CRD

Computational Research Division Report

FEBRUARY 2008

Star Management

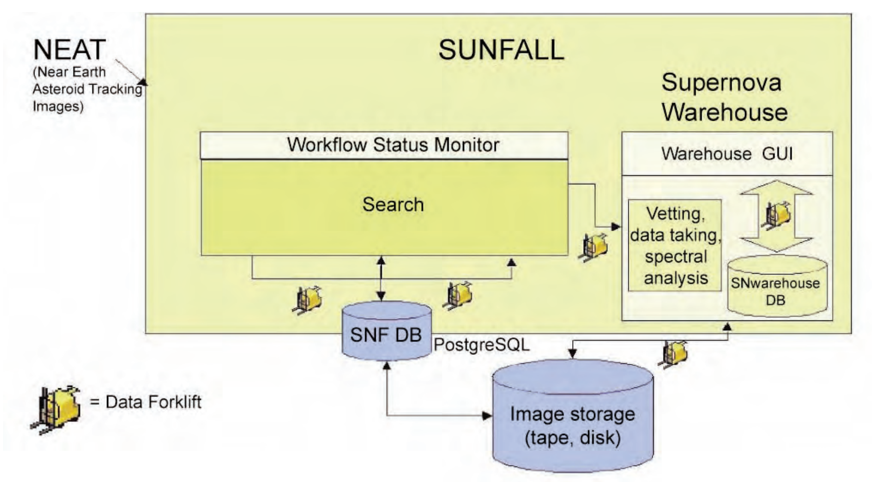
A system to collect and manage a treasure trove of supernova data is winning kudos

More than two years ago, a stellar idea brought together computer scientists and astrophysicists and resulted in an award-winning, interactive visual analysis system for supernova research.

The system, called Sunfall, enables scientists from the Nearby Supernova Factory (SNfactory) to process a large amount of observational data and separate the desired supernovae from other bright objects in the sky. It also offers automated data transfer tools and software that allow the SNfactory scientists — who are located in the United States and France — to easily retrieve, annotate and modify the supernova data for further research.

Headed by Cecilia Aragon in CRD's Visualization Group, Sunfall impressed the judges at the IEEE Symposium on Visual Analytics Science and Technology and won the Best Poster Award last fall.

continued on page 3



Sunfall architecture diagram, depicting the four components (Search, Workflow Status Monitor, Data Forklift, and Supernova Warehouse) and data flow between the components

Bugs for Science

DOE JGI updates its metagenome data management and analysis system

Targeting its ever-expanding user community, the U.S. Department of Energy Joint Genome Institute (DOE JGI) has released an upgraded version of the IMG/M metagenome data management and analysis system, accessible to the public.

IMG/M, developed by the DOE JGI and Biological Data Management and Technology Center (BDMTC) in CRD, provides tools for analyzing the functional capability of microbial communities based on their metagenome DNA sequence in the context of reference isolate genomes. The new version of IMG/M includes five

continued on page 4

A Measure of Success

A network performance and cybersecurity specialist wins $35,000 award

Vern Paxson, a network researcher in CRD's Advanced Computing for Science Department, has been awarded the Association for Computing Machinery's Grace Murray Hopper Award for his work in measuring and characterizing the Internet.

"His innovative techniques are used to assess new communications concepts, improve network performance and prevent network intrusion," according to the ACM. "They provide both the research community and Internet operators with the tools to improve the operation of this increasingly diverse, decentralized com-

continued on page 4

Fiery Prelude

Researchers study thermonuclear flames to understand what could cause a supernova

Intrigued by the process that causes an aging star to explode and release intense energy, Andy Aspden has taken on research on the thermonuclear flames that lead to the explosion of Type Ia supernovae. A year after joining Berkeley Lab as a Glenn T. Seaborg postdoctoral fellow, Aspden is performing high-resolution simulations that reveal unprecedented details of the physics that govern how flames propagate through a white dwarf star.

Aspden, who received a Ph.D. in applied mathematics from the University of Cambridge in the United Kingdom,

continued on page 5

Badge of Honor

Mathematician James Sethian is elected to National Academy of Engineering

James Sethian, head of the CRD's Mathematics Group and a professor of mathematics at UC Berkeley, has been elected to the National Academy of Engineering.

Sethian was one of 65 new members and nine foreign associates whose election was announced on Feb. 8. Sethian was honored "for the development of efficient methods of tracking moving interfaces."

Sethian's research has led to the development of level set methods, which are numerical techniques that can follow the evolution of interfaces, as well as a host of other techniques to track interfaces in various settings. These interfaces can develop sharp corners, break apart and merge together. The techniques have a wide range of applications, including problems in fluid mechanics, combustion, manufacturing of computer chips, computer animation, image processing, structure of snowflakes and the shape of soap bubbles.

"Prof. Sethian is truly deserving of this honor, which reflects his mathematical contributions to solving problems across a wide range of areas," said Berkeley Lab Associate Laboratory Director for Computing Sciences Horst Simon. "Applied math research is an essential component

James Sethian

of our computational science program, and the research by James and his group has contributed to advancements not only at Berkeley Lab but at research organizations across the country and around the world."

Academy membership honors those who have made outstanding contributions to "engineering research, practice, or education, including, where appropriate, significant contributions to the engineering literature," and to the "pioneering of new and developing fields of technology, making major advancements in traditional fields of engineering, or developing/implementing innovative approaches to engineering education."

Sethian received his Ph.D. in applied mathematics from UC Berkeley in 1982, with a dissertation analyzing theoretical and numerical aspects of the numerical approximation of moving curves and surfaces in combustion modeling. He followed with a National Science Foundation postdoctoral fellowship at the Courant Institute of Mathematics, and then returned to UC Berkeley as an assistant professor in 1985. He is currently vice chair for Undergraduate Affairs and professor in the Department of Mathematics.

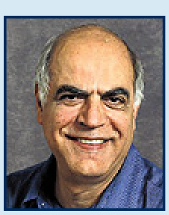
Sethian is the author of many scientific articles and books, and serves as an editor on several journals. He is the recipient of numerous prizes and awards, including the 2004 Norbert Wiener Prize in Applied Mathematics, which is awarded every three years jointly by the American Mathematical Society (AMS) and the Society for Industrial and Applied Mathematics (SIAM), for his work on level set methods and other applications.

Learn more about James Sethian's work.

Hall of Fame

Database Encyclopedia

Arie Shoshani, head of SciDAC's Scientific Data Management Center and leader of CRD's Scientific Data Management Group, has written two essay-type entries for the Encyclopedia of Database Systems, to be published later this year by Springer Verlag.



Arie Shoshani

According to the publisher, "The Encyclopedia of Database Systems is designed to meet the needs of research scientists, professors and graduate-level students in computer science and engineering. Topics for the encyclopedia were selected by a distinguished international advisory board, and written by world class experts in the field."

Based on his work in the field, Shoshani was invited to write entries on "Logical Models of Temporal Data" and "Summarizability Properties of Statistical Databases."

Pushing the Boundary

David Bailey, CRD's chief technologist, gave a Richard B. Pelz Memorial Lecture at Rutgers University in New Jersey that highlighted the research into boosting the ability of high-performance computers to improve their numeric calculations.

David Bailey

continued on page 6

Sunfall *continued from page 1*

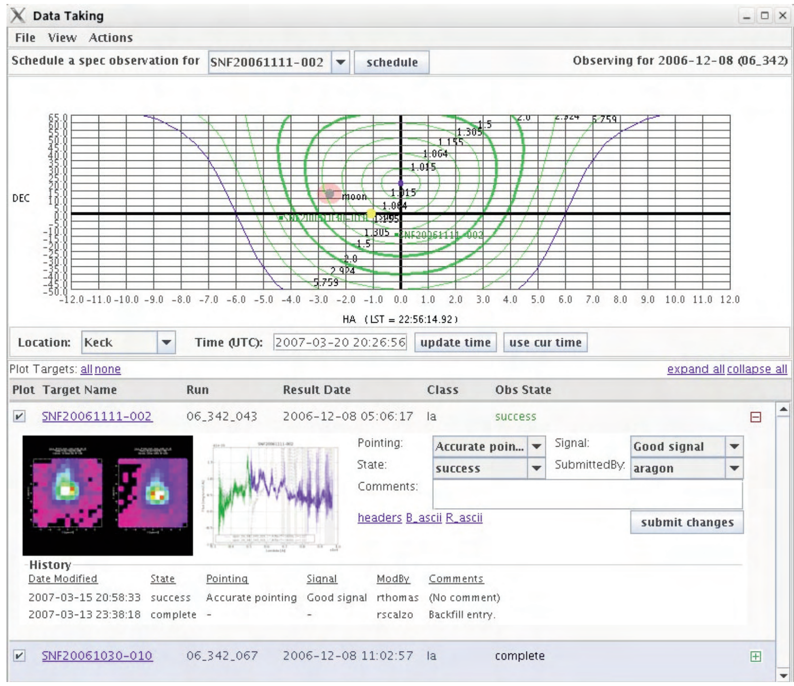
"The project was a success because it was done by an interdisciplinary team of scientists," said Aragon. "As we move into petascale computing and increasingly massive amounts of scientific data, the interdisciplinary approach is the way to solve many problems."

The SNfactory is an international collaboration among Berkeley Lab, Yale University and three research centers in France: Centre de Recherche Astrophysique de Lyon, Institute de Physique Nucleaire de Lyon and Laboratoire de Physique Nucléaire et de Hautes Energies. SNfactory's mission is to create a large database of Type Ia supernovae, which are known for their extraordinary and uniform brightness and their role in a breakthrough research showing that the universe's expansion is accelerating. The data would help researchers to study the mysterious dark energy that propels the expansion.

The SNfactory scientists approached Aragon in 2005 to design Sunfall, which is short for SuperNova Factory AssembLy Line. The first goal was to tackle the growing amount of wide-field image data from the Palomar Observatory in San Diego. SNfactory receives 50 to 80 gigabytes of image data each night, and its researchers must process and examine these data within 12–24 hours in order to get the most from these rare stellar explosions. These supernovae only occur a couple of times in a typical galaxy per millennium, and remain bright enough for detection only for a few weeks.

"To maximize the scientific return, this has to be a very accurate, efficient, and traceable process. The data come in seven days a week for a period of seven months, making the operations very dynamic, and there is no way to take a break to catch up," said Greg Aldering, leader of SNfactory and a Berkeley Lab researcher in the Physics Division. "Before the Sunfall software was developed to integrate this process, it could be an overwhelming job simply to perform the necessary work, but in addition it also was difficult to keep track of whether what had been done was sufficient, much less optimal."

Processing and analyzing these data became a great challenge for the researchers, whose responsibilities require them to sift through vast amounts of image data to search for Type Ia supernovae, and then follow up with spectral and photometric observations of each supernova. Aragon and fellow Sunfall researchers created the system by modifying existing software and developing new algorithms.



SNwarehouse Data Taking window. The observer can follow the targets on the Sky visualization, take notes on the success or failure of each observation, telescope status and weather conditions, and reschedule targets if necessary.

Sunfall provides the software tools for extracting the Type Ia supernovae from the raw images, and statistical algorithms that reduced the number of false-positive supernovae candidates by a factor of ten. Instead of reviewing 1,000 selected images each day, the researchers only have to pore over data from approximately 100 images.

The software provides a visual display of three-dimensional astronomical data in an easy-to-read, two-dimensional format. Another visual display offers the signal strength and other information about each spectrum, making it easy for researchers to analyze the information.

Sunfall also closely monitors and detects any problems that crop up while NERSC supercomputers process the data — there are visual displays of job queues, completion times and other information.

A Sunfall package called Data Forklift then automates data transfers from among different types of systems, databases and formats. Having a reliable and secure data transfer mechanism is critical given that SNfactory researchers are located in different countries and time zones.

By using Sunfall's Supernova Warehouse (SNwarehouse) tool, scientists can easily access, modify, annotate and schedule follow-up observations of the data.

"The new capabilities of SNwarehouse produced an immediate transformation, allowing us to shift our focus onto the science analysis of our follow-up spectroscopy," Aldering said.

Developing Sunfall was no easy task. Aragon and other computer scientists and astrophysicists on the interdisciplinary team met frequently, often daily, to discuss various proposed designs and implementations, report progress, and ask for feedback on the developing system. Ideas for technical solutions often came during regular weekly science meetings in which SNfactory researchers discussed their work.

In addition to Aragon, other Sunfall project members included Stephen J. Bailey, formerly in Berkeley Lab's Physics Division, Sarah Poon and Karl Runge of the Space Sciences Lab at UC Berkeley, and Rollin Thomas, who worked for SNfactory and recently joined CRD's Computational Cosmology Center.

SNfactory researchers have been using Sunfall since 2006.

"It was a really cool project that showed that an interdisciplinary team can have success solving high-data-volume science needs," Aragon said. "Now we have experience in solving a practical problem that can be applied in other scientific domains."

Grace Hopper Award continued from page 1

munications infrastructure.

The award is presented annually to the outstanding young computer professional of the year, selected on the basis of a single recent major technical or service contribution. The candidate must have been 35 years of age or less at the time the qualifying contribution was made. The Grace Murray Hopper Award includes a $35,000 prize, with financial support provided by Google.

Paxson is also affiliated with the International Computer Science Institute and recently joined the faculty at UC Berkeley. The award recognizes his work in the mid-1990s, primarily his Ph.D. dissertation, which "[laid] the groundwork for the exciting resurgence in research in Internet measurement during the last ten years," according to the award citation.

"The fact that so many people today take the speed and the reliability of the Internet for granted is due in large part to Dr. Paxson's early career research into accurately measuring just how traffic moves from point to point on this network of networks," said Michael Strayer, Associate Director of the Advanced Scientific Computing Research in DOE's Office of Science. "This prestigious award is fitting recognition of Dr. Paxson's key role in advancing our understanding of the Internet, which is increasingly critical to our global scientific and economic competitiveness."

"Paxson's Ph.D. thesis — a tome at over 400 pages — is one of the only dissertations that was (and still is) widely read by other researchers in the networking community," according to the ACM. "His early measurement papers are still widely cited, and his style of research adopted as a gold standard for how to measure a complex, heterogeneous network like the Internet and make statistically sound statements about its properties and their implications. The vibrancy of the Internet Measurement Conference (which he co-created in 2001) and the prevalence of measurement papers at other networking conferences are a testament to the influence of his initial research in this area."

The Hopper Award continues a series of honors given to Paxson. In 2007, he was named a fellow of the ACM. The fellowship is given to scientists who have made "outstanding technical and professional achievements in the field of information technology." In 2006, Paxson was awarded the first-ever "test of time" award by ACM's Special Interest Group on Data Communications (SIGCOMM) for his 1996 paper "End-to-end routing behavior in the Internet," citing the paper as the most influential networking paper written 10 years ago.


Vern Paxson

ACM also noted, "Through a series of highly influential papers, Paxson's findings revealed the mismatches between reality and the common assumptions made in analytical and simulation models. By combining the extensive collection of data from many locations with sophisticated statistical techniques, he provided a wealth of useful information about the nature of the Internet and ways to improve its operation."

Paxson is also well known for his leadership in the development of Bro, an intrusion detection system for monitoring and tracing network attacks. He received his M.S. and Ph.D. degrees from UC Berkeley.

Metagenome Data continued from page 1

additional metagenome datasets generated from microbial community samples that were the subject of recently published studies. These include the metagenomic and functional analysis of termite hindgut microbiota (*Nature* 450, 560–565, 22 November 2007) and the single-cell genetic analysis of TM7, a rare and uncultivated microbe from the human mouth (*PNAS*, vol. 104, no. 29, 11889–11894, July 17, 2007).

"IMG/M is a fantastic tool that is incredibly helpful in understanding our data," said Stephen Quake, Co-Chair, Department of Bioengineering at Stanford University, Investigator, Howard Hughes Medical Institute, and senior author on the PNAS study. "We used IMG/M in numerous ways, both to analyze our data and to understand general properties of other relevant bacterial genomes. I look forward to analyzing our new datasets with IMG/M."

IMG/M will be demonstrated at a workshop on March 26 as part of the DOE JGI Third Annual User Meeting. IMG/M contains all isolate genomes in version 2.4 of DOE JGI's Integrated Microbial Genomes (IMG) system, which represents an increase of 1,339 reference genomes from the previous version of IMG/M. Now, IMG/M contains 2,953 isolate genomes consisting of 819 bacterial, 50 archaeal, 40 eukaryotic genomes, and 2,044 viruses.

IMG/M provides new tools for analyzing metagenome datasets in the context of reference isolate genomes, such as the Reference Genome Context Viewer and Protein Recruitment Plot, that allow the examination of metagenomes in the context of individual reference isolate genomes. New Abundance Comparison and Functional Category Comparison tools enable pairwise function analysis (COG, Pfam, Enzyme, TIGRfam) and functional category (e.g., COG category) abundance comparisons, respectively, between a metagenome dataset and one or several reference metagenomes or genomes, and test whether the differences in abundance are statistically significant.

IMG/M has been developed jointly by the DOE JGI's Genome Biology Program (GBP) and BDMTC. The large-scale pairwise gene similarity computations for all the genomes included in IMG/M have been carried out using ScalaBLAST by the Computational Biology and Bioinformatics Group of the Computational Sciences and Mathematics Division at Pacific Northwest National Laboratory.

Story by David Gilbert at the JGI.

Supernova Flames

continued from page 1

has worked closely with John Bell, head of the CRD's Center for Computational Sciences and Engineering (CCSE), and Stan Woosley of UC Santa Cruz, head of SciDAC's Computational Astrophysics Consortium.

Exactly how supernovae explode is a great mystery; however, scientists are working to construct increasingly precise models to determine the likely mechanisms that lead to the runaway thermonuclear reactions, thanks to more powerful computers and better algorithms.

These models have demonstrated that the reaction process isn't uniform. Factors such as fuel density and turbulence levels affect how the star burns before the final explosion that tears it apart.

"To construct models that will permit the study of large-scale supernovae, we need to understand the complex dynamics of small-scale turbulent thermonuclear flames," Aspden said.

The Type Ia supernovae, which are the focus of Aspden's work, are known for their consistent and exceptional brightness. Cosmologists use Type Ia supernovae as "standard candles" because they can use a Type Ia's peak luminosity and the duration of its emission to calculate the supernova's distances from Earth. As a result of this type of research, scientists overturned conventional wisdom in 1998 when they announced that the rate of the universe's expansion is increasing, not decreasing. They made the discovery after determining that the supernovae they had observed were moving away from Earth at a fast clip.

But what leads to the boom in the first place? Scientists have believed the process begins when a white dwarf star accrues enough mass from a nearby star to raise the temperature and pressure, triggering carbon fusion in its core. The ignition creates thermonuclear flames, which burn outward until reaching the periphery of the white dwarf star.

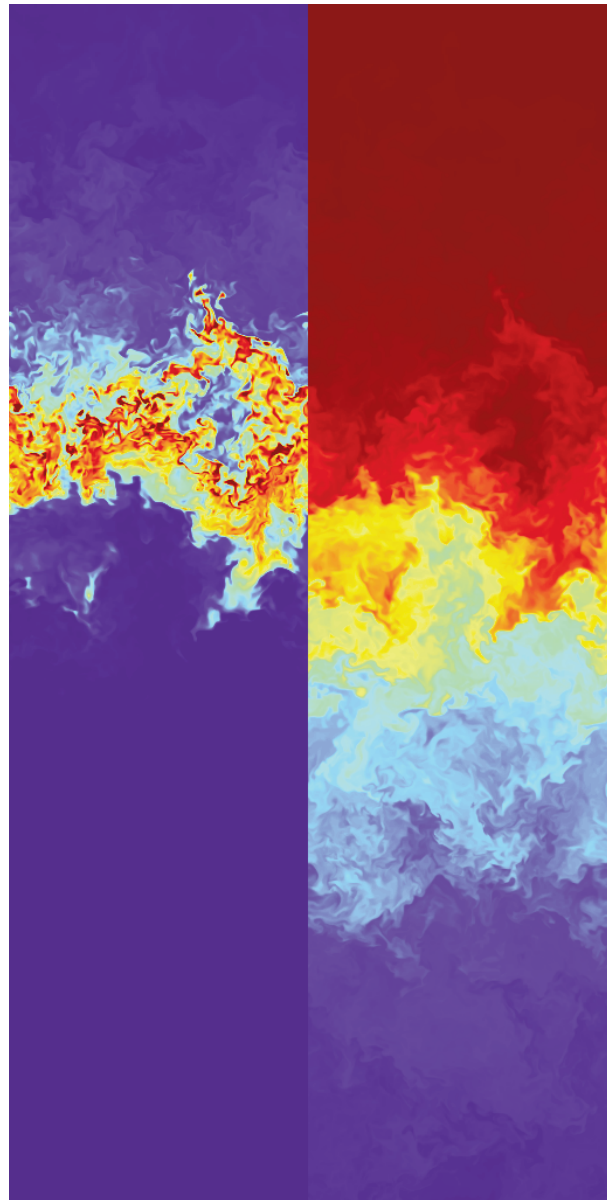
"The physics of the flames, characterized by different burning regimes, presents an interesting and challenging problem," Aspden said. Near the center of the star, the burning is extremely intense and the background turbulent flow within the star has a minimal effect on the flame. As the fuel density decreases, the impact of turbulence on the flame increases, leading to a qualitative shift in behavior known as distributed burning. Scientists have speculated that an optimal condition for the eventual detonation of the star occurs when the flames enter the distributed burning regime.

To understand the details of how these burning regimes work, Aspden has been able to perform simulations that range from 13 microns to 6 millimeters in resolution, in domains up to twelve meters across. These high-resolution models allow Aspden to home in on specific issues and carry out more detailed calculations.

"We are doing data analysis of the interactions of the flames with turbulence to understand how to accurately model the burning in larger scale calculations," Aspden said.

Aspden's work builds on research by Bell and his colleagues, who in 2004 developed a new numerical approach for simulating nuclear flames that makes Aspden's studies possible. CCSE continues to work closely with Woosley's SciDAC Computational Astrophysics Consortium, developing new computational approaches for modeling supernovae. SciDAC, Scientific Discovery through Advanced Computing, provides DOE funding to interdisciplinary research teams that develop sophisticated software and algorithms. These tools enable researchers to make best use of supercomputers for solving complex problems.

"What we are trying to do is to create a model grounded in physics," Bell said.



The two panels show the same slice through a downward-propagating flame in the distributed burning regime. The left-hand panel is the burning rate, and the right-hand panel is the temperature. High turbulence levels stir the cold fuel with the hot products to produce a broad mixed region. The burning occurs at the hot end of the mixed region and is an order of magnitude wider than the corresponding laminar flame.

"These flames don't behave the way terrestrial flames do at all, so you can't study them with terrestrial experiments."

Hall of Fame *continued from page 2*

Bailey's talk in January included a discussion about a paper he co-authored with Pelz on using the 64-digit arithmetic for investigating vortex roll-up. Pelz, a professor of mechanical and aerospace engineering at Rutgers, specialized in computational fluid dynamics. Pelz received honorable mention during the 1998 Gordon Bell Prize competition, which recognizes achievements in high-performance computing. He was born in 1957 and died in 2002.

Although most of the current scientific calculations are carried out in the 16-digit (64-bit) IEEE arithmetic, a number of applications have required calculations with more precision, Bailey said. Aside from vortex roll-up, applications that demand high-precision computation include supernova simulations, planetary dynamics and electromagnetic scattering.

MLK Symposium

Juan Meza

Juan Meza, head of the High Performance Computing Department in CRD, spoke at the Marjorie Lee Browne colloquium at the University of Michigan in January, taking part in a Martin Luther King Symposium.

Meza, in a talk titled "I Want to Be a (Computational) Mathematician," gave his personal account of how he became interested in mathematics as a college undergrad and the research projects he has undertaken since then. The title is a reference to a book by Paul Halmos about his life as a mathematician and the history of the discipline from the 1930s to 1980s.

The colloquium honored Browne, the first African-American woman to earn a Ph.D. in mathematics from the University of Michigan, in 1950.

In his talk, Meza spoke about the rise of the field of computational mathematics and its contribution to science. He gave examples of his recent work on developing new algorithms and using them in massively parallel computers to solve challenging problems.

New Staff

Reese Baird and Ted Kisner recently joined the Scientific Computing Group (SCG), in which they will work with Julian Borrill on researching cosmic microwave background (CMB) radiation.

Baird came to SCG after a two-year tour of duty with the Peace Corps in Cameroon. Previously he was a member of the High Performance Computing Group at Los Alamos National Laboratory. He holds a bachelor's degree in computer science from the University of Texas at Austin.

Reese Baird

Ted Kisner

Kisner worked at the Space Sciences Laboratory at UC Berkeley prior to joining Berkeley Lab. At the Space Lab, Ted already worked very closely with members of SCG on CMB. His past work included the analysis of data from the BOOMERanG experiment. His research currently is focused on the Planck and EBEX experiments. Kisner is working on improving CMB data analysis by developing and implementing new algorithms, as well as creating tools for managing CMB data and the software for analyzing them. He is getting a Ph.D. in physics from UC Santa Barbara.

Reese and Kisner will be part of the newly formed Computational Cosmology Center, which is led by Borrill and Peter Nugent.

About CRD Report

CRD Report, which publishes every other month, highlights the cutting-edge research conducted by staff scientists in areas including turbulent combustion, nanomaterials, climate change, distributed computing, high-speed networks, astrophysics, biological data management and visualization. CRD Report Editor Ucilia Wang can be reached at 510 495-2402 or **Uwang@lbl.gov**. Find previous CRD Report articles at http://crd.lbl.gov/html/news/CRDreport.html.