

NPACI & SDSC

QUARTERLY SCIENCE MAGAZINE

ENVISION

VOL. 18 NO. 1

JANUARY-MARCH 2002



**DESIGNING PERSONALIZED
HEARING DEVICES**

COMPUTATIONAL MEDICINE

Supercomputers Fighting Cancer • Analyzing Brain Tissue

NPACI

Building the Computational Infrastructure for Tomorrow's Scientific Discovery

The National Partnership for Advanced Computational Infrastructure (NPACI) joins 47 partner institutions in 21 states, Australia, Italy, Spain, and Sweden, in creating the foundation for a ubiquitous, continuous, and pervasive computational environment to support research by the world's scientists. NPACI is led by UC San Diego, funded primarily through the NSF's Partnerships for Advanced Computational Infrastructure program, and has its focus of activities at the San Diego Supercomputer Center (SDSC).

NPACI INFORMATION

Fran Berman, Director

David Hart, Communications Director
dhart@sdsc.edu
858-534-8314

SDSC

SDSC is a campus research unit of UC San Diego and the focus of activities for NPACI. Since 1985, SDSC has served the U.S. scientific community as a national laboratory for computational science and engineering.

GENERAL INFORMATION

Phone: 858-534-5000

info@npaci.edu

www.npaci.edu

www.sdsc.edu

Subscribe to *enVISION* and *Online* at:
www.npaci.edu/Press/subscriptions.html

ONLINE

www.npaci.edu/Online

Online is a biweekly, Web newsletter of the latest research and developments from NPACI and SDSC.

enVISION

www.npaci.edu/enVision

enVISION, ISSN 1521-5334, is published quarterly by NPACI and SDSC External Relations. For a free subscription or to make address changes, visit the website or contact the editor.

EDITOR: Rex Graham, rgraham@sdsc.edu,
858-822-5408

DESIGNER: Gail Bamber,
bamberg@sdsc.edu, 858-534-5150

WRITERS: Cassie Ferguson (CF), Mike Gannis (MG), Rex Graham (RG), David Hart (DH), Merry Maisel (MM), Paul Tooby (PT)

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, UC San Diego, or the NPACI partner institutions. All brand names and product names are trademarks or registered trademarks of their respective holders.

PRINTED ON RECYCLED PAPER

Contents

FROM THE DIRECTOR

- 1 Engineering NPACI Software for Usability

COMPUTATIONAL MEDICINE



p. 2

- 2 Designing Personalized Hearing Devices
- 6 Developing Therapy for Non-Hodgkin's Lymphoma with Supercomputer Simulations

- 8 Determining Protein Structures with High-Throughput Techniques

- 10 Dickerson's Formula: Biochemistry's Equivalent to Moore's Law

- 11 Biochemical Modeling Helps Explain Complex Neural Junction in Chicken Embryo

ENERGY AND THE ENVIRONMENT

- 14 Exploiting Aging Oil Fields with Advanced Computational Tools



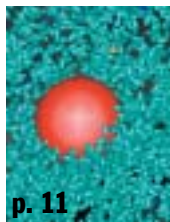
p.14

EDUCATION

- 17 From Proteins to Galaxies: NPACI's Growing Outreach to Museums

NEWS

- 20 Strategic Advisors Selected
- 20 Richard Moore Appointed NPACI Executive Director
- 20 Ellisman's Group Wins Award at SC2001



p. 11

- 20 12 Companies Adopt Globus Toolkit as Standard Grid Technology Platform
- 20 PDB Scientists Release Alpha Version of OpenMMS Toolkit

- 21 Girls Gather for SDSC-Sponsored Science Night

- 21 Taylor Moving to University of Warwick

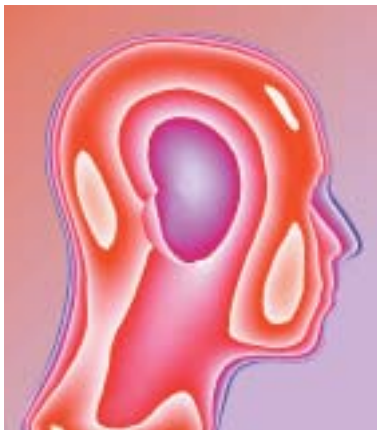
- 21 SDSC Chooses SAN Solution from Sun

- 21 NPACI Portal Adds Data Technology

- 21 Calendar

BACK COVER:

Clarifying Fluid Behavior in Pores



CHANDRAJIT SAJJU

FRONT COVER: DESIGNING PERSONALIZED HEARING DEVICES

The path that acoustic waves follow around and through the human head subtly alters them in a way that allows people to subconsciously locate the source of a sound. One of the drawbacks of hearing aids is that although they can amplify sounds, they can't provide the perceptual cues required for sound localization. The first step in filling in the missing piece of hearing for the millions of people who wear hearing devices is to develop a thorough understanding of how sound interacts with the head, a project pursued by NPACI researchers. One computational approach to plotting sound around the head essentially involves draping it with a digital mesh consisting of fixed points, then aiming acoustic waves at those points and recording the flow. The images on the front cover depict simulations of acoustic pressure around the head at a variety of frequencies.

Engineering NPACI Software for Usability

FROM THE DIRECTOR

Consider VCR clocks. For years, consumers complained about the blinking 12:00...12:00...12:00s while manufacturers insisted that they had provided instructions for setting the digital timepieces. Many consumers found the instructions confusing and didn't bother setting their VCR clocks. It turned out that there was a reasonable alternative: manufacturers could use time signals from TV channels to set the VCR clocks automatically. Once they began doing that, VCR clocks kept accurate time more frequently.

The evolution of VCR design to provide more effective and accessible time-keeping provides a lesson to developers of consumer appliances and complex software. Ultimately, our software must be easy to use, minimize the possibility of error, and enable users and application developers to access the full functionality of the underlying system. The NPACI partnership is focusing on usability as a fundamental characteristic of our software infrastructure so that our computational and data management resources are more accessible and effective for existing and new users.

The process of making software effective, efficient, and satisfying to use is called usability engineering.

This process makes it possible to:

- Find ways to capitalize on what users know so that less learning is required to use software.
- Minimize opportunities for error by structuring operations so they match users' common-sense notions about software behavior.
- Help the user recover easily when errors do occur (for example, by incorporating "undo" functions).

Usability engineering is not just a matter of developing a more aesthetic user interface or writing better documentation. It involves understanding the tasks users need to accomplish and how they think about those tasks. Drawing on that base, usability engineering establishes what functionalities should be available to the user at specified points and how those functionalities can be more obvious and natural.

With NPACI's increasingly complex programming environments and the challenge of building software for coordinated distributed platforms like the TeraGrid, it's even more important that our software support users as they try to solve the next generation of scientific questions. Usability engineering should be incorporated in the design and development of software from the start, but it is also critical in readying software for widespread use. As software designers shift from prototype development to production, usability engineering is part of the necessary hard work that makes the final application easy to use.

Over the last year, we began to apply structured usability assessment and improvement to NPACI-related software. An important example of this is the Storage Resource Broker (SRB), which is undergoing usability engineering with NPACI partner and Earth Systems Science thrust co-leader Cheri Pancake.

The SRB technology is being used in the production mode to support digital libraries, data grids, and

persistent archives. These systems use either C Library interfaces or Unix shell commands to manage their digital holdings. To enable the dynamic creation of collections, a new Web-based interface called MySRB is being developed to manage and explore data across distributed heterogeneous sites.

A key factor for MySRB effectiveness is the need for special data, called metadata, to track where and how data sets are stored, what they mean, and how they should be used. Metadata is needed for data to be used by anyone other than the original data providers. The higher the quality of the metadata, the more likely the digital archives will be used in the future.

The MySRB interface provides functions for creating a collection, adding metadata attributes, and manipulating the associated digital entities, all of which makes usability engineering of the interface design essential. An initial usability review showed that the functions should be simplified to avoid ambiguity, reduce complication, and minimize user error.

We are restructuring the MySRB user interface to do a better job of guiding users through the process of developing and managing metadata. Usability engineering is being applied to restructure what users see and when they see it to better match how they go about creating, changing, and browsing metadata. The new interface eliminates many opportunities for user error and improves self-instruction and usability by presenting MySRB's functionality in ways that are more obvious to scientific users.

The information and data management infrastructure being developed by NPACI partners is intended for broad use by computational scientists, engineers, students, and others. Developing, hardening, and engineering our products for usability are key steps in helping users access our resources easily and effectively. Usability engineering methods can be applied to community codes, application environments, data repositories, and even VCRs.

Experience in science and industry has shown that usability engineering can make a dramatic difference in how well, and how widely, software is adopted. While usability has become a fundamental tool in developing consumer electronics, it is also essential to NPACI as increasing numbers of our collaborative research teams rely on environments supported by distributed data archives, grid computing, remote sensors, personal digital assistants, and many other technologies. ▼



By FRAN BERMAN
NPACI and SDSC Director

Designing Personalized Hearing Devices

PROJECT LEADERS

LESZEK DEMKOWICZ, CHANDRAJIT BAJAJ
University of Texas, Austin

PARTICIPANTS

ADAM ZDUNEK
Aeronautical Institute of Sweden

RICHARD CHARLES
SDSC

JAMES DEMMEL
UC Berkeley

WALDEK RACHOWICZ
Cracow University of Technology, Poland

PETER MONK
University of Delaware

ULRICH LANGER, JOACHIM SCHOEBERL
University of Linz, Austria

JOHN VOLAKIS, EDWARD DAVIDSON
University of Michigan

DANIELE BOFFI
University of Pavia, Italy

IVO BABUSKA, ANDRZEJ BAJER, SATISH CHAVVA, HAO LING, DEAN NEIKIRK, WALDEK RACHOWICZ, JESSICA SUN, TIMOTHY WALSH, GUOLIANG XU
University of Texas, Austin

JACK LANCASTER
University of Texas Medical Center, San Antonio

Imagine hearing a rattlesnake but not knowing which direction to run. Proper localization is an essential part of hearing perception, and losing that ability is an unsettling aspect of hearing impairment. Hearing aids can amplify sounds, but they can't provide the perceptual cues required to ascertain their origin. As part of an effort to restore this sensibility to people who wear hearing devices, NPACI researchers in the Engineering thrust are using mathematically intensive techniques to trace the path acoustic waves take around and through the head. The path subtly alters sounds in a way that allows people with normal hearing to locate the source of sounds. The findings could be used to build customized hearing devices that would enable wearers to locate the sources of sounds.

"Hearing aids are there to improve hearing, but they don't attempt to address the localization problem," said Chandrajit Bajaj a professor of computer science at the University of Texas, Austin. Bajaj also is the chair in visualization at the Texas Institute of Computational and Applied Mathematics (TICAM), an NPACI partner. Currently, designing and tuning a hearing device that provides localization requires incorporating clues related to location and frequency, a process that involves a trial-and-error approach that depends on the feedback of the wearer. However, for many people, particularly young children, properly tuning a hearing device is virtually impossible.

Bajaj and Leszek Demkowicz, a fellow researcher at TICAM, have been working on models that determine the pressure on the eardrum depending on the location and frequency of the sound waves. Their work may someday be used to determine how hearing aid design will be based on the size, shape, and other characteristics of an individual's head.

The human head is shaped irregularly with bone, muscle, and fat—materials that are challenging for researchers to model in order to understand the pattern of acoustic waves passing around it. Modeling requires a twofold approach—carefully mapping the geometry of the head's surface and describing its interaction with acoustic waves. The computational approach involves draping the head with a digital mesh consisting of fixed points, aiming acoustic waves at those points, and recording the flow.

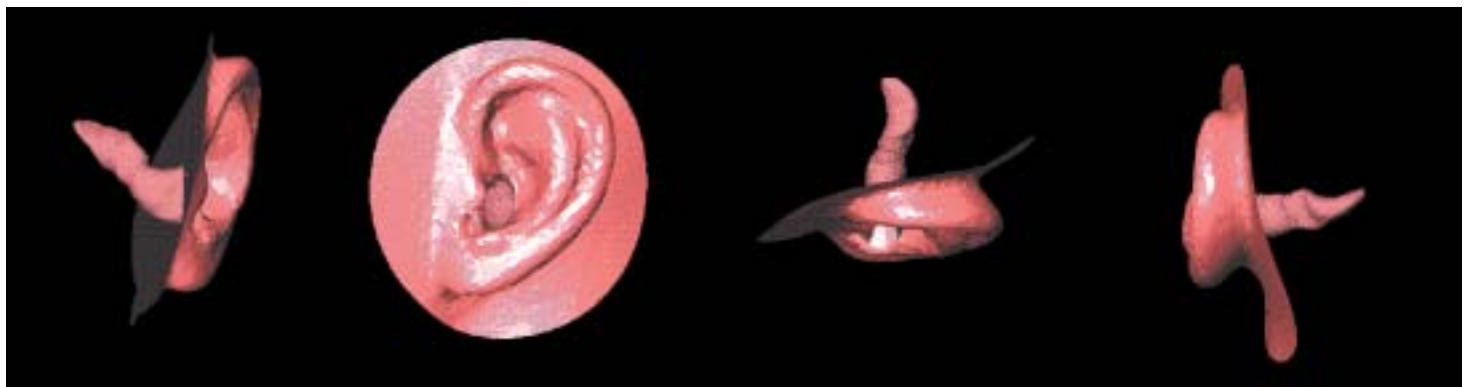
"We looked at MRI (magnetic resonance imaging) data and used that to reconstruct a model of the human head," said Bajaj. "We did the meshing based on that to show that we could do this in a customized setting. So if one wanted a patient-specific calculation, you could do that."

Bajaj and his colleagues initially worked on mesh geometry, while Demkowicz and his group focused on solving the propagation of waves through it. "To solve the problem accurately, we tweak the modeling of the human head to come up with smooth geometries, and balance that with faster solution techniques so the numerical integration does not suffer," said Bajaj.

MODELING MESH

Bajaj uses a toolbox of geometric modeling and visualization methods to solve the problems. His mesh-generating tools start with a set of two-dimensional images, such as MRI slices, and extract 3-D geometries from them. The object's geometry is expressed as a mesh, a lattice of tetrahedra or hexahedra (blocks) called finite elements. Such imaging data can be dense, and the goal is to extract the geometry adaptively, providing high resolution only where necessary. The challenge is to do it rapidly for large data sets.

Simulation results must be plotted onto such underlying finite-element meshes. Accelerated isocontouring tools analyze the physical features in the results—such as flow, temperature, stress, or electro-



SOUND CHANNEL

These are views of the ear canal, as connected to an ear mesh. Produced early in a project to model acoustic pressure around the human head by Leszek Demkowicz and his colleagues, the simulation uses parallel hp-boundary element method generate an image of the middle ear.

magnetic fields—and plot isocontours of functions on the surface mesh. Points on the mesh with the same value are given the same color and transparency, similar to elevation contours on a topographical map. The basic visualization technique is used often by scientists for two-dimensional and 3-D simulation data.

“There are difficulties in modeling the human head, which is the problem of building conforming and smooth meshes from imaging data,” said Bajaj. “There also are difficulties in modeling and accurately computing the integration for acoustics scattering.”

Currently, Bajaj uses a 128-processor Compaq cluster and a 24-processor Silicon Graphics Onyx 2 at TICAM to perform his simulations. Bigger machines wouldn’t alter the outcomes, but they would more quickly generate the models, which would be important in a clinical setting when a patient is waiting to have a hearing aid customized.

The project was proposed by Rich Charles, then at SDSC, after Demkowicz investigated a variety of acoustic and electromagnetic scattering problems. Demkowicz’s graduate student, Tim Walsh, then picked up the project for his doctoral thesis. For several years, Bajaj’s group, which included visiting scientist Guoliang Xu, worked on the geometric modeling of the head, while Walsh and Demkowicz worked on the solver. Charles provided additional assistance in constructing a geometrical model for the ear canal.

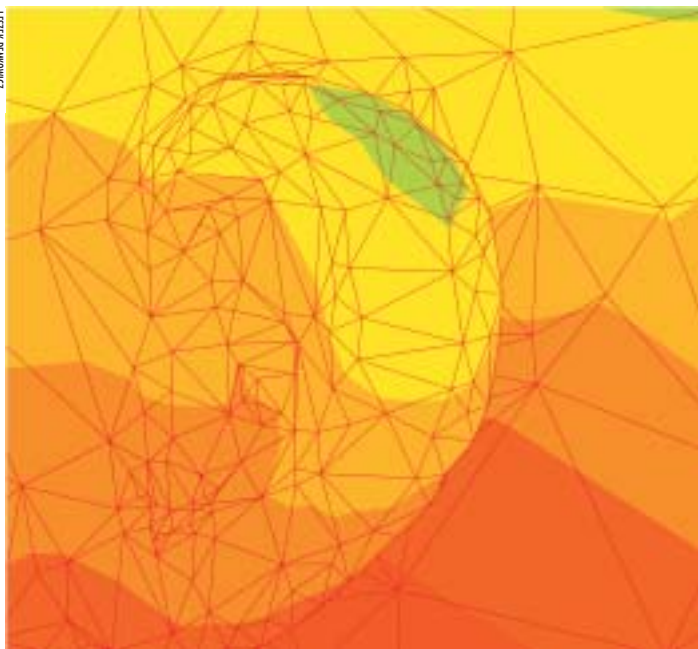
After Walsh graduated, Bajaj’s group took over the solution portion of the problem. “In some sense, all the balls are in our court now, which includes all the modeling, meshing, and the solution of the Helmholtz equation,” said Bajaj. The Helmholtz equation, one of



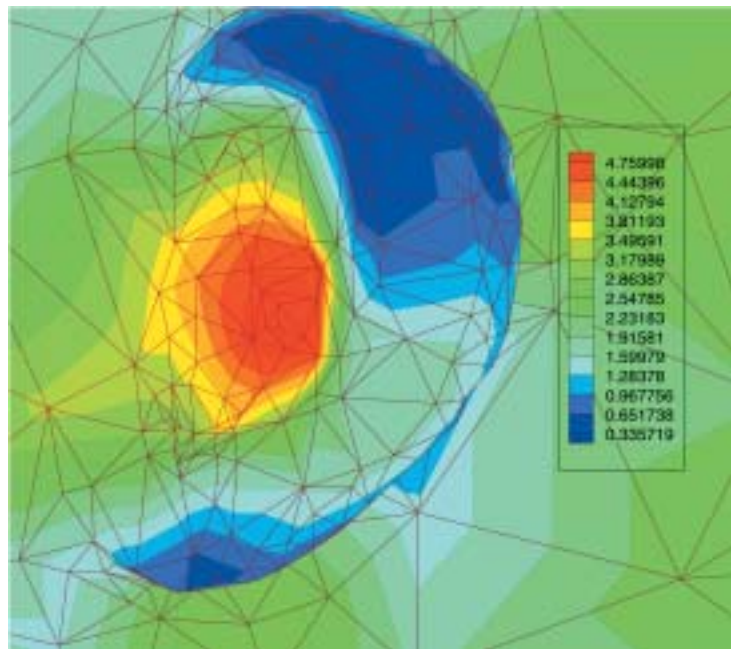
LESTER DEMKOWICZ

INNER EAR SIMULATION

The geometrical complexities associated with the ear canal, along with the possibility of singular solutions in such domains, necessitate the use of adaptive finite elements. This simulation shows the pressure distribution within the ear canal in resonance mode, 2900 Hz. Sounds of this frequency create different levels of pressure along the canal, which are represented by the blue-to-red color gradient.



ZIMKOWICZ



ZIMKOWICZ

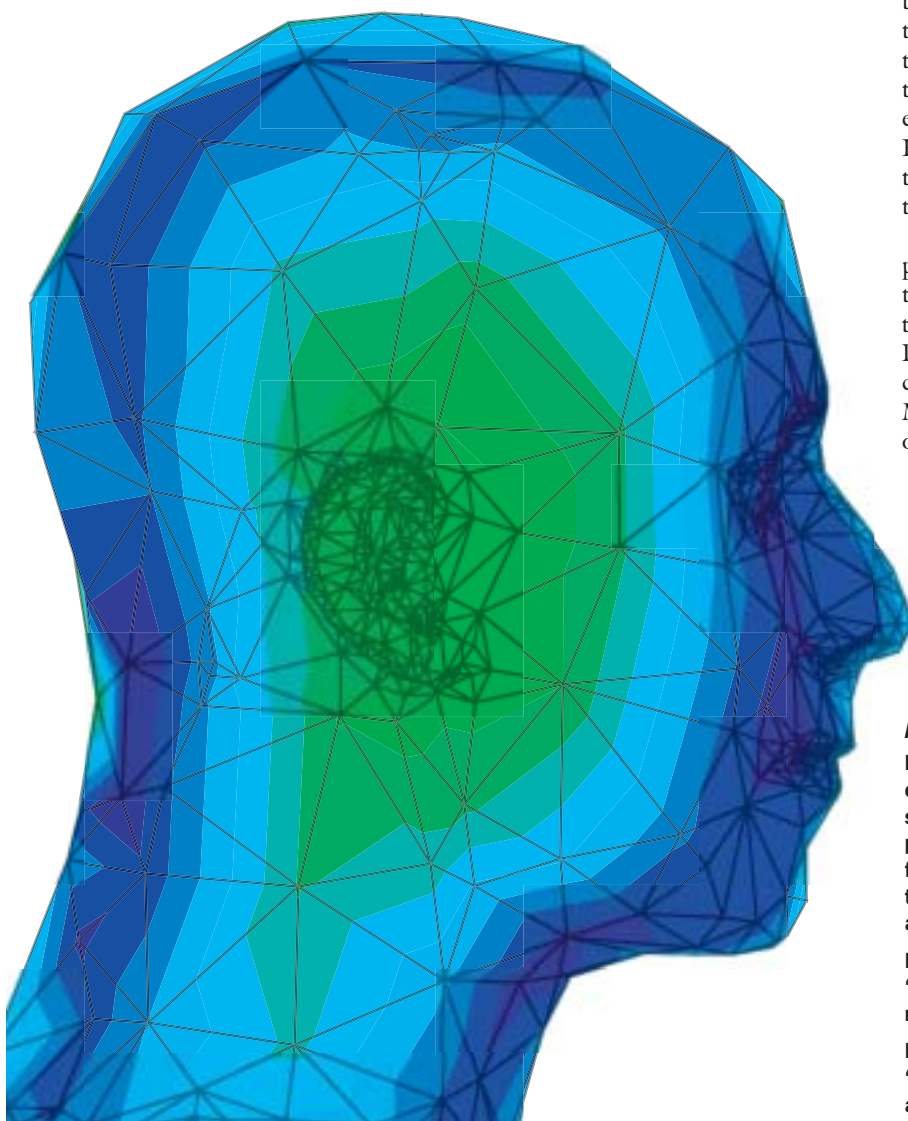
IMPACT ZONES

Most sounds are made up of a complicated mixture of vibrations. A sound spectrum is a representation of a sound—usually a short sample of a sound—in terms of the amount of vibration at each individual frequency. It is usually presented as a graph of either power or pressure as a function of frequency. As the frequency increases, the areas of highest pressure become more and more localized on the impact zone. The warmer colors indicate areas of higher pressure, with red being the most intense, while the cooler colors show the area of lower pressure, with green being the lightest. At higher frequencies (right) the pressure becomes greatest at the ear’s opening, or concha.

SDSC’s “grand challenge” equations, is used in acoustics and electromagnetic studies. It arises, for example, in the analysis of vibrating membranes, such as the head of a drum. “We’ve taken Leszek’s work, and we’re pushing it further. We are developing parallel cascadic solvers based on surface and 3-D recursive subdivision schemes. We don’t have the most efficient solver yet, but we are getting there.”

Bajaj’s group is fine-tuning its code for the solution, and coupling it to the meshing and visualization codes. The efficiency of the code depends on domain modeling and the fast convergent technique used to compute the numerical solution.

The problem of modeling the human head began in Demkowicz’s lab as part of a larger project working out better ways of solving acoustic and electromagnetic scattering problems. For two decades, J. Tinsley Oden, director of TICAM and leader of the NPACI Engineering thrust, and Demkowicz, a professor in the Department of Aerospace Engineering and Engineering Mechanics at Texas, pioneered the use of adaptive methods that automate the mesh construction process to deliver superb accuracy at low cost.



CHANDRAJIT BAJAJ

Working with meshes generally involves changing either the mesh size or the order of approximation. The mesh size is usually referred to by the variable h , and the order of approximation with the variable p , so these methods are usually called h -adaptive and p -adaptive. However, certain common types of engineering problems, such as modeling acoustic and electromagnetic waves, require methods that automatically vary both the size and order of approximation—called hp -adaptive methods.

Demkowicz’s work grew out of earlier research by renowned TICAM mathematician Ivo Babuska, who demonstrated theoretically that an hp -adaptive approach could deliver exponential convergence rates for both regular and irregular solutions. “For the last 13 years, I have been one of rather few people who have tried to translate that theoretical result into a practical engineering tool,” said Demkowicz. “Coding hp -methods is extremely challenging.”

FROM THEORY TO SOFTWARE

To illustrate the difficulty of the problem, Demkowicz tells the story of a 150-page paper written by German mathematician Leo Korn on a fundamental mathematical result now known as Korn’s inequality. Hermann Weyl, a leading mathematician at the time, said he was able to read only the first 50 pages of the article. “That paper provided a foundation for the entire modern theory of elasticity,” said Demkowicz. He notes that functional analysis and partial differential equations, which were developed later, now allow the same theory to be proven in less than two pages.

Oden, Demkowicz, and their colleagues have incorporated advanced mathematics into versions of the two-dimensional and 3-D hp finite-element code for the first time. The codes eventually resulted in PHLEX, the first hp -adaptive finite-element commercial software, developed at the Computational Mechanics Company. Applications included a variety of complex problems in solid and fluid mechanics, focusing especially on supersonic compressible flows.

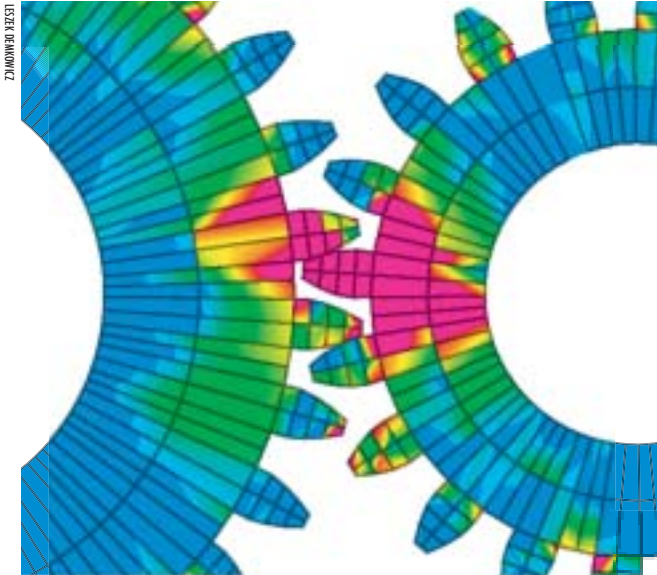
For the past four years, they have been working on parallel implementation on memory-distributed platforms, such as the Cray T3E, which has included an evolution through three programming languages and versions of the code. “The idea has been to minimize changes in the data

ACOUSTIC PRESSURE PROFILES

NPACI researchers have been applying hp -adaptive boundary element methods to the acoustics of the human auditory system. The goal of the project is to determine the acoustic pressure on the eardrum. The so-called acoustic transfer functions are used to design and tune hearing devices. The tuning is especially challenging in children for whom the “trial and error” approach is virtually impossible.

LEFT: This head shows the pressure distribution in the “shadow” zone resulting from a plane wave at 500 Hz, at normal incidence.

RIGHT: This image shows the pressure distribution in the “impact” zone at the same frequency. The peak in pressure around the ear is referred to as the “bright spot.”



FROM EARS TO GEARS

The same code used to model the human auditory system, *hp*-adaptive technology, can be applied to planetary gears. Dynamical behavior of planetary gear train, consisting of a sun gear, four planets, and a carrier, is shown modeled as general dynamic contact/impact of elastic, nearly rigid bodies. The problem reduces to a successive solution of a large number of single steps. Each single step involves a solution of a large system of linear equations and inequalities. The developed parallel simulator is based on 2Dhp90—the two-dimensional *hp* code—and has been developed for the Cray 3TE. The work constituted the Ph.D. dissertation of Andrzej Bajaj in Demkowicz's lab.

hensive model of the human auditory system. Clearly, modeling the acoustic pressure on the eardrum is just one step. The next thing is to model the dynamic of how the sounds get amplified within the tunnel and finally how it gets beautifully spread out into a spectrum in the cochlea.

"It's little epsilons that we are doing," said Bajaj. "As our ability to use computers to prototype various physical phenomena improves, including acoustic scattering, we can make

those solutions very numerically robust and customize them for individual patients." —CF ▼

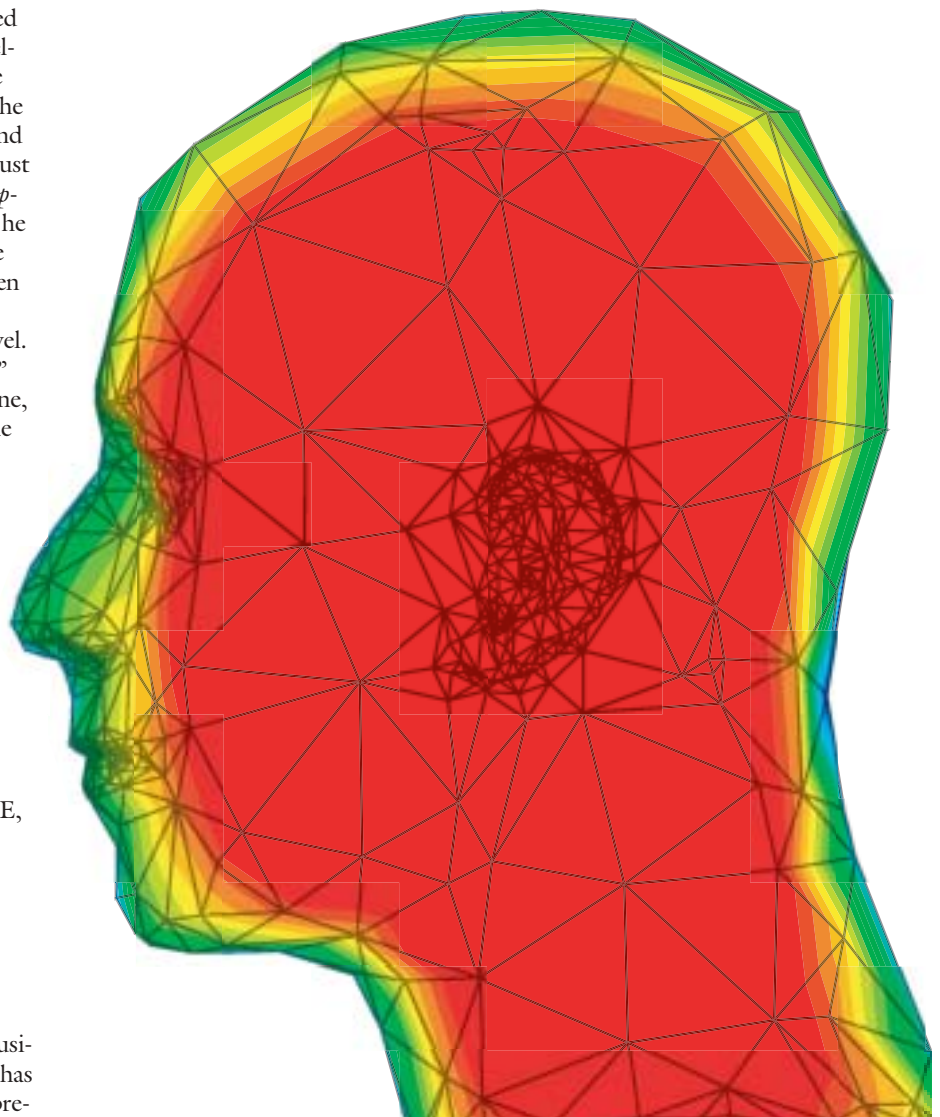
structure and to recycle as much as possible of our existing software," said Demkowicz. "It is a painful learning process."

Part of that pain was trying to build on a successful code for modeling two-dimensional objects. Inspired at an NPACI All-Hands Meeting, Demkowicz developed a new 3-D data structure. "After that came the breakthrough with automatic adaptivity," he said. The code became "smart" enough to control the error and modify the discretization accordingly. Rather than just using the *h*-refinements (breaking elements) or the *p*-refinements (increasing locally polynomial degree), he could use both. His simulations begin with a coarse mesh, solve the problem, estimate the error, and then generate an adaptively refined mesh. The process is repeated until the error declines to an acceptable level. "With *hp*-methods, the choice is much more subtle," he said. "You have to choose not only where to refine, but also how to refine. Many people working on the subject say this is impossible to accomplish."

Four *hp* finite-element codes (2Dhp90, 2Dhp90_EM, 3Dhp90, and 3Dhp90_EM) have been developed, documented, and placed on the Web. The Swiss Federal Institute of Technology in Zurich, the Aeronautical Research Institute of Sweden in Schlumberger, the University of Cracow, the U.S. Navy, and other institutions use them for teaching and research. The codes also have provided a basis for specific applications cofunded within NPACI: modeling acoustics and the human auditory system, modeling of planetary gears, and 3-D electromagnetic simulations using FE, and coupled FE/IE methods.

THE NEXT STEP

Bajaj wants to model the entire human auditory system. Once the framework for understanding the interaction of acoustic pressure on the eardrum is complete, the next step will be to continue farther along the ear canal. "It's like you have this entire musical system," he said. "Everything along the pipeline has to be modeled if you want to come up with a compre-



CHANDRAJIT BAJAJ

Developing Therapy for Non-Hodgkin's Lymphoma with Supercomputer Simulations

PROJECT LEADER

YUNI DEWARAJA
University of Michigan

PARTICIPANTS

KENNETH KORAL
University of Michigan

MICHAEL LJUNGBERG
Lund University, Sweden

ABHIJIT BOSE, RANDY CRAWFORD
University of Michigan

AMITAVA MAJUMDAR
SDSC

The overall mortality rate of cancer in the United States is declining, but increasing numbers of patients are dying from non-Hodgkin's lymphoma, the nation's fifth leading cause of cancer. Successful clinical trials of therapies using monoclonal antibodies to deliver radioactive isotopes to tumors are offering hope to patients with non-Hodgkin's lymphoma. Now, University of Michigan nuclear medicine researcher Yuni Dewaraja and her colleagues are developing a method to improve the effectiveness of that therapy with patient-specific supercomputer simulations that will more precisely estimate the amount and distribution of radiation absorbed. "Usually, we apply supercomputers to advance fundamental science that is some distance from practical applications," said SDSC's Amitava Majumdar. "It's not often that you can use a supercomputer in a way that could eventually directly help a cancer patient."

Non-Hodgkin's lymphomas develop in the lymphatic system, a network of thin tubes that branch like blood vessels into tissues and organs throughout the body. The tumor cells often crowd out normal cells within bone marrow, lymph nodes, and the thyroid gland. The usual chemotherapies for these tumors are nonspecific, indiscriminately harming normal and cancerous tissues alike. However, new radioimmune therapies treat patients with a source of radiation such as the isotope iodine-131 (¹³¹I) in a way that selectively targets the cancer cells, largely sparing normal tissue. This provides more effective treatment with fewer side effects.

Radioimmune therapies rely on monoclonal antibodies, proteins made in the laboratory by immortal lines of antibody-producing cells. Many monoclonal antibodies are used in cancer therapy; each recognizes and binds to a different protein on the cancer cell such as those of non-Hodgkin's lymphoma. Monoclonal antibodies can be coupled, or conjugated, in the laboratory to a toxin, in this case ¹³¹I.

Once absorbed by the tumor, ¹³¹I performs two important functions. The radioactive isotope emits short-range beta particles, or electrons, which are ionizing and kill nearby cells. The ¹³¹I isotope also emits gamma photons that can pass through the body. They can be detected by a gamma camera, which rotates 360° around the patient's body, like an X-ray CT scan, detecting the gamma photons at many positions. Reconstructing the data using Single Photon Emission Computed Tomography (SPECT) produces a 3-D image showing the location within the patient of the radiation and its quantity.

Michigan physicians Mark Kaminski and Richard Wahl have conducted phase II clinical trials of monoclonal antibodies conjugated with ¹³¹I to test the safety of the new drug and evaluate how well it works. In these trials, tumors disappeared in 48 of 76 non-Hodgkin's

lymphoma patients, and shrank at least 50 percent in another 26 patients. ¹³¹I has also shown promise in treating the advanced metastatic stage of the deadly childhood cancer, neuroblastoma.

TREATING WITH LEAST HARM

The success of the clinical trials has spurred research sponsored by the National Cancer Institute to develop imaging methods capable of predicting absorbed radiation dose and revealing dose-response relationships. "Our goal is to be able to more accurately correlate the amount of radionuclide administered with the absorbed dose in the tumor and how the patient responds," said Dewaraja. The potential benefits of reliable early prediction include better screening to identify which patients will benefit from these treatments, and earlier planning of subsequent therapy. In addition, accurate predictions of the dose of radioactivity absorbed by healthy tissue will improve estimates of the maximum dose of such therapy the patient can tolerate. "While often less toxic than chemotherapy, physicians still need to know the maximum dose they can give without harming the patient," said Dewaraja.

When a patient diagnosed with non-Hodgkin's lymphoma comes for treatment, physicians first order a standard X-ray CT scan to image the size and location of the patient's tumor and internal organs. Next, a small tracer dose of about five millicuries of ¹³¹I-labeled antibody is given to help assess how much of a larger therapeutic dose will be absorbed by the patient. Then, in a series of steps the researchers are developing, physicians will combine the X-ray CT scan and the SPECT tomography derived from the tracer dose to guide the building of a digital representation of the individual patient's body and tumor—a detailed 3-D virtual patient, or "voxel phantom."

Using the voxel phantom in a supercomputer simulation of a full therapeutic dose, the researchers will then be able to estimate the true radioactivity distribution in the real patient. Currently, the researchers are validating this simulation method, including developing better quantification of the SPECT gamma data. "Eventually, our research will also use simulations of

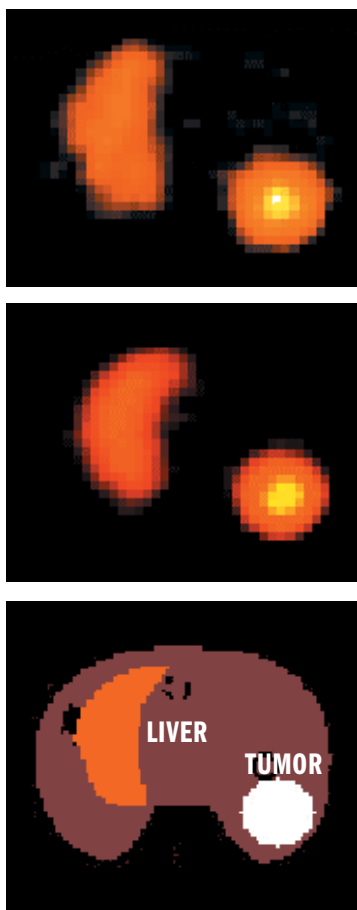


FIGURE 1. SIMULATION AND TOMOGRAPHY
To verify supercomputer simulations using the SIMIND code, researchers have compared them to a known radiation distribution within an experimental patient (top), the radioactivity distribution measured with SPECT gamma tomography (center), and the simulated distribution using the SIMIND code (bottom). They found good agreement among simulation, measured, and true values.

the transport of electrons in the tumor to determine patient-specific radiation dose distributions,” said Dewaraja. “This should be much more accurate because we will simulate the actual energy deposition in the tumor of each patient.”

BIGGER COMPUTERS ARE BEST

To simulate the radioactivity distribution and gamma photon paths, the researchers began with the SIMIND serial Monte Carlo code, which runs on a single computer processor. Michael Ljungberg of Lund University in Sweden developed the code. The simulations model the path of each gamma photon—from its emission from an ^{131}I isotope to its entry into the camera. In computing each photon’s path, the simulation must model absorption and scattering within the patient’s tissues. The Monte Carlo statistical method has proven highly accurate in modeling these interactions. It randomly assigns photon absorption and scattering events based on known probability density functions for how photons interact with the various types of tissue. In addition to requiring a highly accurate parallel random number generator, a major challenge in using the Monte Carlo method has been the long computation times required to simulate large numbers of photons.

“We found there were large errors even when using simple shapes like cylinders to approximate the tumor, especially for smaller tumors,” said Dewaraja. That is, the radioactivity distribution depends sensitively on the specific geometry of the patient and tumor. For large, spherical tumors the quantification error of the estimate is less than 3 percent, but for a smaller or irregular tumor the error can reach 50 percent. Thus, accurate simulations require the use of a more realistic, individualized 3-D virtual patient.

Accurate simulations of the ^{131}I radioactivity distribution are computationally demanding not only because of the need to use a high-resolution 3-D virtual patient, but also because of the need to use Monte Carlo methods to simulate millions of individual photons. Because of these factors, CPU time can exceed one month using the single-processor SIMIND code. To cut the time required, Dewaraja and Ljungberg collaborated with Abhijit Bose, also at NPACI partner Michigan, in developing a parallel version of the SIMIND code that initially ran on NPACI’s IBM SP2 at Michigan.

“The problem is embarrassingly parallel,” said Bose. “Physically, it’s a number of independent events, so to run a large number of photon simulations you just divide them among the processors.” Each processor performs the entire simulation for all photons assigned

to it and reports the results to the host processor, which calculates the final result.

Even with an efficient parallel code, it’s necessary to simulate millions of photons to achieve sufficient accuracy, and the researchers found a major speed increase from running the parallel code. A simulation that would have taken a month with one processor, takes four to five hours on 512 processors of NPACI’s Blue Horizon. The researchers expect full-machine runs on the 1,152-processor Blue Horizon to take as little as two hours. “We anticipate it will easily scale to 1,000 processors, and more—it looks like a natural migration to the clusters of the TeraGrid,” said Majumdar.

NPACI has also contributed to the research through a Strategic Applications Collaboration. In addition to integrating the Scalable Parallel Random

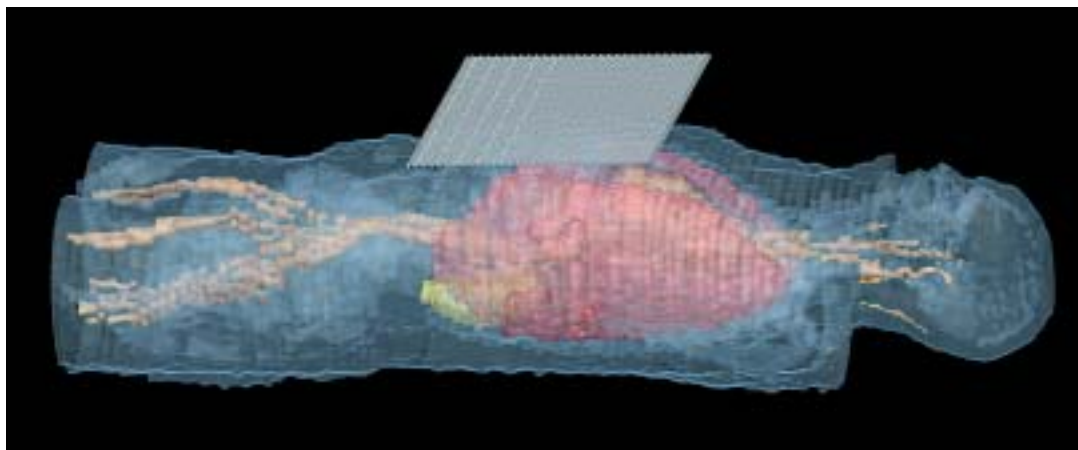


FIGURE 2. VIRTUAL PATIENT

This “voxel man” is a high-resolution, 3-D digital representation of a patient. The Zubal Phantom, obtained from researchers at Yale University and based on X-ray CT and MRI images, is being used by Michigan researchers to verify quantification of radioactivity uptake in tumor and organs by Single Photon Emission Computed Tomography (SPECT). The gray area is the collimator for the SPECT gamma camera.

Number Generator to enable accurate parallel Monte Carlo simulations, another important goal has been to bring the benefits of NPACI’s advanced parallel supercomputers to the medical imaging community, which until now has made little use of them.

TOWARD CLINICAL APPLICATION

“Previously, people in our field thought it was impossible to simulate an individual patient because it would simply take too much computer time,” said Dewaraja. “But through the faster parallel code and having access to large NPACI machines we’re showing that it’s within sight to do realistic simulations for individual patients.”

One of the researchers’ goals is to produce a tool that clinicians can use to accurately quantify the uptake and absorption of radioactivity in the tumor and healthy tissues of each patient. Other simulations will model the activity and dose of ionizing beta particles within the tumor. “But the most rewarding part is creating an application that may eventually have clinical use in helping cancer patients,” said Dewaraja.

—PT ▼

REFERENCE

Y. Dewaraja, M. Ljungberg, A. Majumdar, A. Bose, and K. F. Koral: A parallel Monte Carlo code for planar and SPECT imaging: Implementation, verification and applications in ^{131}I SPECT, *Journal of Computer Methods and Programs in Biomedicine* (in press).

Determining Protein Structures with High-Throughput Techniques

PRINCIPAL INVESTIGATOR

IAN WILSON
The Scripps Research
Institute

CO-PRINCIPAL INVESTIGATORS

JOHN WOOLEY
UCSD

KEITH HODGSON
Stanford University, the
Stanford Synchrotron
Radiation Laboratory

PARTICIPANTS

ADAM GODZIK
SDSC

PETER KUHN
Stanford University, the
Stanford Synchrotron
Radiation Laboratory

RAYMOND STEVENS
The Scripps Research
Institute, the Genomics
Institute of the Novartis
Research Foundation

SUSAN S. TAYLOR
UCSD

Scientists have sequenced the genomes of more than 800 organisms, and as the list grows researchers in California and elsewhere are racing to understand what the tens of thousands of proteins encoded by the DNA blueprints actually do. The crucial first phase of that formidable task is structural genomics, the process of determining the 3-D structure of all the proteins to gain insights into their functions and interrelationships. Structural genomics will enable researchers to understand how flawed proteins cause diseases and how to design drugs to treat those diseases.

In recognition of the enormous potential to benefit humanity, the National Institutes of Health (NIH) has funded nine multi-institution consortia seeking to determine protein structures with assembly-line approaches (Figure 1). One of these “high-throughput” projects, the Joint Center for Structural Genomics (JCSG), involves SDSC and UCSD, The Scripps Research Institute, the Genomics Institute of the Novartis Research Foundation (GNF), Stanford University and the Stanford Synchrotron Radiation Laboratory, and the Salk Institute. The NIH’s National Institute of General Medical Sciences has awarded five-year grants to the nine projects as part of its Protein Structure Initiative. “Our ambitious initiative in the United States is part of a worldwide effort,” said John Norvell, program director for the Protein Structure Initiative.

Major structural genomics consortia have formed in Japan, Germany, the United Kingdom, and Canada, with other efforts under way in Sweden, France, Italy, Brazil, and China. The International Structural

Genomics Organization was founded in 2001 to coordinate the goals of all the projects. The U.S. effort is focused on finding the structures for representatives of all the families of proteins and protein “folds.”

SCIENCE OF STRUCTURE DETERMINATION

Proteins are strings of amino acids. The order of the amino acids along the string can be determined by several techniques, but the 3-D structure of any protein is hard to predict because amino acids distant from one another often come in close contact. An understanding of proteins in living cells is possible only when their 3-D structures have been determined experimentally, revealing folds and other features. “The goal is to fully yet efficiently characterize the range of protein folds,” said Norvell.

A previously unknown protein fold, which also is a possible new drug target, was among the first findings in JCSG’s highly promising past year. “We’ve hit the ground running,” said project director Ian Wilson of Scripps. “There are more surprises in store.”

The objectives of JCSG are being carried out in a systematic, step-by-step fashion (Figure 2). “Efficient coverage of families and fold space requires strategic selection of target proteins for solution,” said Susan S. Taylor, a professor of chemistry and biochemistry at UCSD, chairperson of JCSG’s target selection effort, and a member of the National Academy of Sciences. With global coordination, she said, the community could minimize the number of structure solutions required to serve as templates for solving others.

“Success in these projects will change the science of structure determination,” said Wilson. The advantages will include a tenfold reduction in the cost per protein solution, a vast increase in data available to biomedical scientists, and an ability to answer more complex questions. “We will be able to ask not only ‘What does this protein do?’ as we can today,” Wilson said, “but also ‘What do all these proteins do collectively?’ or ‘Why do some metabolic pathways of protein activity occur in some organisms and not others?’ and ‘How might we specifically interfere with them in microbial pathogens?’—questions at a more holistic level.”

A YEAR OF DISCOVERY

“This year we have identified and, in many cases, opened up several bottlenecks in the process of high-throughput structure determination, but not all of them,” said Raymond Stevens, leader of the JCSG Crystallomics Core at Scripps. Stevens and Peter Schultz, also at Scripps, are speeding up structural genomics with robotics designed for rapid expression, purification, and crystallization of proteins.

Some of the first structures solved by JCSG are produced by the bacterium *Thermotoga maritima*, one of the planet’s oldest life-forms, which was discovered in geothermally heated marine sediments. The bacterium’s 1,877 genes code for relatively small proteins that remain active at 80° C and crystallize relatively easily in the laboratory. “Indeed, we chose this system because it provides an opportunity to do structural genomics on an entire organism, and it enables an ideal end-to-end test of our system,” said Wilson.

The GNF team, led by Scott Lesley, successfully



FIGURE 1. HANDLING CRYSTALS ROBOTICALLY

A device at the lower end of this robotic arm withdraws a crystal from a canister before placing it into a beamline at the Stanford Synchrotron Radiation Laboratory. The diffraction of X-rays passing through the crystal reveals underlying molecular structure.

amplified 1,791 of the 1,877 *T. maritima* genes in the high-throughput system, and 1,369 were cloned and expressed in a host bacterium. The proteins were purified, then crystallized, a necessary step for structure determination by X-ray crystallography.

Historically, crystallization of a protein has been a trial-and-error enterprise that occasionally took years. Automation and miniaturization allows crystals to form faster from smaller quantities of protein. The second-generation robotic crystallization system developed by Stevens and engineers at GNF and Syrrx, Inc., can produce 60,000 crystallization trials a day, across a range of temperatures and other conditions. The system uses “nanodrops” (50 nanoliters) of protein solution. A robot scans all trials for crystal formation. By January 2002, the *T. maritima* trials had produced more than 300 high-quality crystals.

The crystals are sent to Peter Kuhn, who leads the Structure Determination Core at the Stanford Synchrotron Radiation Laboratory. Kuhn and his colleagues at the laboratory have designed and built an automated sample-changing system (Figure 1) to select, characterize, and transfer premounted crystals grown in nanodrops onto the synchrotron’s X-ray beamline for analysis. Data sets can be taken for up to 285 crystals without further human intervention.

In a beamline, X-rays pass through each crystal, a repetitive array of protein molecules, producing a diffraction pattern of numerous “reflections,” sets of X-ray spots. The position and intensity of the spots depend on the type and 3-D arrangement of atoms in the crystal. Many reflections are collected for each crystal and examined computationally to produce a structural solution: a map of the distribution of each atom in space. Each solution is further validated, and researchers add the final 3-D coordinates to the Protein Data Bank, the international repository of protein structures.

One of the JCSG’s solved protein structures is TM0449, an enzyme needed by *T. maritima* for replication and repair of its DNA. The enzyme has a completely novel protein fold, which has been added to the Protein Data Bank and international databases of protein topologies. “The significance is that, in pursuit of presumably simple proteins that have survived as long as life has been on Earth, we find that we can still discover something new,” said John Wooley, a JCSG co-principal investigator and associate vice chancellor for research at UCSD. The new protein fold was also found in a number of pathogens (including the deadly anthrax bacterium). The molecule may become a target for new antimicrobial drugs.

THE ROLE OF BIOINFORMATICS

In addition to miniaturization and process automation, bioinformatic technologies are making high-throughput genomics faster and more informative. “It isn’t just quantity we’re after,” said Adam Godzik, leader of JCSG’s Bioinformatics Core, located at SDSC. “We make the whole process faster because our database records successes and failures at each stage, information that enables us to raise the success rate. Our processes and results are accessible on our website.” As required by the International Structural Genomics Organization, all U.S. structural genomics projects have target lists accessible on the Web. Each project is committed to rapid deposition (within 4–6 weeks after completing refinement) of new structures in the Protein Data Bank.

“We have been rolling out variations on existing tools and developing new ones to aid the high-throughput model,” said Godzik. On the website, homology and structure analysis tools are augmented by other tools, including a Data Acquisition Prioritization System. The location of the Bioinformatics Core at SDSC also allows JCSG to take advantage of SDSC’s resources and expertise in manipulating large data sets.

“We have begun on a very promising note,” said Wilson. “Our efforts are now complementary to those pursuing structures one by one: we all find that the X-ray structure solutions often suggest functional roles for proteins. High-throughput structural biology will also facilitate comprehensive studies of complete metabolic pathways. It will lift all our questions onto a philosophically higher plane. We will be able to find out how complex assemblies of protein components provide a blueprint for the operation and function of a cell. Protein structure determination, per se, will no longer be the main obstacle in elucidating the molecular basis of biology.” —MM ▼

REFERENCES

R.C. Stevens, S. Yokoyama, and I.A. Wilson (2001): Global efforts in structural genomics, *Science* 294, 89-92.

R.C. Stevens and I.A. Wilson (2001): Industrializing structural biology, *Science* 293: 519-520.



FIGURE 2. THE STRUCTURAL GENOMICS PIPELINE

The Joint Center for Structural Genomics cloned a specific gene (TM0449) of the bacterium *Thermotoga maritima*, purified and crystallized the protein, and exposed crystals to an X-ray beamline as part of a high-throughput process of determining the 3-D structures of proteins.

Dickerson's Formula: Biochemistry's Equivalent to Moore's Law

PARTICIPANTS

ARTHUR ARNONE
University of Iowa

HELEN BERMAN,
JOHN WESTBROOK
Rutgers University

PHIL BOURNE
SDSC

RICHARD DICKERSON
UCLA

GARY GILLILAND
*National Institute of
Standards and
Technology*

Twenty-four years ago, Richard Dickerson came up with a mathematical formula that optimistically predicted an accelerating pace of discovery in the burgeoning field of protein structure determination with X-ray crystallography (see story, page 8). Dickerson, then a professor of physical chemistry at Caltech, noted that the number of protein crystal structures had risen from one solved by the end of 1961 to 23 solved by the end of 1977. His formula predicted that by March 2001, scientists would have solved the 3-D structures of a grand total of more than 12,000 proteins. He was very close.

Nobody knew how close until Arthur Arnone, a biochemistry professor at the University of Iowa, checked. Arnone found that the equation predicted that there would be 12,066 crystal structures solved by March 27, 2001. By that date, the Protein Data Bank (PDB), the international online repository of protein data, had posted 12,123 protein structures, only 57 more than Dickerson's forecast. The Dickerson formula was accurate to within 0.5 percent.

Arnone notes that the PDB tally on February 5, 2002, showed 13,635 X-ray structures for proteins, peptides, viruses, and protein-nucleic acid complexes. However, including the non-protein structures (the 601 nucleic acid structures and 14 carbohydrate structures) the total number of experimental X-ray structures in the PDB was 14,250. Dickerson's equation predicted 14,201.

"I think that Dickerson's equation is to protein structure what Moore's Law is to semiconductor chips," said Arnone. "Even the time frames are the same. Dickerson's equation and the current form of

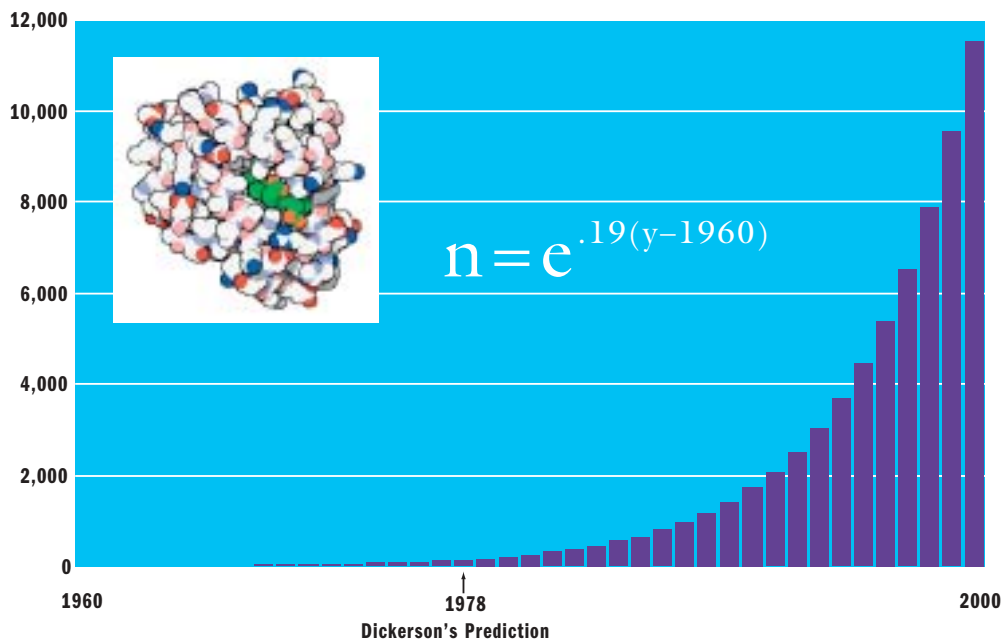
Moore's Law were both derived in the late 1970s." Moore's Law, created by Intel cofounder Gordon Moore, has forecast, with remarkable accuracy, that the number of transistors per square inch on integrated circuits would double about every 18 months.

Dickerson, now a professor of biochemistry at UCLA and a pioneer in X-ray crystallography, is astounded by his own predictions. "To move from the accidental to the ridiculous, if Dickerson's Law holds, the year 2024 will see the production of 1 million new protein and nucleic acid structures, or 83,000 new structures per month," said Dickerson. "Somehow, I doubt that. Even the genomics and proteomics researchers who claim that their goal is to sequence all the DNA of an organism, and solve all the protein structures coded by that DNA, don't go quite that far. But who knows? In short, I seem to have won the lottery. Pity there's no jackpot."

Phil Bourne, co-director of the PDB and director of Integrative Biosciences at SDSC, describes Dickerson's prediction as impressively insightful. "He put together the pieces and came up with this equation that basically predicted the exponential growth of the PDB. Back then, the numbers would have seemed staggering."

The details of just a single protein can easily fill a large book. Fortunately for crystallographers and other scientists, the growing encyclopedic catalog of proteins easily fits into the PDB, the world's largest repository and distribution center of protein data (www.rcsb.org/pdb).

Dickerson's equation predicts that a combined total of 24,667 structures will be found by 2004. "We're just starting to see an explosion of growth as the field of proteomics comes into its own," he said. "When I started out, it took years to solve a structure; now they're solving a couple thousand per year." —CF ▼



DICKERSON'S FORMULA PREDICTS RISING NUMBER OF SOLVED 3-D PROTEIN STRUCTURES

The first protein structure solved by scientists in 1961 was myoglobin (inset). In the equation developed by Richard Dickerson for the number of new protein structures solved per year, *n* is the number of new structures in a given year, *e* is the natural logarithm 2.71828, and *y* is the year (for March 27, 2001, *y* = 2000.25).

Biochemical Modeling Helps Explain Complex Neural Junction in Chicken Embryo

UCSD neurobiologist Darwin K. Berg and his coworkers have achieved a groundbreaking molecular-level description of the functioning of a specialized synaptic junction in the ciliary ganglion of embryonic chicks by combining results from anatomical, neurochemical, microscopic, and computer modeling studies. The study, presented at the annual meeting of the Society for Neuroscience in November 2001, combined the work of Berg's neurobiology group at UCSD, neuroscientists at UCSD's National Center for Microscopy and Imaging Research (NCMIR), and computational neurobiologists at the Salk Institute. "We have found this to be a particularly productive collaboration," said Berg. "We knew the power of the microscopy, but we were surprised by the power of the modeling."

The biochemical modeling program that Berg and his collaborators used is called MCell, which runs on SDSC's IBM supercomputer, Blue Horizon. The program is also at the center of an NPACI alpha project and an NSF Information Technology Research project led by NPACI Director Fran Berman. "I think we've shown that MCell is becoming a valuable means of checking on the results of chemical, microscopic, and electrophysiological studies of complex events occurring rapidly on the scale of microns to millimicrons," said Salk researcher Thomas Bartol. The ciliary ganglion is a small tangle of neurons that supplies nerve connections to the muscle controlling the pupil of the eye and to muscle in the eyeball. Berg, graduate student Richard Shoop, and postdoctoral researcher Jay Coggan spent several years characterizing the fine structure and anatomy of their target junction, elucidating the detailed functioning of neurotransmission across it. "We were studying the ciliary ganglion of the chick embryo because vision systems in birds are both well developed evolutionarily and more straightforward to study than in mammals," said Berg. "The ciliary ganglion is an exemplary form of a complex type of neural junction, quite distinct from the fine, dendritic junctions of brain neurons and from the much larger and coarser neuromuscular junctions found in other parts of the body."

SERENDIPITOUS COLLABORATION

"We had been studying it for several years as an interesting class of hybrid vertebrate synapses," Berg said. His group was using its own well-equipped chemical lab, while also taking advantage of electron microscopes available in the UCSD

Division of Neurobiology. "At one point, a few years ago, those microscopes went down for a rebuild," Berg said. "That was when we turned to NCMIR," in the lab of UCSD Medical School neuroscientist Mark H. Ellisman. "We became aware of MCell almost by accident," Berg said.

"We were delighted to work with the Berg group," said NCMIR Associate Director Maryann E. Martone. The NCMIR researchers had just finished a study of the large-scale neuromuscular junctions in the body—also modeled by Bartol and the group at Salk—and

PROJECT LEADER

DARWIN K. BERG
UCSD

PARTICIPANTS

JAY S. COGGAN,
MARK H. ELLISMAN,
EDUARDO ESQUENAZI,
STEPHAN P. LAMONT,
MARYANN E. MARTONE,
RICHARD D. SHOOP,
NAOKO YAMADA
UCSD

THOMAS M. BARTOL,
TERRENCE J. SEJNOWSKI
Salk Institute

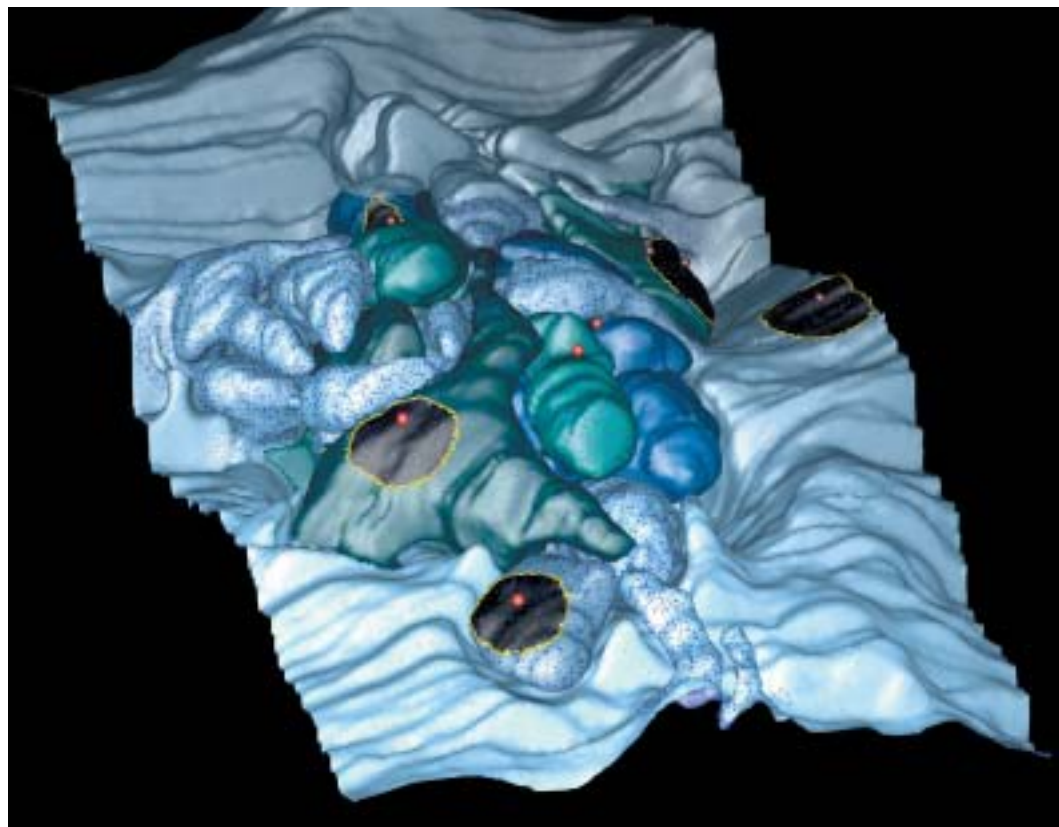


FIGURE 1. SPINE MAT MODEL

This picture of the somatic spine mat modeled by MCell shows neurotransmitter vesicles (large red dots) located at the end of a neuron (not shown). These vesicles release the neurotransmitter acetylcholine, which finds receptors (small blue specks) distributed across the spine mat.

ALL IMAGES: TOM BARTOL

Berg's smaller and more complex target seemed a logical next step. NCMIR, funded by the National Institutes of Health's National Center for Research

Resources, introduced Berg's group to their 400 keV electron microscope, which allows study of much thicker sections of tissue than those Berg had been using across the campus. "These make 3-D reconstruction of the tissue sections much easier," said Martone. NCMIR scientists have worked with a constellation of other scientists, including many from SDSC, to link the powerful electron microscope with a computational system in their own lab (now part of the Keck I satellite cluster of SDSC). NCMIR has also pioneered the development of programs for visualizing neuronal structure in three dimensions, using accurate computational reconstructions of whole neurons from serial sections.

A complete 3-D reconstruction of a portion of an embryonic chick ciliary ganglion was carried out by the collaboration in 1999 and 2000. The reconstruction showed the somatic spines, branching structures for receiving neural signals that emerge from and are tightly folded against the soma (body) of the neuron (Figure 1). "Somatic spines grouped in discrete clumps or mats are the hallmark of this kind of junction," Berg said. In the reconstruction, arbitrary colors were used to distinguish one spine from another, "but they almost certainly function together as a group," he said.

SYNAPSE SIMULATION

Neuromuscular junctions of all types are able to deliver high-frequency, reliable stimuli to the postsynaptic muscle fiber. In the chick ciliary ganglion, the neurotransmitter acetylcholine is released from a calyx, so called because the tiny structure is reminiscent of a flower part having the same name. The calyx envelops and surrounds the somatic bodies of the ganglion and overlies the mats of spines.

In their chemical studies, Berg and his group had found that the spine mat surfaces have two types of receptors for the neurotransmitter acetylcholine. One type, carrying a sequence called the alpha-7 gene product, is particularly permeable to calcium ions, allowing them into the cell, where they regulate many processes. The other, called alpha-3, is much less permeable to calcium. Paradoxically, the alpha-7 receptors

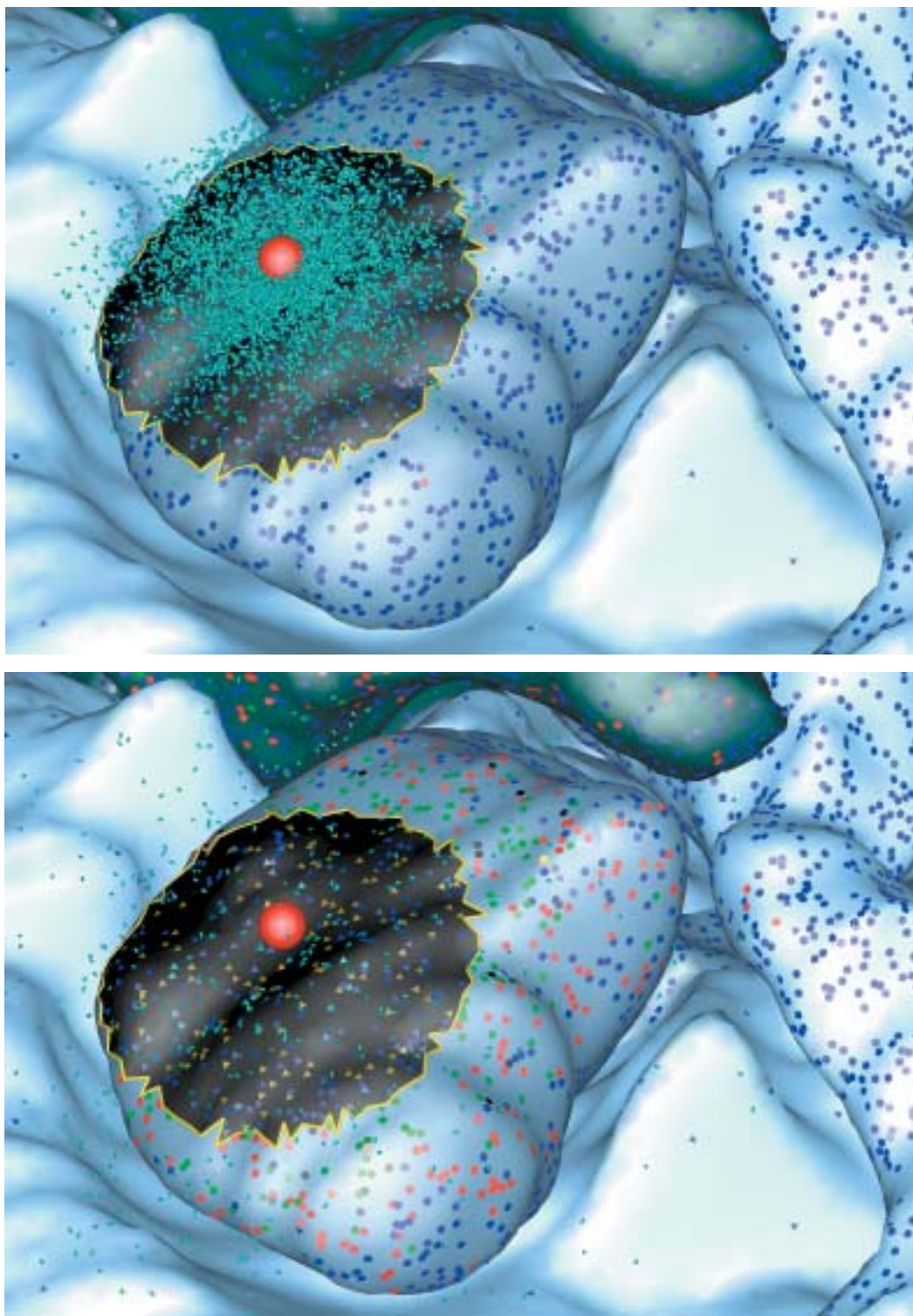


FIGURE 2. A POSTSYNAPTIC DENSITY

This pair of images highlights activity in the region of the postsynaptic density (black circular area) 10 microseconds (upper image) and 200 microseconds (lower image) after release of acetylcholine (small green ovals) from a synaptic vesicle (large red dot). Alpha-3 receptors (triangles) predominate in the postsynaptic density area while alpha-7 receptors (pentagons) are distributed outside this area. Colors indicate state of activation—prereceptive (blue), activatable (red, green, and brown), fully activated (yellow and orange), or deactivated (black). Diameter of the postsynaptic density area is about 0.5 microns.

were widely distributed across the somatic spines, while the alpha-3 receptors occupied only a few small areas called postsynaptic densities—key entryways into the cell. “We knew that both types of receptor were involved in neurotransmission across the ganglion,” Berg said, “but we needed to know exactly how they functioned—together or in competition with one another.” Martone and Ellisman suggested that the Berg group consider using Bartol’s MCell code to model the detailed, dynamic, time-dependent behavior of the receptors.

MORE DETAILED GEOMETRY

Bartol also was excited by the prospect. “MCell does detailed biophysical modeling, and the ciliary ganglion seemed an ideal next project to work on after the larger neuromuscular junction project,” he said. The researchers went from the visually reconstructed spine mat covered with receptors to a much more detailed geometry. “We had to put numbers on it,” Bartol said. This massive task was undertaken by Eduardo Esquenazi, a recent UCSD graduate working in Ellisman’s lab. Under the direction of Coggan, Martone, and NCMIR staff scientist Naoko Yamada, Esquenazi performed detailed measurements of the electron microscope data and added the chemical information from Berg’s lab.

Computationally, the geometric reconstruction for MCell also took advantage of the capabilities of a code called Xvortex, developed by Stephan Lamont, a computational visualization specialist in Ellisman’s laboratory. Lamont worked with Bartol, Esquenazi, and others on the project to make the necessary modifications to Xvortex.

CONFIRMING A PREDICTION

Some of the results of the modeling with MCell are shown in Figures 2 and 3. “What we found helped to confirm our idea that the two receptor types operate both cooperatively and competitively to modulate the overall response to neural signal in a way that makes the entire ciliary ganglion extremely sensitive,” said Berg. Thus, it is able to modulate the opening and closing of the chick pupil in response to changing light levels. “Despite the fact that the junction appears as a rather messy set of spine mats, its overall functioning can be rapid and sensitive enough to permit very fine

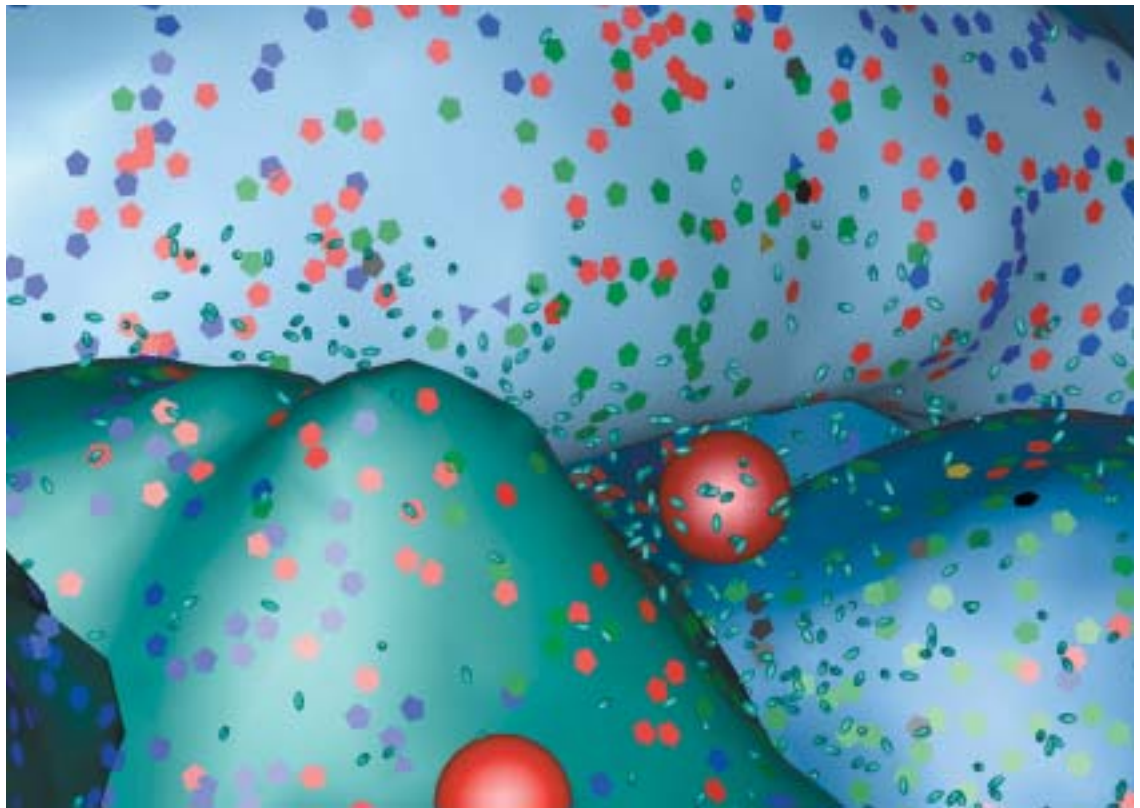


FIGURE 3. CLOSE-UP OF SPINE MAT

The location of two overlying vesicles is represented by the red dots. In this close-up, the neurotransmitter acetylcholine (small green ovals) was released from the upper vesicle 200 microseconds ago. The alpha-7 receptors (pentagons) are more numerous in areas outside of postsynaptic densities, while alpha-3 receptors (triangles) are scarce here (see Figure 2 color code). The distance across the image is 0.5 microns.

control of the pupil,” Berg said.

The MCell model allows “parameter sweep” studies evaluating the neurochemical behavior of complex systems. “In the dynamic modeling,” Bartol said, “we can estimate many parameters that would otherwise be very difficult to determine, like the rate of transmitter-receptor binding at each point.” Martone noted that the study points the way to a new kind of multi-laboratory collaboration that can synthesize the results of very different kinds of study—anatomical, neurochemical, microscopic, and modeling. “It was a perfect problem for NCMIR, because it pushed us to develop new tools,” she said.

“MCell gives us another way to obtain biological answers to fundamental questions,” said Berg. “We had predicted that the receptors would work in certain ways, and the modeling provided a spectacular confirmation.” —MM ▼

REFERENCES

- Shoop, R. D., E. Esquenazi, N. Yamada, M. H. Ellisman, and D. K. Berg (2002): Ultrastructure of a somatic spine mat for nicotinic signaling in neurons, *Journal of Neuroscience* 22: 748-756
- E. Esquenazi, J. S. Coggan, T. M. Bartol, R. D. Shoop, T. J. Sejnowski, M. H. Ellisman, D. K. Berg (2001): Computer simulation of synaptic ultrastructure and microphysiology in the chick ciliary ganglion, *Society for Neuroscience annual meeting poster*.

Exploiting Aging Oil Fields with Advanced Computational Tools

PROJECT LEADER

MARY F. WHEELER
University of Texas, Austin

PARTICIPANTS

STEVEN BRYANT, RYAN MARTINO,
MALGORZATA PESZYNSKA
University of Texas, Austin

ALAN SUSSMAN
University of Maryland

JOEL SALTZ
Ohio State University

MANISH PARASHAR
Rutgers University

The fortunes of Slav immigrant Anthony Lucas and the state of Texas rose together on January 10, 1901, when a 150-foot geyser of oil erupted from a well near Beaumont. However, nothing like the 75,000-barrel-a-day Lucas Gusher spurting from the ground near Borger, Midland, or Beaumont in 2001. Even though half the known petroleum reserves under Texas (and other oil-producing states) remain underground, extracting them has become increasingly difficult. To help, NPACI and its partners in Texas, Maryland, New Jersey, and Ohio are integrating computational tools to recover petroleum more efficiently from the nation's dwindling reserves. "The key questions for oil and gas production companies are plumbing questions," said Steven Bryant, associate director for Microstructure and Pore Scale Modeling at the Center for Subsurface Modeling at the University of Texas, Austin. "What is the rock like between existing wells, and how is the reservoir connected? Two tools we're working on to answer these questions are reservoir simulation and searching quickly through mountains of simulation data."

Plumbing is vital to an industry running out of new discoveries. "There are still virgin territories where companies and countries are searching for new oil fields, but there are fewer and fewer of those areas and they are in increasingly difficult environments," said David Lumley, one of the world's leading authorities on oil field reservoir monitoring technologies and president of 4th Wave Imaging Corp., of Aliso Viejo,

CA. "There is a growing emphasis on becoming more efficient in extracting oil from reservoirs that have

already been discovered."

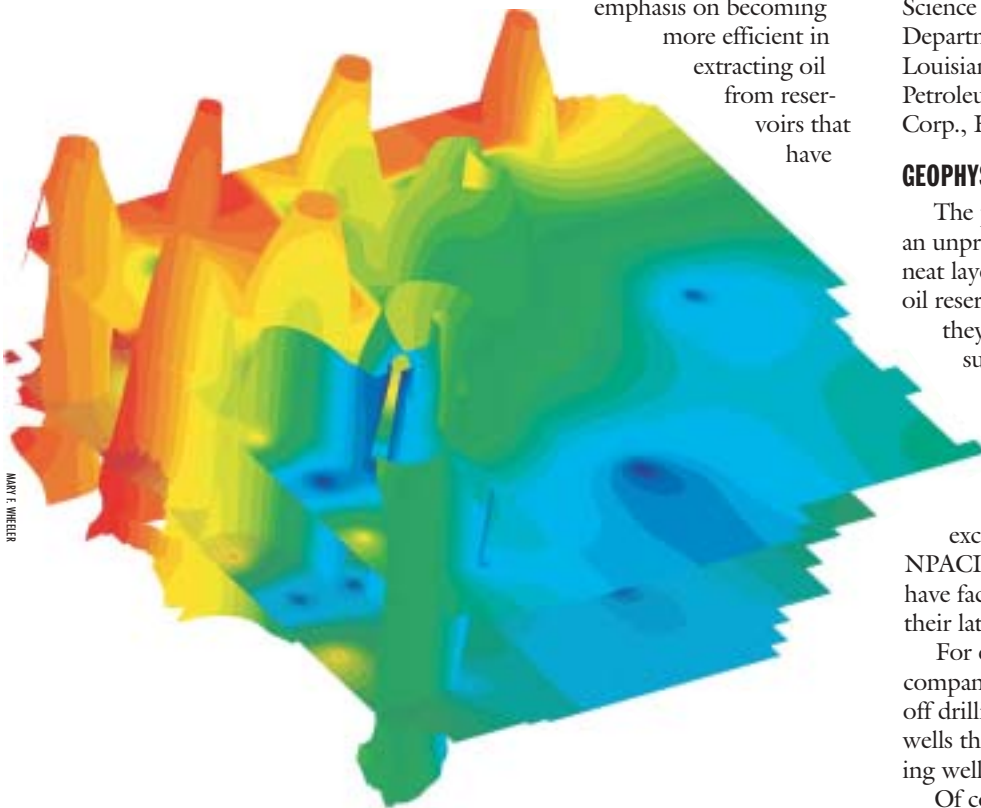
The shift has led to a burgeoning interest in oil-reservoir modeling and data-mining techniques, areas of expertise of an NPACI alpha project, Multi-Component Models for Energy and the Environment. Research by the alpha project's members is being applied to protect groundwater, but the same geophysics applies to oil extraction. The alpha project is supported directly and indirectly by the National Science Foundation, the Department of Energy, the Department of Defense, the states of Texas and Louisiana, and industrial partners Aramco, British Petroleum, ChevronTexaco Corp., ExxonMobile Corp., Halliburton, IBM, and others.

GEOPHYSICAL AND FINANCIAL FORCES

The plumbing of subsurface oil reservoirs is often an unpredictably contorted labyrinth rather than a neat layer cake of porous and nonporous rock. No two oil reservoirs are the same. Researchers model as best they can both the variations in the permeability of subsurface rock and the geophysical forces that control the movement of oil, gas, and water through the fickle medium. An oil company could simply drill more wells in an existing field to produce more petroleum, but the costs of such an approach could far exceed any additional revenue. This is why NPACI partners at the University of Texas, Austin, have factored economic and financial constraints into their latest models.

For example, some simulations show that an oil company in need of sustained income would be better off drilling a few strategically placed water-injection wells that "sweep" oil like a squeegee toward producing wells.

Of course, the oil business is inherently risky, which may explain why some executives try to improve their odds with a mathematical technique called Monte Carlo simulation. The methodology was named for Monte Carlo, Monaco, home of roulette wheels, card games, and other forms of gambling.



RESERVOIR SIMULATION

In this placement, high pressure (red) corresponds to areas with injection (water) wells, and low pressure (blue) to production (oil, gas, and water) wells. In this example, the Center for Subsurface Modeling at the University of Texas conducted simulations of placements of several well patterns over five years of water-flooding in order to determine the best drilling scenario.

Monte Carlo simulation randomly picks variables with uncertain values for particular times or events within a known range. It is a “brute force,” computationally intensive technique used to solve complex problems when no other method is effective. Researchers who would never stack chips on red at a roulette table, routinely use Monte Carlo simulation to investigate everything from particle physics to cancer treatment (see story, page 6).

In oil reservoir simulations, the technique can calculate many scenarios of a given geostatistical model by repeatedly sampling values from the probability distributions for the uncertain variables. The technique uses those values to create subunits, or “cells” of a reservoir. In other words, an oil reservoir (typically hundreds of cubic meters) can be subdivided into hundreds, thousands, or millions of cells and “sampled” in random configurations to describe the reservoir as a whole. The more cells sampled, the better.

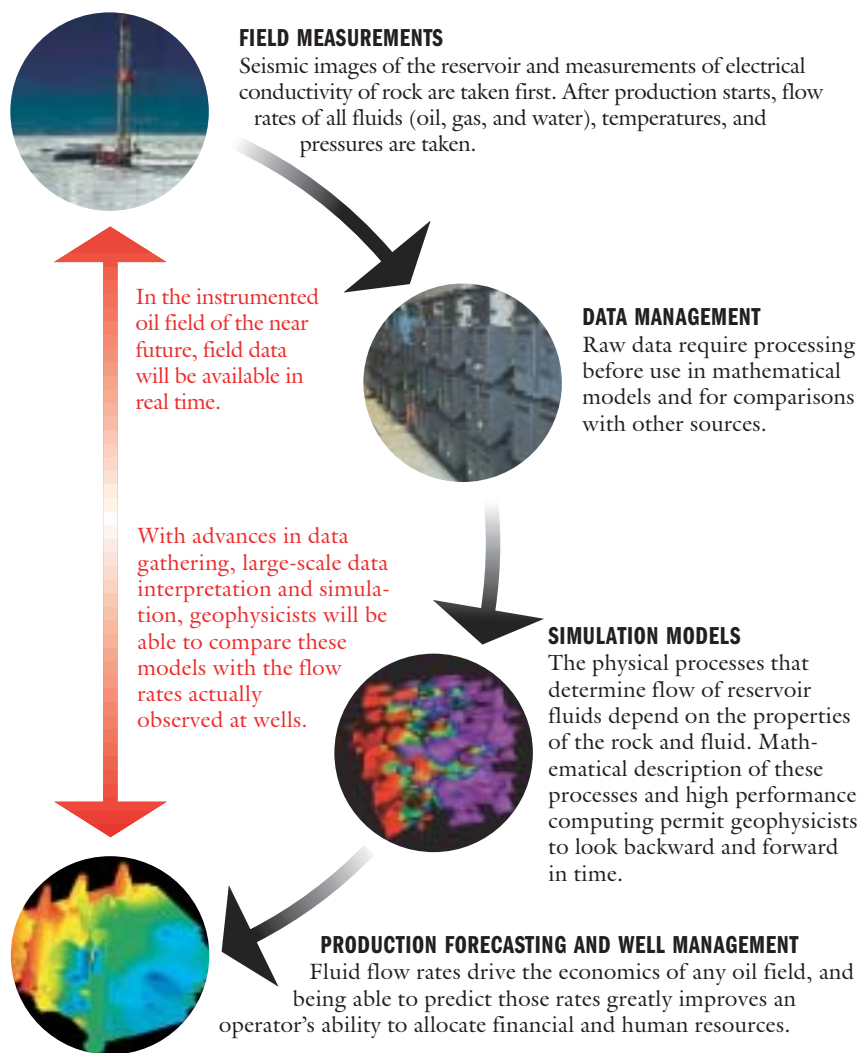
TEXAS LEGEND

Lucas had drilled into Louisiana salt domes, hitting oil in two wells. He was obsessed by the sulfur smelling water near Beaumont, which had been bottled for years to treat various human ailments. Lucas guessed that the bitter fumes were a harbinger of oil, and he used nearly all of his personal funds to buy leases at the springs. The rest is Texas history.

Modern production companies rely on geological assessments, seismographic studies that reveal subsurface rock formations, and past experiences. Some oil executives use back-of-the-envelope calculations, but not arcane-sounding mathematics. But fluctuating oil prices can squeeze profits, prompting more companies to apply computational approaches, particularly in cases where performance distinctions are subtle.

Oil reservoir modeling uses measurements of porosity, quartz content, and other properties of the rock perforated by drilling. Geophysicists also use measurements from nearby outcroppings, and time-lapse seismic data that can reveal the year-to-year flow of subsurface fluid. With those data and random “realizations” made with Monte Carlo simulation, geophysicists create a 3-D model of an oil reservoir. “This approach requires a tight integration between the mathematical sciences, the engineering sciences, and the geological sciences,” said James Jennings, a research scientist at the University of Texas, Austin’s Bureau of Economic Geology.

Mary F. Wheeler, director of the Center for Subsurface Modeling at the University of Texas, Austin, has led a large group of researchers in the development of the Integrated Parallel Accurate Reservoir Simulator (IPARS), which is designed for large-scale parallel simulations of multiphase, multi-component flow and transport in the subsurface. IPARS is taking Monte Carlo simulations to a higher level of sophistication by integrating financial and economic models. “We can look at the total oil production from a reservoir versus the total water injection, which is very helpful because the injection can be costly,” said Malgorzata Peszynska, associate director for



Subsurface Modeling in Wheeler’s group. “We know that the rate at which you produce oil from a reservoir will affect its total production; if you produce very fast you may get a large, immediate financial return, but you may also leave lots of oil and gas behind, which can’t be recovered unless you drill more wells. Our model takes these and many other factors into account. It’s extremely complicated.”

DEMONSTRATING IPARS

During the Supercomputing 2001 conference in Denver, CO, an IPARS demonstration was prepared by members of the Multi-Component Models alpha project, including: Peszynska, Wheeler, Bryant, and Ryan Martino, all with the Center for Subsurface Modeling; Joel Saltz, an Ohio State University professor and co-leader of the alpha project; Tahsin Kurc and Umit Catalyurek, assistant professors at Ohio State; and Alan Sussman with the University of Maryland. The IPARS demonstration rapidly solved a challenging benchmark problem designed by the Society of Petroleum Engineers (SPE).

As a group of SC2001 attendees watched, Saltz queried 1.5 terabytes of IPARS modeling data, which

had been generated earlier and stored at the University of Maryland. The data included the results of 200 realizations of the SPE benchmark problem. As part of the solution of the problem, DataCutter, a middleware infrastructure developed by researchers in the University of Maryland and the Johns Hopkins Pathology Informatics Departments, enabled the sub-setting and user-defined filtering of multi-dimensional data sets stored in archival storage systems across a wide-area network.

“We had 9,000 grid cells, which is not very big—

we’ve made runs with a couple million grid cells,” said Martino, a graduate student in Wheeler’s group. “In the realizations, we used the same well patterns, the same initial conditions. The only difference was how the oil flowed because of the differences in the permeability field.”

Within seconds of each of Saltz’s queries, results moved from a 50-node storage cluster at Maryland to Saltz’s computers at Ohio State where the results were processed with visualization software to create easy-to-understand images, which then flashed onto a display screen at the

Denver Convention Center. “An interesting part of this demonstration was the computation of pockets of bypassed oil,” said Sussman, an assistant professor at Maryland. “An oil company might well want to drill wells at particular places if the concentration of oil is high enough and the oil is present in enough cells for a sufficient length of time.”

Each demonstration worked in seconds—so fast that some conference attendees may not have realized the scale of the computation. “With the advent of PC clusters, anybody could generate gigabytes or terabytes of data, but people who actually do it always say, ‘I’ll never do that again—I don’t have time to look at all of the data,’” said Bryant, chuckling. “What Saltz and his collaborators have created is a tool that allows us to access huge data sets in a cleverly indexed fashion and rapidly query and visualize the results of the queries.”

THE INSTRUMENTED OIL FIELD

The Texas group’s IPARS model is slated to become even more powerful. In October 2001, Wheeler and her colleagues won a three-year, \$1.4 million Information Technology Research grant from the National Science Foundation to design technologies to “better monitor and optimize oil and gas production.” The grant will help Wheeler’s group play a major role in designing “the instrumented oil field of the future.” Other participants in this research include Peszynska, Mrinal Sen, Paul Stoffa, and Clinton Dawson at University of Texas, Austin; Rick Stevens and Mike Papka at Argonne National Laboratory; Manish Parashar at Rutgers University; Saltz, Kurc, and Catalyurek at Ohio State; and Sussman at Maryland.

Wheeler’s group plans to integrate seismic information, well sensors, fiber optics, and remote-control operations. “A major outcome of the proposed research is a computing portal that will enable reservoir simulation and geophysical calculations to interact dynamically with the data and with each other,” said Wheeler. “Since the proposed research is directed toward the general problem of modeling and characterization of the Earth’s subsurface, it has immediate application to other areas, including environmental remediation and storage of hazardous wastes.” The research also has major economic implications for the nation because the total remaining oil reserves in the U.S. may exceed 200 billion barrels—about 70 years worth at the current rate of consumption.

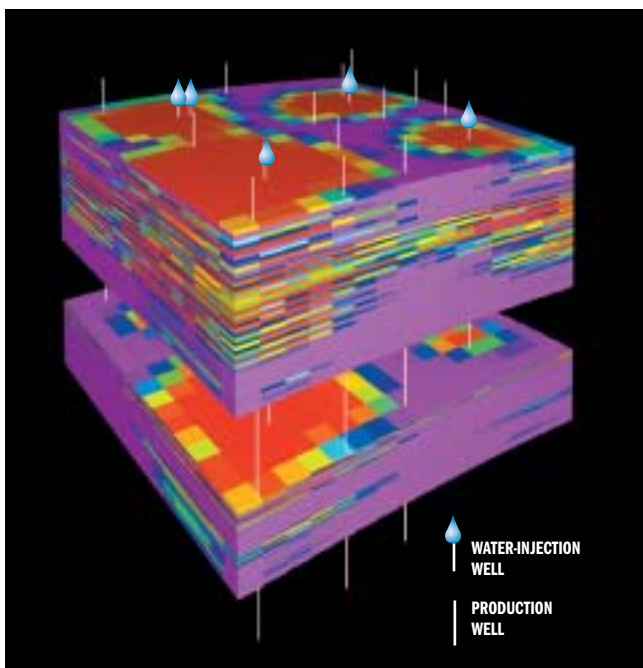
PETROLEUM LEFT BEHIND

Increasingly powerful parallel computers are needed to make the instrumented oil field a realistic possibility with more detailed, more realistic simulations. “If you do enough realizations, you can be more confident of an upper and lower bound,” said Martino.

The simulation estimated the costs and potential benefits of various production scenarios, such as injecting water to sweep oil toward production wells. Such “water flooding” is so common that some reservoirs in Texas now produce 100 barrels of water for every barrel of crude. “If you own an oil company, there are a lot of things you could try, but you can’t estimate which one is going to be best without running a reservoir performance prediction,” said Jennings, one of several academic researchers collaborating with the Center for Subsurface Modeling. “You simply must run some sort of reservoir performance prediction to evaluate candidate proposals in order to determine how best to operate a reservoir.”

For the instrumented oil reservoir of the future, multi-sensor array measurements of pressure, saturation, and other parameters—including seismic images—will travel as a constant bit stream of data to remote computer sites. “All that data will be needed to be computationally processed in order to update reservoir models in real time,” said Lumley. “Continuous monitoring measurements will allow you to test and refine your reservoir model on a continuous basis and thus make it more accurate over time. This is the Holy Grail we are all striving for.” — RG ▼

JAMES SHIMMICKS, U. OF TEXAS BUREAU OF ECONOMIC GEOSCIENCE



PERMEABILITY PATTERNS

These 14-foot-thick slabs of color-coded rock, about 0.5 mile on a side and between 6,000 and 8,000 feet underground, reveal how highly stratified a petroleum reservoir can be. In this case, the highest permeability area is a thin stratum near the top of the bottom layers—red tracer in a “water-flood sweep” has almost completely swept away the purple color—but just below that layer is a low permeability bed with lots of oil left behind by the flooding.

From Proteins to Galaxies: NPACI's Growing Outreach to Museums

The San Diego Natural History Museum lends exhibits of everything from seashells, leaves, and fossilized fish to beetles, butterflies, and bobcats. Three recent exhibits added to the museum's lending rotation are 3-D models of undersea terrain and continental landforms. The computer-designed models were created in SDSC's Design Visualization Lab. The models represent one of many exhibits provided by NPACI and SDSC researchers to museums, from visualizations of proteins to maps of bird habitats. "In addition to supporting the many worthy endeavors of educators and historians, these outreach efforts are helping to inspire the next generation of scientists and engineers while also increasing the public's appreciation of science," said Fran Berman, director of NPACI and SDSC.

HANDS-ON GEOGRAPHY

SDSC's Design Visualization Lab fabricates a wide variety of solid models, which are used as unique forms of scientific visualization. The undersea terrain and continental landform models were made by a Laminated Object Manufacturing (LOM) machine from thousands of sheets of laser-cut paper bonded together. The models look and feel like wood.

The first of the geographical models to go into the lending rotation was a flat, cylindrical-projection map of the world, with vertical elevation exaggerated by a factor of 40. "With only a 12-mile spread between the lowest and highest points on an area 25,000 miles across, it's necessary to exaggerate the data to show the land forms," said Mike Bailey, SDSC senior principal scientist and director of the lab. "Without the exaggeration, the map would look like a slightly bumpy piece of plywood."

Bailey picked a flat projection for ease of handling by students; the spherical model of the Earth would be about the size and weight of a bowling ball. "A flat map also allows students to directly compare mountain ranges across the world," said Bailey, who has contributed other terrain maps to the museum.

The idea to contribute models sprang from a meeting of the master teachers of San Diego City Schools in 1999. After a presentation by Bailey, a teacher suggested that SDSC produce

LOM models and give schools access to them via the specimen library program. A cooperative relationship between SDSC and the museums of San Diego's Balboa Park quickly developed. "I would love it if more teachers requested models that would enhance the lessons they're teaching," said Bailey.

A larger 3-D model from Bailey's lab—a solid topographic map of the continental United States (Figure 1)—is on display in the U.S. Library of Congress in Washington D.C. "The model helps us depict and document new computer-assisted techniques of cartographic production," said James Flatness of the library's Maps Division.

POPULARIZING PROTEINS

"The Art of Science" exhibit, which included images from the Protein Data Bank (PDB), was one of the more popular shows at The Gallery, a space dedicated to art exhibits at Rutgers University. "The exhibit looked at the beauty inherent in the three-dimensional structures of proteins," said Christine Zardecki, research assistant at the PDB and curator of the exhibit. "These structures have a very strong visual component."

Highlighted proteins included those available

PARTICIPANTS

CARTER EMMART, NEIL DE GRASSE TYSON, ERIK WESSELAK
American Museum of Natural History

RALPH RAMIREZ
California Center for Military History

FERNANDO HERNANDEZ
California State University, Los Angeles

CHRIS MIHOS
Case Western Reserve University

LARS HERNQUIST
Harvard-Smithsonian Center for Astrophysics

DONNA COX
National Center for Supercomputing Applications

LEONARD KRISHTALKA, DAVID VIEGLAIS
University of Kansas

MICHAEL J. BAILEY, DRU CLARK, DAVID R. NADEAU, DAVID STOCKWELL
SDSC

www.cmhep.org

biodi.sdsc.edu/www_home.html

dvl.sdsc.edu

www.rcsb.org

www.amnh.org/rose/haydenplanetarium.html



FIGURE 1. TERRAIN MODEL

This 3-D terrain map of part of North America, including the continental United States, is on display in the U.S. Library of Congress in Washington, D.C.

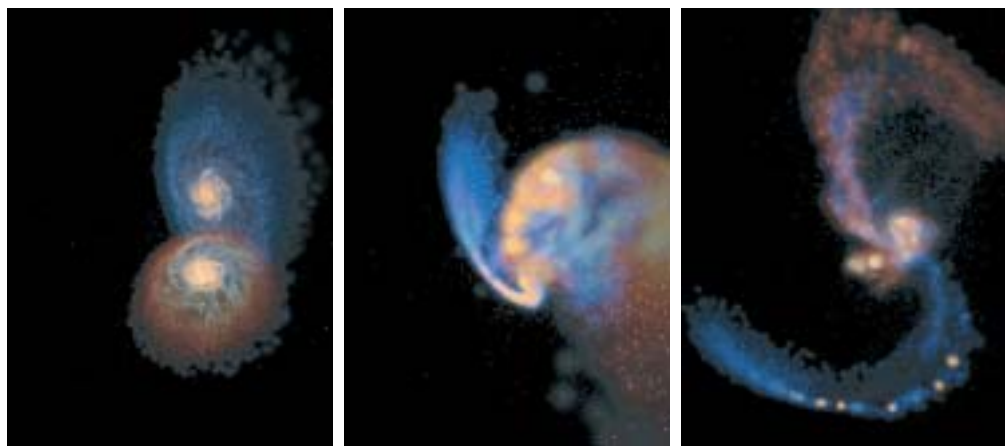


FIGURE 2. GALAXIES IN COLLISION

Frames from the IMAX movie *Cosmic Voyage*, showing the collision and merger of two spiral galaxies.

EDUCATION

from PDB Structure Explorer pages, images of collagen made by Jordi Bella, assistant research scientist at Purdue University, and pictures from the PDB's Molecule of the Month (Figure 3) series by David Goodsell, assistant professor of molecular biology at The Scripps Research Institute.

The PDB is operated by the Research Collaboratory for Structural Bioinformatics—a joint activity of Rutgers University, SDSC, and the National Institute for Standards and Technology. The PDB is supported by the National Science Foundation, the Department of Energy, the National Institute of General Medical Sciences, and the National Library of Medicine.

COSMIC VISTAS

A century's telescope observations of a pair of colliding galaxies amounts to only a "freeze-frame" in a crash sequence that spans hundreds of millions of years.

Simulations are key to understanding such events. For more

than a decade, astrophysicists have been using SDSC's supercomputers to model the forces and physical processes involved in galactic encounters and collisions.

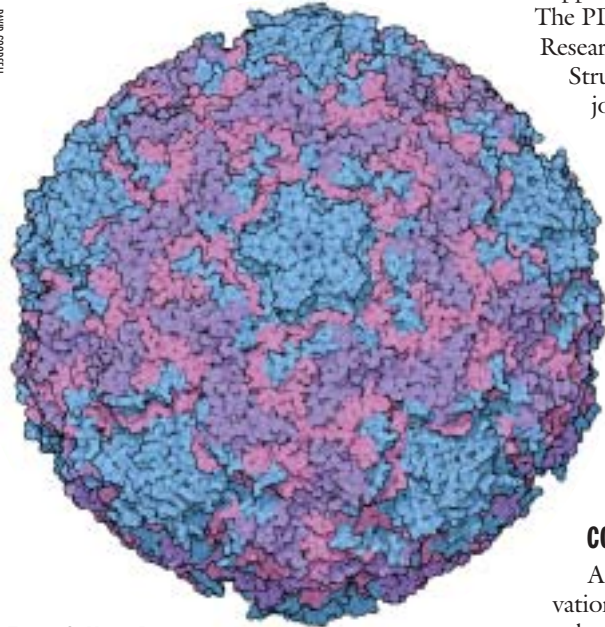


FIGURE 3. VIRUS ART

"The Art of Science" exhibit at Rutgers University displays images from the Protein Data Bank, including this illustration of a rhinovirus.



FIGURE 4. USS CALIFORNIA

Among the materials kept in the repository of The California Military History Museum Education Project is the story of the USS California. The battleship, commissioned in 1921 as the fifth vessel to bear the state's name, was sunk in Hawaii's Pearl Harbor on December 7, 1941.

In the mid-1990s, Lars Hernquist and Chris Mihos, both then at UC Santa Cruz, collaborated with SDSC and computer graphics director Donna Cox at the National Center for Supercomputing Applications (NCSA) to create a high-resolution video sequence of a galaxy encounter (Figure 2). It was the basis of the IMAX movie "Cosmic Voyage," which debuted at the Smithsonian National Air and Space Museum in 1996. One of the first IMAX films to use data-driven supercomputing and the first to feature extended scientific visualization sequences, "Cosmic Voyage" was nominated for an Academy Award. The simulation required 750 CPU hours on a Cray supercomputer at SDSC and generated 65 gigabytes of raw data. Cox and her graphics team at NCSA used software from Pixar Studios to render the visualization in IMAX image format (Figure 4).

One of NPACI's best-known museum collaborations is an all-dome video presentation, made with the help of SDSC and shown at the Hayden Planetarium of the American Museum of Natural History in New York. "Passport to the Universe," the planetarium's premiere production, features a starship voyage through the heart of the Orion Nebula 1,500 light-years away. The strikingly beautiful computer graphics sequence maintains rigorous scientific accuracy, using data from NASA's Hubble Space Telescope and the European Space Agency.

For nearly two years, SDSC researchers worked closely with the Hayden staff to create realistic views of diffuse astronomical objects. "We had to create unique enhancements to the renderer's capabilities to handle astronomical data," said David Nadeau, the leader of SDSC's technical team working with the planetarium. "The VISTA renderer and the tools we've developed for scene construction are applicable to many other problems in scientific visualization, and we believe other researchers and graphic artists will want to use them."

"Our science education message for 'Passport to the Universe' led to our desire to recreate the nearest stellar nursery to our solar system," said Neil de Grasse Tyson, Frederick P. Rose, director of the Hayden Planetarium. "It all came together in the cyberspace of SDSC—without their participation we could have shown some pretty pictures, but they wouldn't have had the impact of the full three-dimensional journey through the volume-rendered nebula."

In March, a new film created with Nadeau's help and NPACI computing resources debuted at the Hayden. Journeying from the depths of Earth's oceans and onward to planets outside our solar system, "The Search for Life: Are We Alone?" depicts how scientists are searching for signs of life beyond our world.

CALIFORNIA MILITARY HISTORY

Although World War II predates grade-school children and most of their parents, they can learn about it from the veterans via the Web. The California Military History Educational Project, an outreach effort by the California Military Museum in Sacramento, will be hosted on SDSC's servers as part of the center's

Science and Technology Outreach program.

“The idea is to have military history, with a California perspective, in a place where it can be accessed worldwide,” said Ralph Ramirez, co-director of the project and deputy commander of the southern region for the California Center for Military History.

The project collects and stores oral histories of survivors, includes a “virtual field trip” on the Web, and provides online lesson plans, teacher resource kits, photo archives, data banks, special collections, and links to everything from California infantry soldiers to the USS California (Figure 4). Downloadable social studies lesson plans were developed in accordance with the California State Content and Performance Standards, and include supplementary materials, teacher rubrics and strategies, discussion questions, and examinations.

THE WORLD ACCORDING TO GARP

Museums traditionally have not coordinated their collections with other institutions or made the data available on the Internet, but NPACI researchers are in the forefront of a revolution. *Science* magazine featured the efforts of NPACI Earth Systems Science partners to put museum collections on line for the taxonomy and systematics communities as driving a “taxonomic revival.” Leading the efforts to integrate the collections data from the world’s natural history museums with online tools are Leonard Krishtalka and David Vieglais at the University of Kansas Natural History Museum and Biodiversity Research Center, along with David Stockwell of SDSC, scientists in the Partnerships for Enhancing Expertise in Taxonomy (PEET) project, and the Kansas Natural History Museum.

The Species Analyst software package developed by Vieglais queries and retrieves collection data from museums that have made their collections Internet-accessible and integrates those data for further analysis. The Species Analyst then forwards the geographic information about a given species to GARP (Genetic Algorithm for Rule-set Production), a program developed by Stockwell that runs on computers at SDSC. GARP maps the known geographic species occurrence data, combines it with environmental data about the general region, and predicts the geographic distribution of the species (Figures 5–7).

GARP has been incorporated into WhyWhere, a Web-based resource that provides quantitative answers to the question “Where is it found and why?” for any species, anywhere on the globe. An example of modern informatics infrastructure, WhyWhere supports researchers by giving them unprecedented access to environmental data, software for analysis and visualization, and high-end computing power, all integrated in an easy-to-use Web interface.

“WhyWhere demonstrates how advanced information infrastructure can help us reach out to a broader range of users and help them develop more reliable predictions of species distribution for regions of the terrestrial and marine environments that have been little studied,” Stockwell said. —CE, MG ▼

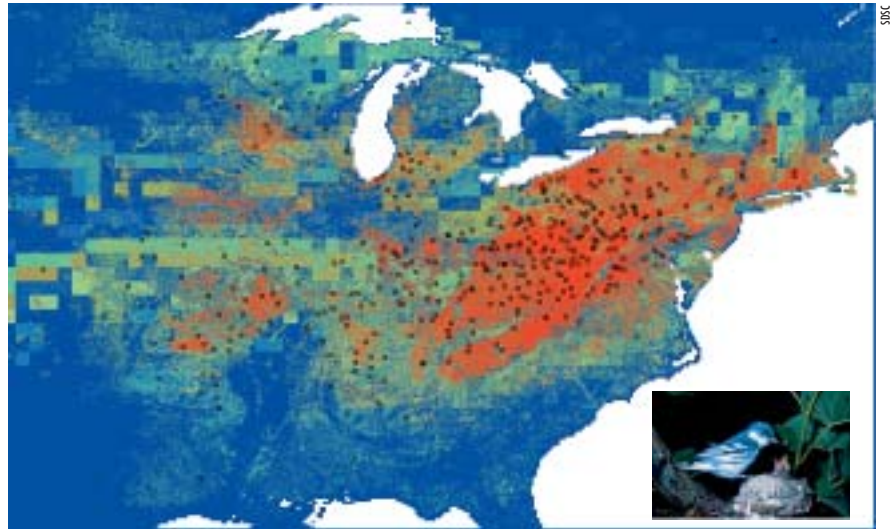


FIGURE 5. FINDING WARBLERS WITH WHYWHERE

Statistical methods identify which environmental variables are the most accurate predictors of such species as the Cerulean Warbler.



FIGURE 6. FISH PREDICTION

The observed species occurrences (black squares) of the tonguefish (inset) coincides with a map of the predicted distribution of the marine flatfish (red regions) by the WhyWhere application.

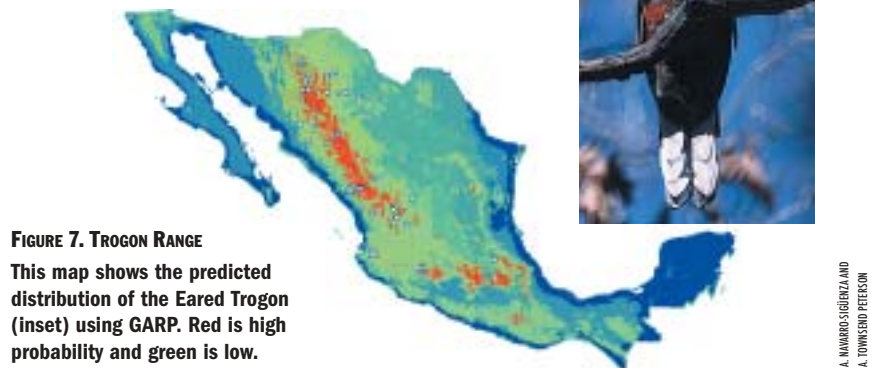


FIGURE 7. TROGON RANGE

This map shows the predicted distribution of the Eared Trogon (inset) using GARP. Red is high probability and green is low.

Strategic Advisors Selected

Fran Berman, director of SDSC and NPACI, has selected six scientists to provide her counsel on a regular basis. "I look forward to working with this distinguished group and to expanding our set of strategic advisors in the future," said Berman. The strategic advisors are:

- Helen Berman (no relation to Fran) is a Board of Governors Professor of Chemistry and a member of the Waksman Institute at Rutgers University.
- Thomas A. DeFanti is a Distinguished Professor of Computer Science at the University of Illinois at Chicago.
- Ed Lazowska, the Bill and Melinda Gates Chair in Computer Science at the University of Washington, is a visiting professor of computer science and engineering at the UCSD Jacobs School of Engineering.
- Paul C. Messina recently retired as director of the Center for Advanced Computing Research and assistant vice president for Scientific Computing at Caltech.
- Cherri M. Pancake is a professor and Intel Faculty Fellow in the Department of Computer Science at Oregon State University.
- Rick Stevens is division director of Mathematics and Computer Science at Argonne National Laboratory and professor of Computer Science at the University of Chicago.
- John C. Wooley is associate vice chancellor for research at UCSD and a former deputy associate director for Biology and Environmental Science and chief of staff and associate director of the Office of Energy Research at the Department of Energy (DOE). (p 6.1)

Richard Moore Appointed NPACI Executive Director

NPACI has selected aerospace research executive Richard L. Moore as its new executive director, succeeding Peter Arzberger, who has assumed new responsibilities in the office of the Vice Chancellor for Research at UC San Diego. Moore, who moved into his office at SDSC in mid-February, will provide strategic direction and

management as well as day-to-day supervision of NPACI operations.

"This is an important year for NPACI and Richard's combination of strategic, management, leadership, and scientific skills will greatly benefit the partnership and help take us into the future," said NPACI and SDSC Director Fran Berman. "We would also like to thank Peter for his service, leadership and outstanding contributions to NPACI." (p 6.2)

Ellisman's Group Wins Award at SC2001

A team of researchers representing NPACI, the National Center for Microscopy and Imaging Research (NCMIR), and UCSD received a top award at the Supercomputing 2001 conference for having the "Best Network-Enabled Application." The group, led by Mark Ellisman, leader of NPACI's Neuroscience thrust area and director of NCMIR at UCSD, transferred a live video stream from a high-energy transmission electron microscope in San Diego to the conference in Denver, CO.

SC2001's Bandwidth Challenge encouraged researchers to push SciNet's 14.5-gigabit wide-area network connection over multiple OC-48 links with demonstrations of leading-edge computer applications. Judges of the competition added a category for the "Best Network-Enabled Application" to recognize Ellisman's "Telescience for Advanced Tomography Applications" project.

The telescience application included a remote-controlled electron microscope that imaged spiny dendrites, the branch-like structures that transmit nerve impulses between cells in the brain. The judges recognized the demonstration for transmitting data in new ways, as opposed to simply moving as much data as possible. "This application was head and shoulders above the others in terms of using a network to integrate and deliver an effective application," said one judge. (p 5.24)

12 Companies Adopt Globus Toolkit as Standard Grid Technology Platform

The open source Globus Toolkit has become the de facto international standard in grid computing as 12 leading computer vendors and software providers in the United States and Japan announced that

they will port and/or support the product.

The University of Southern California's Information Sciences Institute and Argonne National Laboratory developed Globus, which is a focus of research in the NPACI Grid Computing Thrust Area.

Compaq, Cray, SGI, Sun Microsystems, and Veridian in the United States, and Fujitsu, Hitachi, and NEC in Japan, will adopt Globus and develop an optimized form of it for their platforms as their preferred path toward secure, distributed, multi-vendor, grid computing. Three other American companies, Microsoft, IBM, and Entropia also announced stronger commitments to Globus. Platform Computing Inc. also announced that it will collaborate with Globus to provide a commercially supported version of the toolkit.

"Grid computing makes possible new and previously unimagined collaboration and applications," said Bill Blake, vice president of High Performance Technical Computing at Compaq. "Open source standards are crucial to making this vision a reality, and the Globus team is doing essential work in this area." (p 5.24)

PDB Scientists Release Alpha Version of OpenMMS Toolkit

A consortium of scientists managing the Protein Data Bank (PDB)—the world's central online archive containing detailed structural data on proteins, nucleic acids, and protein-nucleic acid complexes—has released a software toolkit that provides more seamless access to this information. The OpenMMS (Open Macromolecular Structure) Toolkit, released by the nonprofit Research Collaboratory for Structural Bioinformatics (RCSB) consortium, allows researchers to use the PDB more efficiently with the Common Object Request Broker Architecture (CORBA) standard.

Researchers rely on the PDB to pursue a wide range of biological investigations, such as drug discovery and design, molecular comparison analysis, and studies of evolution at the molecular level. "The RCSB consortium is committed to providing PDB data in the most effective way possible, and the OpenMMS toolkit does that," said Philip Bourne, co-director of RCSB and director of Integrated Biosciences at SDSC, and professor of Pharmacology at UCSD. (p 5.25)

To view the full Online article, append the issue number to the URL: www.npaci.edu/online/v5.x

Girls Gather for SDSC-Sponsored Science Night

Hundreds of San Diego girls and their families recently spent an evening at the San Ysidro (CA) Community Center working with computers, microscopes, bridge building kits, and soda bottle rockets as part of the SDSC-sponsored Science Enrichment Program. The November 16, 2001, event was intended to give parents of minority girls in grades 4–8 an opportunity to become more aware of their daughters' talents and interest in science.

"Parental support is very important to further girls' interest in science and technology," said Rozeanne Steckler, senior staff scientist at SDSC and a founding director of the Science Enrichment Program. "Some may say that higher education can distance minorities from their culture, and we're showing that that is not always the case."

Family Science Night at San Ysidro, which lies north of the border with Mexico, also gave parents an opportunity to meet their daughters' teachers and participate in hands-on science projects. NSF Director Rita Colwell called the program "an exemplary education initiative." (p 5.25)

Taylor Moving to University of Warwick

Peter Taylor, SDSC's Program Director for Computational Science, has announced plans to take on new roles with Great Britain's University of Warwick. Taylor, who joined SDSC in 1992, has accepted an appointment as Royal Society Wolfson Professor of Chemistry and chief scientist of the Centre for Scientific Computing at the University of Warwick. The Centre was established in late 2001 to provide an environment in which computational scientists, computer scientists, and mathematicians can interact and collaborate.

"This is a great opportunity for Peter," said NPACI and SDSC Director Fran Berman. "Peter has shown outstanding scholarship and great leadership in his work at the center, and we will greatly miss him. We wish Peter much success in his new role and look forward to continued collaboration with him in the future." (p 6.3)

Subscribe to Online or ENVISION: www.npaci.edu/Press/subscriptions.html

SDSC Chooses SAN Solution from Sun

SDSC has selected Sun Microsystems, Inc., to build the center's data-intensive portion of the National Science Foundation's (NSF) Distributed Terascale Facility, called the TeraGrid. SDSC will use Sun's iForce(SM) High Performance Computing Storage Area Network (HPC SAN) solution to allow multiple computers—using a range of operating systems—to seamlessly share data. SDSC chose the Sun platform and its iForce HPC SAN solution for its proven performance and scalability, and will leverage Sun Professional Services for architecture guidance.

The \$53 million TeraGrid project includes four partners: SDSC; the National Center for Supercomputing Applications at the University of Illinois, Urbana-Champaign; Argonne National Laboratory in Argonne, IL; and the Center for Advanced Computing Research at the California Institute of Technology in Pasadena.

SDSC works on some of the world's most challenging biological, environmental, and computing issues. Sun technologies offer the performance needed to support data-intensive requirements. (p 5.25)

NPACI Portal Adds Data Technology

GAMESS, a favorite software tool of computational chemists, is incorporating the advanced file handling and storage management of the SDSC Storage Resource Broker (SRB). The GAMESS Web portal provides seamless access to computational resources across the country, allowing scientists to investigate *ab initio* quantum chemistry without worrying about which machines are working on the problems.

"Integrating the SRB into the GAMESS Web portal will give researchers access to one of the best data management systems available," said Jerry Greenberg, a programmer analyst with the National Biomedical Computation Resource at SDSC.

"Computational chemistry has reached the point where problems generate enormous amounts of data on many different types of machines and the SRB provides an excellent and easy-to-use way to control that information." (p 6.4)

APRIL 2002

- 4-5 MentorNet Partners Forum, Intel Corporation, Santa Clara, CA
- 11-12 27th AAAS Colloquium on Science and Technology Policy, Washington, DC
- 14-18 High Performance Computing Symposium (HPC 2002), San Diego, CA
- 17-19 Internet2 IPV6 Workshop, San Diego Supercomputer Center, La Jolla, CA
- 21-24 International Conference on Computational Science 2002, Amsterdam, The Netherlands
- 22-23 2nd Joint Center for Structural Genomics Annual Meeting, Stanford University, Menlo Park, CA

MAY

- 5-9 CENIC 2002 Annual Conference, San Diego, CA
- 13 Introduction to Parallel Computing with MPI, San Diego Supercomputer Center, La Jolla, CA
- 20-21 Scaling to New Heights: An NPACI workshop exploring the issues in scaling codes to thousands of processors, Pittsburgh Supercomputing Center, Pittsburgh, PA
- 21-24 2nd IEEE International Symposium on Cluster Computing and the Grid, Berlin, Germany

JUNE

- 19-20 Network for Earthquake Engineering Simulation National Workshop, San Diego Supercomputer Center, La Jolla, CA
- 21-25 Global Grid Forum (GGF)-5, Edinburgh, Scotland
- 23-27 IEEE INFOCOM 2002, New York, NY
- 24-27 International Multiconferences in Computer Science, Las Vegas, NV

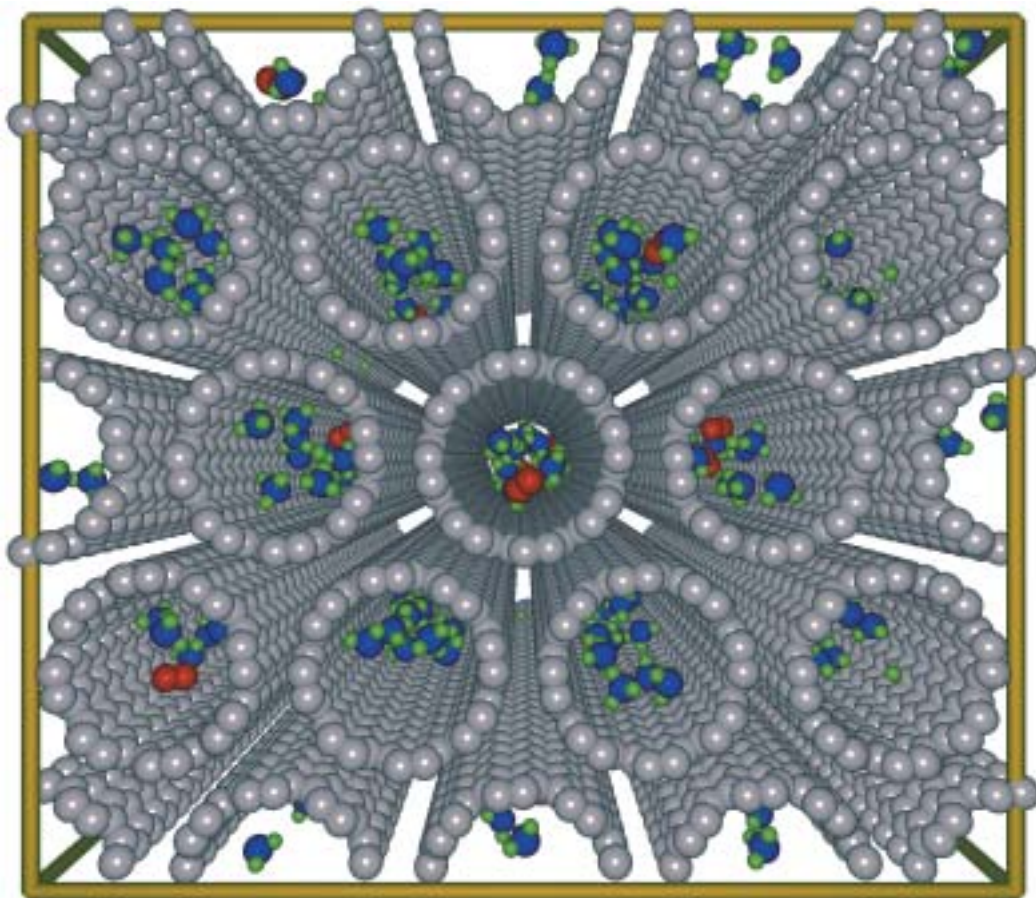
For more information and events, see the SDSC calendar on the Web:

www.sdsc.edu/Calendar

BACK COVER

NPACI PARTNERS

University of California, San Diego/
San Diego Supercomputer Center
California Institute of Technology
University of Texas
University of Michigan
University of California, Berkeley
Baylor College of Medicine
California State University/San Diego State University
University of California, Davis
University of California, Los Angeles
University of California, Santa Barbara
University of California, Santa Cruz
University of Houston/Keck Center
The Johns Hopkins University
University of Kansas
Long-Term Ecological Research Network/
University of New Mexico
University of Maryland
Mississippi State University
Montana State University
New York University
Ohio State University
Oregon State University
Rice University
Rutgers, The State University of New Jersey
The Salk Institute for Biological Studies
The Scripps Research Institute
University of Southern California
Stanford University
University of Tennessee
University of Virginia
Washington University
University of Wisconsin
Center for Advanced Research in Biotechnology
University of Colorado
Jet Propulsion Laboratory
Kitt Peak National Observatory
Lawrence Berkeley National Laboratory/National
Energy Research Scientific Computing Center
Lawrence Livermore National Laboratory
Los Alamos National Laboratory
University of Massachusetts
Michigan State University
Pacific Northwest National Laboratory/Environmental
Molecular Sciences Laboratory
University of Pennsylvania
BioComputing Unit, Centro Nacional de Biotecnología,
Madrid, Spain
Center for Research on Parallel Computation and
Supercomputers, Naples, Italy
Computer Engineering Department, University of
Lecce, Lecce, Italy
Parallel Computing Center, Royal Institute of
Technology, Stockholm, Sweden
University of Queensland, Brisbane, Queensland,
Australia
Research Institute for the Management of Archives and
Libraries, University of Urbino, Italy



CLARIFYING FLUID BEHAVIOR IN PORES

Solids with pores of nanometer dimensions, such as zeolites, activated carbons, and silicas, play prominent roles in chemical processing; however, neither their behaviors nor the influences of surface forces on such small scales are fully understood. Experimental observation of confined fluids is difficult, and porous materials are often poorly characterized.

Researchers at North Carolina State University, led by Keith Gubbins, a chemical engineer and W.H. Clark Distinguished University Professor, have employed NPACI resources to develop molecular models that describe these sieve-like materials, which are used in separations and as catalysts and catalyst supports. This is a simulation snapshot from the ammonia synthesis reaction equilibrium in a bundle of single-walled carbon nanotubes. The cylindrical structures increase the yield of this reaction by 40 percent over the conversion in the bulk gas phase. ▼

SUBSCRIPTIONS

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

Gretchen Rauen, NPACI/SDSC
University of California, San Diego
9500 Gilman Drive, MC 0505
La Jolla, CA 92093-0505

cradmin@sdsc.edu, 858-534-5111

NAME _____

TITLE _____

INSTITUTION _____

ADDRESS _____

CITY, STATE, AND ZIP _____

COUNTRY _____

E-MAIL _____

NPACI NATIONAL PARTNERSHIP FOR ADVANCED
COMPUTATIONAL INFRASTRUCTURE

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

ADDRESS SERVICE REQUESTED

PRESORTED STANDARD
U.S. POSTAGE

PAID

WESTERN GRAPHICS