chromodynamics theory and make predictions to complement experiments, physicists in the MILC collaboration use progressively smaller lattice spacing, shown in this artist’s representation, in their simulations on SDSC’s Blue Gene supercomputer. These demanding simulations reveal greater detail (right) of the quarks and related particles normally hidden inside a proton (left). Douglas Toussaint, U. Arizona.

SDSC CONNECTS FAMILIES WITH LOVED ONES IN IRAQ

This spring, SDSC began hosting video teleconferencing sessions connecting San Diego military families with loved ones serving their country in Iraq. Utilizing existing satellite uplink technology and staffed by SDSC volunteers, the Center has opened its doors each Saturday, reaching to date more than 100 families,



SDSC helps connect families with loved ones in Iraq. Ashley Wood.

giving them the opportunity to spend about 30 minutes virtually face-to-face with distant loved ones.

The program is a collaboration between SDSC and the nonprofit Freedom Calls Foundation, which provides the satellite uplinks at locations in Iraq.

"It's great to live at a time when technology makes it possible for people to be together, even though they're half a world apart," explains Dick Bartlett, volunteer and VTC coordinator. "Video teleconferencing isn't a hug and a kiss, but it's the next best thing. If you want proof of its powerful impact, just watch the faces of the children when they see their parent who is in Iraq on the television screen—it melts your heart."

"Arranging and coordinating VTCs is probably the most satisfying thing I've ever

been involved in and I'm very thankful to have SDSC's assistance on this program," Bartlett continues. "For people looking for a way to support the troops, I can't think of anything more meaningful than to volunteer to assist at one of the VTCs." For more information, see www.freedomcalls.org.

SDSC'S PHIL PAPADOPOULOS NAMED TO NATIONAL NETWORKING RESEARCH COUNCIL

Phil Papadopoulos, SDSC's Advanced Cyberinfrastructure Laboratory (ACiL) Director, has joined a team of leading researchers and technologists as part of the National Networking Research Council (NNRC) of National LambdaRail (NLR). The Council is a distinguished national group tasked with guiding and directing NLR's support of networking research.

National LambdaRail is a consortium of leading U.S. research universities and private sector technology companies that is deploying a nationwide optical, Ethernet, and IP networking infrastructure.

NLR provides researchers unprecedented control over a nationwide network infrastructure

with up to 40 individual lightpaths, each of which can transmit data at 10 gigabits per second and be used to deploy dedicated side-by-side, but physically and operationally separate, production and experimental networks.

SDSC COLLABORATES WITH STATE OF CALIFORNIA ON DATA PRESERVATION

In a first-of-its-kind collaboration, SDSC and the State of California have been awarded funding by the National Historical Publications and Records Commission (NHPRC), a

branch of the National Archives and Records Administration (NARA), to develop a preservation cyberinfrastructure for California Geospatial records. The visionary project will deploy persistent archival infrastructure across State agencies and create a working prototype.

Led by Richard Marciano, director of SDSC's Sustainable Archives and Library Technologies (SALT) lab, and Reagan Moore, director of the Data-Intensive Computing Environments (DICE) group at SDSC, the funding will support the deployment of a data grid infrastructure at the California State Archives, the California Resources Agency's California Environmental Resources Evaluation Systems (CERES) office, and SDSC.

PATENT AWARDED FOR PERSISTENT ARCHIVE SYSTEM

Six SDSC inventors were recently awarded a patent for a "Persistent Archives" software system that provides a method for ingesting and validating a knowledge-based archive of data objects. The team included Reagan Moore, Arcot Rajasekar, Chaitan Baru, Bertram Ludaescher (now at UC Davis), Amarnath Gupta, and Richard Marciano. The newly patented system enables the building of a preservation environment that can take advantage of new generations of technology, making possible the sustainability of data collections over time. Managing long-lived data is fundamental to advancement in the research and education community, as described in the compelling National Science Board report, "Long Lived Digital Data Collections: Enabling Research and Education in the 21st Century."

INCA SET TO LAUNCH VERSION 2.0

Inca, the flexible framework for the automated testing, benchmarking, and monitoring of Grid systems, will introduce Inca 2.0 in mid-summer 2006. The new version will offer improved archiving, installation and configuration control, as



SDSC is collaborating with the State of California on digital preservation research funded by the National Archives and Records Administration through the NHPRC. NARA.

well as security, with new and enhanced features.

Led by SDSC's Shava Smallen and originally developed for the TeraGrid project, Inca is a general framework that can be adapted and used by other Grids. A pre-release version of Inca 2.0 was introduced in February 2006 and currently has nine active users and seven grid projects around the world including TeraGrid, DEISA (Distributed European Infrastructure for Supercomputing Applications), Cineca (Italy), LRZ (Germany), NGS and WRG (UK National Grid Services), as well as APAC (Australian Partnership for Advanced Computing).

SDSC AND NCSA PROVIDE COMBINED TERAGRID RESOURCES

Since April 1, 2006 users requesting high-performance computing resources from the National Science Foundation have had seamless access to all computational resources at SDSC and the National Center for Supercomputing Applications (NCSA) within the TeraGrid environment.

TeraGrid now provides the national user community access to the Tungsten, Radium, and Copper systems at NCSA and SDSC's DataStar and Blue Gene systems, bringing the TeraGrid's total computational performance to a peak of about 102 trillion calculations per second.

The move represents a large-scale integration effort by NCSA and SDSC, as well as the TeraGrid partners, to provide a national grid-enabled computational and data management facility. Integrating new hardware into TeraGrid carries with it a large-scale software challenge—integrating the software systems of the new hardware as well as extending security, accounting, scheduling, and other systems.

Another benefit is that TeraGrid users with "roaming" allocations now have access to all NSF-supported systems at NCSA and SDSC, as well as the TeraGrid systems at other Resource Provider (RP) sites.

SDSC and NCSA, two of the founding sites in the TeraGrid, join six other participating Resource Providers: the University of Chicago/Argonne National Laboratory; Indiana University; Oak Ridge National Laboratory; Pittsburgh Supercomputing Center; Purdue University, and the Texas Advanced Computing Center.



Massive damage from the 1906 San Francisco earthquake. SDSC visualization researchers created a movie of a USGS simulation of the earthquake. Ervin Scharff, Online Archive of California.

INTRODUCING NATIONAL ALLOCATIONS FOR DATA STORAGE

For the first time, the application and approval process for storing data collections larger than one terabyte will be managed by a NSF-sponsored national allocations process, pioneered through DataCentral at SDSC (see article on page 10 of this issue).

SDSC is making available for allocation tens of terabytes of disk, targeting researchers who need to store large data collections and make them widely available. Eventually, significant additional data resources at SDSC and other TeraGrid sites will be allocated through the national process using the Partnerships Online Proposal System (POPS).

The next allocation meeting is scheduled for June 2006, for allocations beginning July 1, 2006. For more information on how to apply for a data allocation, see SDSC DataCentral at datacentral.sdsc.edu.

SDSC VISUALIZES THE GREAT 1906 SAN FRANCISCO EARTHQUAKE

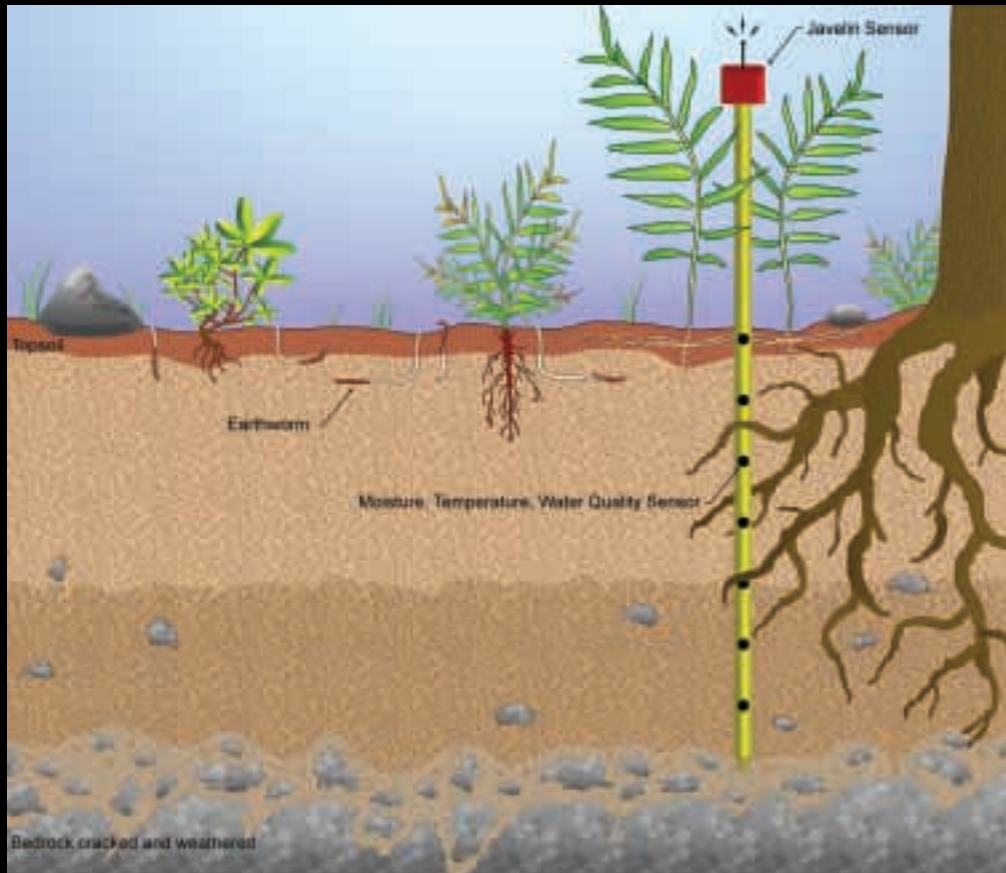
Striking before dawn on April 18, 1906, a magnitude 7.9 earthquake began off the San Francisco coast and ruptured the ground for some 300 miles along the San Andreas Fault. Creating a sustained shaking that lasted more than 40 seconds, the earthquake and resulting fires left much of the Bay Area in ruins.

To commemorate the 100th anniversary of this earthquake, SDSC visualization experts have created a movie showing in extraordinary detail what happened that day, based on the latest simulation performed by scientists at the USGS (United States Geological Survey), Menlo Park. Developed by SDSC researchers Steve Cutchin and Amit Chourasia, the movie was unveiled on April 18, 2006 at the 100th Anniversary Earthquake Conference in San Francisco.

The conference was hosted by the Earthquake Engineering Research Institute (EERI), the Seismological Society of America (SSA), and the California's Office of Emergency Services. For more information see www.1906eqconf.org.

BUILDING ECOLOGICAL CYBERINFRASTRUCTURE FOR NEON

Much of the power of the National Ecological Observatory Network (NEON) will lie in the vast volume and variety of ecosystem data generated, integrated, and delivered to ecologists and others. The San Diego Supercomputer Center (SDSC) is playing an important role in delivering advanced cyberinfrastructure for NEON, building upon experience with NSF projects such as GEON, SEEK, and NEES. The soil sensor array shown will characterize a suite of soil biological and chemical parameters that contribute to the biogeochemical cycling of nutrients and plant growth, including growth and respiration of roots and microbes, soil chemistry, temperature, and water content. Jason C. Fisher.



SUBSCRIPTIONS

For a free subscription to *ENVISION*, send the information requested below to:

Gretchen Rauen, SDSC
 University of California, San Diego
 9500 Gilman Drive, MC 0505
 La Jolla, CA 92093-0505
 eradmin@sdsc.edu, 858-534-5111
 www.sdsc.edu

NAME

TITLE

INSTITUTION

ADDRESS

CITY, STATE, AND ZIP

COUNTRY

E-MAIL



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

ADDRESS SERVICE REQUESTED

PRESORTED STANDARD
 U.S. POSTAGE
PAID
 WESTERN GRAPHICS