Who we are
The National Center for Supercomputing Applications (NCSA) opened its doors in January 1986. NCSA earned and maintains an international reputation in high-performance computing and networking and in developing innovative software applications—like NCSA Mosaic, the first readily-available graphical Web browser. Since 1997, NCSA has been the leading-edge site for the National Computational Science Alliance (Alliance), one of two partnerships of the NSF’s Partnerships for Advanced Computational Infrastructure program. The Alliance is a partnership among about 50 academic, government, and industrial organizations from across the United States.

In August 2001, NCSA—as part of a four-institution team—was tapped by the National Science Foundation to build one of the world’s first computational grids and put it to use. It will be the most comprehensive grid yet deployed for open scientific research, spanning the country and providing the backbone from which tomorrow’s global grid can grow. With software developed to make it all work in concert, this TopGrid will offer the fastest unclassified supercomputers as well as an unparalleled array of visualization tools, sensors and instruments, and mass storage devices. These resources will be linked via a network four times faster than today’s fastest.

Major support for NCSA and the Alliance is provided by the National Science Foundation’s Partnerships for Advanced Computational Infrastructure program. Additional funding for NCSA comes from the state of Illinois, the University of Illinois, industrial partners, and other federal agencies.

Cover
An image of the spiral galaxy NGC 6603 produced by the Hubble Space Telescope. NGC 6603 belongs to the massive Centaurus cluster, one of the largest spiral galaxy collections in the nearby universe. NCSA researchers involved with the National Virtual Observatory are currently developing an infrastructure that will allow astronomers to easily access and visualize enormous amounts of archival information about such massive clusters and their sky regions. Photo credit: Jeffrey Newman (University of California at Berkeley) and NASA STScI-PR09-19 May 25, 1999.
The Regional Oceanic Modeling System helps scientists study coastal systems in all their complexity.

University of Kansas researchers model the structure and dynamics of a calcium-binding protein responsible for passing messages throughout the body.

The Regional Oceanic Modeling System helps scientists study coastal systems in all their complexity.

University of Kansas researchers model the structure and dynamics of a calcium-binding protein responsible for passing messages throughout the body.

Nanoparticle research provides fuel for future catalyst designs.

University of Illinois scientists, along with collaborators around the world, draw a bead on the subatomic muon, hoping to better understand the Standard Model of physics.

Argonne National Laboratory's phytoremediation program and a University of Illinois environmental engineering professor use NCSA's D2K to find patterns in large, inconsistent datasets.
ARCHIVING THE UNIVERSE

by

Kathleen Ricker

Using NCSA computing resources, scientists help construct the National Virtual Observatory, a digital framework that will make astronomical data accessible for research on a scale never before achieved.

Abell 1689, one of the most massive clusters in the known universe, in a recent Hubble photograph. Image courtesy of NASA, N. Benitez (JHU), T. Broadhurst (The Hebrew University), H. Ford (JHU), M. Clampin (STScI), G. Hartig (STScI), G. Illingworth (UCO/Lick Observatory), the ACS Science Team, and ESA.
Imagine you're a researcher attempting to gather all the information ever published on a particular subject. You begin with the largest, most central information clearinghouse you know of: the Library of Congress and its catalog of 120 million volumes. Wading through thousands of citations, abstracts, and book indices, most of which initially look promising but eventually turn out to be largely irrelevant, quickly becomes exhausting and overwhelming.

By the time you've finished, the Library of Congress has doubled in size. Suddenly there are now five or six libraries of comparable magnitude through which you'll also have to comb for information, as well as countless smaller, more obscure libraries and archives that you'll have to visit individually, because they, too, might have information important for your research.

This is the problem that confronts many astronomers working with archival data. Generally, once observing astronomers are finished looking for something specific in the data they have collected, they make it available publicly. But astronomical data have a long life. Data that have already served their original research purpose are still a rich source of new information for researchers asking other questions.

"We are just beginning to open up the multiwavelength universe," says Robert Brunner, an assistant professor in the astronomy department at the University of Illinois at Urbana-Champaign and a research scientist at NCSA. "Data are increasing at an exponential rate." In addition to ever-increasing optical and radio astronomical data repositories, he says, vast amounts of high-energy and infrared data are also accumulating.

Searching archival material thoroughly and efficiently—whether you are looking for all existing information about a single object or an entire region of the sky—is fast becoming as daunting as attempting to search the whole known universe.

One way to access archival data is simply to search known repositories, such as the Sloan Digital Sky Survey, or warehouses such as the NASA Extragalactic Database. These archives, accessed via Web browsers, are likely to be frequently bookmarked by researchers. However, "certain kinds of questions are not easily answered by bookmarks," says Ray Plante, a radio astronomer and research programmer at NCSA.

Often, says Plante, valuable information may be found in more obscure but equally useful catalogs or observatories, but the search process can be time-consuming. "Once you discover these sources, where do you find information about your question, and how do you get at that information efficiently? You may find a thousand different text links, but should you visit them all?"

Plante is helping to build the infrastructure for the National Virtual Observatory, a project that will incorporate the Sloan Digital Sky Survey.
A cumulative image of Abell 1656. Symmetric, or elliptical, galaxies tend to be concentrated toward the cluster's center, while asymmetric, or spiral galaxies, tend to be more evenly distributed throughout the cluster.

and many other repositories, both famous and less well known. It will make astronomy research based on archival data a whole lot easier and more efficient by making use of Grid services and NCSA computing resources, particularly the new Linux clusters being installed as part of the TeraGrid.

The value of metadata

The advantages of retrieving astronomical data using a grid model based on interoperable standards are clear: instead of visiting a number of different archives, explains Plante, an astronomer will be able to send a single query to many different servers, which will then respond with the results of that search.

It may sound like a typical Google search, but astronomy researchers’ needs are different from those of the average Internet user. “The NVO’s interfaces must be intelligent enough to have some understanding of astronomical content,” says Plante. For example, a Google-like search for infrared images of a specific quasar would return text documents containing the keyword “infrared,” and possibly a source name. But it probably would not contain any other valuable information, such as the name of a specific kind of telescope, for example, or the location in the sky where the data were originally recorded.

As chair of the NVO Metadata Working Group, Plante is developing a framework that would organize astronomical data in ways that would allow astronomers to zero in on the information that they need. This means cataloging data retrieved from a broad variety of repository sites using a universally machine readable standard, like XML.

XML, short for Extensible Markup Language, is perhaps best known as a hypertext markup language both more universal and customizable than HTML. However, it is also an industry standard for marking up metadata, or “data about data.” This information can document how and when datasets are collected and describes the format in which they are encoded. NVO metadata, for example, may include such information as telescope names, frequency ranges, and sky positions.

Most astronomy data are stored in a format called FITS, developed in the 1980s and used widely throughout the astronomy community. XML won’t replace FITS as the standard format for NVO data. Instead, it will provide important information about data encoded in FITS to speed the query and retrieval process, a development that is becoming more and more necessary as data files increase in magnitude from the gigabyte to the terabyte to the petabyte. “Moving data around physically is one of the hardest problems” facing researchers and NVO architects, says Brunner. “It can be archived, but getting it out of archives easily and quickly is a serious challenge.” The NVO, however, would permit an astronomer to access the metadata describing an image of a half-gigabyte or more and decide whether or not downloading the entire image would be relevant to her research.
A trial run

To give researchers an idea of how they might eventually use the NVO, Plante and his group unveiled a prototype at the American Astronomical Society meeting in January of this year. The prototype focuses on a particular research problem: identifying the relationship between the morphologies, or shapes, of galaxies within a given cluster and the dynamic state of that cluster as a whole.

When a cluster is analyzed, the metadata are cataloged internally using a flexible XML table format called VOTable, designed especially for astronomical data. VOTable allows metadata and data to be stored separately but linked so that researchers who would find the metadata useful are not also required to download enormous chunks of data. The format is able to handle different kinds of data, including XML, FITS, and binary data. VOTable is also scalable and capable of managing varying sizes and quantities of data records.

In addition to the table, image cutouts are created containing single galaxies to help eliminate all but the most relevant information about those galaxies. Using Chimera, which Plante describes as "an Alliance-produced technology for creating and executing computational recipes for data," three parameters for each galaxy are calculated in real time. These parameters include the average surface brightness, the light concentration index, and the asymmetry index (which describes whether a galaxy is spiral or elliptical in shape). Chimera uses Globus technology, also Alliance-produced and integral to the development of grid infrastructure, to manage the data and submit the processing requests to Condor pools (here, computing clusters located at NCSA, the University of Wisconsin at Madison, the University of Southern California, and Fermilab). The resulting analysis produces two files of FITS data and a catalog in VOTable format containing the calculations that users can download and view using a visualization tool such as Aladin or OASIS modified to read VOTable files.

By itself, the information imparted through the prototype analysis is already known well to researchers who study galaxy clusters. But, says Plante, "what we can do is not only make this measurement for one cluster but for many clusters. We can analyze not just hundreds of galaxies but tens of thousands of galaxies. We can build up a lot of statistics that show how the distribution of symmetric galaxies changes with the age of the galaxy, the distance, or many other variables."

Access to enormous quantities of data, using the diverse and powerful computing capabilities of the Grid, could very well catapult researchers in astronomy and other fields into an entirely new way of framing scientific questions, says Brunner. "We're talking here about questions that we never would have thought about asking before because we didn't have access to the resources and technology that we have now."

A prototype of the future NVO Web service. In this example, users choose a cluster from a selected list for analysis. The position in the sky, coordinates, and links to optical and X-ray images of the cluster are automatically returned.

This research is supported by the National Science Foundation and NCSA.

Access Online http://access.ncsa.uiuc.edu/coverstories/nvoatncsa/

For further information, and to see the prototype in action: http://www.us-vo.org/prototypes/galaxymorphology.html

Team members
Robert Brunner
Roy Plante
Harold Ravin
Ramon Williamson
Multiscaling a cell membrane

University of Utah researchers bridge the gap between atomic-scale simulations and whole-cell models.

Two mesoscale simulation snapshots of small vesicles acquired over the course of 800 picoseconds. The mesoscale simulation covers an interval about 20 times longer than the molecular dynamics simulation, permitting simulations that can follow membrane behavior over hundreds of nanoseconds. (Individual molecules cannot be seen in this representation.) The “hatchmark” membrane surface indicates mesoscale interaction and resolution. A mesoscopic “particle” can be found at each intersection of the network. Instead of chemical bonds, mesoscale particles interact via material bonds more akin to larger-scale material properties such as elasticity. The parameters of these particles were calculated from microscopic-level molecular dynamics simulation data.

by

Kathleen M. Wong
According to the old adage, oil and water don't mix. As simple as it sounds, this principle is the construction secret behind one of evolution's most successful structures—the cell membrane. It acts as both guardian and gate, controlling what nutrients and ions, chemical messages, and drugs pass into and out of the cell.

To do its job, the cell membrane must be strong yet flexible, fluid yet robust. It must shrink, bend, and swell like a rubber ball in response to changes in its contents and the external environment. This elasticity allows it to protect the building blocks essential to all organisms—genes and organelles, nutrients and energy—against the hostile world outside.

Understanding how cell membranes accomplish these feats would help scientists design more efficient drug delivery systems, develop accurate models of cellular function, and better understand basic biochemical processes such as ion transport. The small size of cells, however, makes it problematic to observe the behavior of real cell membranes in laboratory experiments. Determining both how easily a membrane expands and contracts and how a membrane's composition affects its elasticity demands pricey equipment and impeccable technique.

The alternative is to run experiments on virtual cells. Using a computer, scientists can observe the behavior of every atom in a cell membrane and its surrounding solvent. Rather than inferring what the molecules are doing from laboratory results, researchers using molecular dynamics simulations can observe all the molecular interactions firsthand. This method allows them to trace membrane behaviors back to the movements of a few constituent molecules. But molecular dynamics modeling comes with its own set of difficulties. Because cell membranes have a relatively large surface area, the number of molecules required to model an entire cell membrane plus its surrounding solvent molecules is unbelievably high. Not even a supercomputer can keep track of all the atoms involved.

Now chemistry professor Gregory Voth and research assistant professor Gary Ayton of the University of Utah have found creative ways to sidestep this problem. Using methods typically applied in statistical mechanics and engineering, they have been able to bridge the gap between atomic-scale simulations and whole-cell membrane models. The results, obtained on Alliance supercomputers, represent views of cell membrane behavior unparalleled in any petri dish.

Oily film

The cell membrane is made up of a double layer of oily molecules known as lipids. In living cells, the lipid bilayer is studded with cholesterol molecules and membrane-spanning proteins such as ion channels. All of these are in constant motion, jostling and bobbing like a flock of rubber ducks in a tub. At the same time, they bump into, repel, and interact with a sea of solvent molecules on both sides of the membrane.
In a watery solution, lipids will naturally assemble into sheets to avoid interacting with the water. The interactions between cell membrane lipids and surrounding water molecules also help stabilize the membrane against the stresses incurred when cells swell and shrink. For this reason, the interactions of every atom in the system can potentially affect the behavior of the entire cell membrane.

Voth and Ayton began their simulations by replicating experiments done in the laboratory on real membranes. On NCSA's Platinum Linux cluster, the researchers built a virtual version of the simplest cell membrane possible: a bilayer made solely of DMPC, a type of lipid molecule. Their atomic-scale model showed the membrane swelling in response to a decrease in exterior osmotic pressure. The flexibility of their virtual membrane turned out to be very similar to previously measured experimental values, confirming their model was quite accurate.

But just as the range of a standard camera lens is too limited to capture the panorama seen by the eye, the molecular dynamics simulation was limited by computing capabilities to just small snippets of membrane. The same constraints meant the simulation followed the membrane for only a few nanoseconds—a period too brief to observe longer-lasting effects playing out in their entirety. To catch any effects that ripple across larger portions of the cell surface, the researchers would have to expand the model's scale.

To do this, Voth and Ayton applied the techniques of statistical mechanics. A very broad and diverse field, statistical mechanics can take impossibly large problems and reduce them to a manageable size. But the cost of reducing the problem is that some information is lost. The trick is determining which information is critical to keep in raw form and which information can be averaged to simplify the model. “If you average too much, you average away everything that's interesting,” Voth says. “That's the art in this technique.” In the case of cell membranes, the most critical information turned out to be the membrane’s flexibility.

**Mesoscale matters**

With flexibility parameters in hand, the researchers then attempted to skip directly to a whole-cell scale. Soon, however, they realized they were missing a critical view of membrane behavior: the mesoscale. Midway between the atomic and cellular scales, the mesoscale is a domain where the kinetic energy of all those jostling and bumping molecules causes measurable perturbations in the membrane. “This mesoscale methodology turned out to be absolutely essential. It wasn’t just icing on the cake,” Voth says.

Though run under the same conditions as the atomic-level simulation, the mesoscale membrane behaved very differently. Instead of the relatively uniform expansion observed in the atomic-scale model, the mesoscale view revealed ripples propagating across the membrane like waves traveling across the surface of the ocean. The higher the temperature, the larger the waves grew. Voth and Ayton found that such long wavelength undulations have a major effect on overall membrane behavior.
Other researchers had predicted the presence of these waves based on previous experiments. Some had even suggested that the phenomenon of ion channels flickering on and off unpredictably is a result of mesoscale bilayer undulations. "Undulations of the membrane may close over the top of an ion channel, causing it to quit working temporarily," Voth says. Membrane undulations could also help explain how ion channels sensitive to mechanical stress operate: when the membrane is stretched or stressed, the channel pore opens and allows molecules to stream in and out. "It's the perfect example of how mechanical motion is coupled to electrical or concentration gradients or flows of ions."

Using the mesoscale data, the researchers were able to construct an accurate model of the entire lipid bilayer. In this whole-cell, or continuum-level, model, individual molecules are no longer discernible. Instead, the model shows how undulations interact to either cancel one another or deform entire segments of the membrane. "This bridging has been one of the first successes, showing how a limited amount of molecular dynamics data can give you enough information to bridge out to a wider range of motions," Voth says.

Confident that their bridging techniques were robust, the researchers turned their attention to more realistic membrane compositions. The membranes of living cells contain much more than just lipids. Complex ion channels, cholesterol molecules, chemical receptors, and more stud its oily domain. When Voth and Aytan added cholesterol molecules to their virtual membranes, they found that the large sterol molecules increased the stiffness of the membrane at all scales and that the pure lipid membrane expanded the most during the simulation runs. For this reason, the concentration of sterols like cholesterol affect the overall elasticity of the membrane far beyond the atomic scale. The researchers reported their bridging results in *Biophysical Journal* in 2002.

The multiscale simulation techniques of Voth and Aytan should be applicable to plenty of other areas in biology. "Many phenomena people want to look at bridge length and time scales over many orders of magnitude, from angstroms to millimeters or femtoseconds to seconds," Voth says. "The grand vision is that by beginning with atomistic information, you will be able to work your way upward to simulating on a computer critical aspects of a living cell."

This research is supported by the National Institutes of Health.

**Access Online**  [http://access.ncsa.uiuc.edu/CoverStories/mesoscale/](http://access.ncsa.uiuc.edu/CoverStories/mesoscale/)

**For further information:**  [http://voth.hec.utah.edu/](http://voth.hec.utah.edu/)

**Team members**

Gary Aytan  
Gregory Voth
The Regional Oceanic Modeling System helps scientists study coastal systems in all their complexity.
Jim McWilliams knows there are many ways to look at an ocean. You can stand on the shoreline and judge surfing conditions at the local beach. You can fly over miles of ocean at 5,000 feet and see where the waves get rough or fly over hundreds of miles at 30,000 feet to see the effects of currents and upwells caused by wind patterns.

McWilliams helped develop a simulation and modeling code that accounts for the many scales that must be considered when studying broad ocean phenomena such as coastal circulation systems. Called the Regional Oceanic Modeling System (ROMS), the code is used by a UCLA-based research group to study the many complex events that are part of a coastal circulation system. These events include the cycling of nutrients locally and regionally and their effect on local ecologies, sediment movement and how it mixes with the waters, and localized circulation patterns caused by surface storms and resulting upwells, coastline terrain, and inflow from coastal rivers.

McWilliams, a professor of oceanography in UCLA’s department of atmospheric and oceanic sciences and Institute of Geophysics and Planetary Physics, is the principal investigator for this research project. The team also includes Nicolas Gruber, assistant professor of atmospheric and oceanic sciences, Keith Stolzenbach, professor of civil and environmental engineering, and seven other research scientists.

“We are not looking at a single question, but a whole system,” says McWilliams. “We try to fit together all the various pieces to see how a regional coastal system works and how regional and global systems influence each other.”

The team began studying coastal regions off the North American West Coast (NAWC) about five years ago and has been running simulations on NCSA’s Origin2000 supercomputer for about three years, using several hundred thousand hours of compute time each year. Most recently, the researchers began generating simulations of circulations, ecosystems, and the geochemistry of the entire Pacific basin and then scaling down those simulations to examine conditions specific to the NAWC. The work promises to shed light on how very large-scale phenomena, such as the periodic warming of the sea surface temperature in the tropical Pacific known as El Niño Southern Oscillation, come to be. It will also show how more localized currents and conditions influence each other.

According to McWilliams, the work tends to be done in pieces. A simulation might model carbon cycles and show their relationship to phytoplankton blooms, for example. Another might illustrate sediment circulation. The team looks at these pieces of the whole and puts them together to understand the bigger picture. The goal is to create an accurate model of the NAWC regional system using both ROMS and data from sensors and sampling.
A matter of scale

ROMS requires masses of computing power partly because it examines phenomena that need to be measured on different time scales. Air temperature and wind speed can change very quickly. Carbon dioxide levels in the water change more slowly and salinity levels slower still. In addition, other coastal phenomena happen on greatly different spatial scales. Local circulation patterns are influenced by localized winds, storms, coastline irregularities, and turbulence in larger-scale currents. These local patterns must be simulated with relatively high resolution, with each measurement covering perhaps 5 kilometers. Larger-scale circulation patterns, however, must be simulated with coarser resolution, say 15 or 20 kilometers per measurement.

For these reasons, ROMS was developed with a time-stepping algorithm that allows for a great variety in the size of steps and with numerical methods that allow high resolution locally and coarser resolution for larger-scale, or mesoscale, phenomena such as storm systems or circulations associated with a particular coastal area.

“The ability to have different spatial resolutions is very important in this kind of modeling because many of these events, such as cycling of biological materials, happen both locally and on a larger scale,” explains McWilliams. “Localized events have an impact on mesoscale events and vice versa.”

If the ROMS research team is able to account for these interactions among local and mesoscale events in their simulations, those simulations will be more meaningful and closer to what actually happens in regional coastal systems, he adds.

From parent to child

For some ROMS models, the researchers use field data to set the initial conditions; for others they create hypothetical conditions. The simulation is then built as a series of grids of different resolution used to measure variables that must be examined on different scales. These nested grids work much like adaptive mesh refinement techniques. The simulation starts with the coarsest grid (the parent grid), which measures a mesoscale phenomenon. Finer “child” grids are then embedded within the parent.

This hierarchical embedding of grids works on many resolution scales over many levels of embedding, explains McWilliams. For studies of the Pacific basin, for example, the researchers start with a very coarse-grain grid that measures at 50-kilometer spatial intervals and works down to a much finer local scale. For local studies around Monterey Bay and the Southern California Bight, the researchers use a three-level embedded grid. The first grid measures at 15-kilometer intervals. Advancing this parent grid by one step determines the boundary conditions for the second level of the grid—in this case a grid with a 5-kilometer resolution. A third grid (the second child grid) looks at phenomena at a 1.5-kilometer resolution. The simulation continues by advancing each child grid forward in space and time enough so that it covers the same space and time as the parent grid. The parent grid is then updated with the information obtained from the child grids to create a more accurate model that measures both large- and small-scale events over time.
For example, to understand how a major winter storm affects the movement of sediment in a local area of the coastal system, a ROMS researcher might look at the regional currents on one level, outflow of river water and debris into the system at a second, finer level, and localized eddies at a third, still finer level. Using an embedded grid, says McWilliams, researchers can see the impact each phenomenon has on the movement of sediment as well as how the phenomena interact and influence each other.

So far, the ROMS research team has run simulations that use as many as four grid levels, with the finest grid measuring at 500-meter intervals in the Santa Monica Bay. The embedded method can handle an unlimited number of grids, but the finer the grid, the more computationally intensive the simulation becomes, explains Patrick Marchesiello, a member of the ROMS research team. For example, it takes 27 timesteps at 1.5-kilometer resolution to equal one timestep at 50-kilometer resolution.

"Embedded gridding can be done on many levels. A very large-scale event like El Niño would be measured on a scale of 100 kilometers," McWilliams says. Although ROMS is a regional modeling system, a similar global ocean modeling system could be developed when the computing power to support it exists.

Putting together the pieces

The UCLA ROMS team has already come up with results that underscore how local, regional, and global phenomena interact and impact each other. The group's simulations show that localized eddies that are highly irregular have more impact on the dilution of pollutants in the regional coastal system than large-scale currents—a finding that could impact pollution control policies, sewage treatment practices, and efforts to control pesticide runoff. Results also indicate that mesoscale currents have the most influence in determining the overall equilibrium of regional coastal systems—another finding that could affect public policy and understanding of global ocean and climate systems.

In the months and years to come, the ROMS team plans to further refine their techniques and perhaps look at other coastal regions or other phenomena within the NAWC region.

"The really big picture means putting all the work and all the pieces together to get an understanding of global systems and global climate change," says McWilliams.

This research is supported by the National Science Foundation, the Office of Naval Research, the National Oceanic and Atmospheric Association, NASA, the NASA Jet Propulsion Laboratory, the California Sea Grant, and the Caltech President’s Fund.

Access Online: http://access.ncsa.uiuc.edu/CoverStories/ROMS/

For further information:
http://www.atmos.ucla.edu/web/research/ocean_coast.html
http://www.atmos.ucla.edu/~patrickm

Team members
Meinte Blaas James C. McWilliams
Xavier Capet John Oram
Harmut Frenzel Kaspar Plattner
Nicolas Gruber Alexander Shchepetkin
Patrick Marchesiello Keith Stolzenbach

The UCLA ROMS team.
MESSAGE RECEIVED

University of Kansas researchers model the structure and dynamics of a calcium-binding protein responsible for passing messages throughout the body.

by

J. William Bell
Schoolhouse Rock, that staple of Saturday-morning TV in the 1970s and '80s, taught us about the body's telegraph line. "You got yours and I got mine," the cartoon ditty explained. It regaled us with a story of the electric impulses that flash through the nervous system. The impulses send signals from your body to your brain and back, according to the song, with your hand sensing that there's a hot stove and your brain giving the order to pull that hand away.

But another critical form of communication within the body went unsung, at least by the Schoolhouse Rockers. Cellular signaling relays information about the surrounding environment to cells, passing ions from the outside to the cell, and serves as a messaging system within individual cells as well, passing those ions along once inside. Various processes rely on different ions—potassium, calcium, and sodium, to name a few.

In many cases, all this ion movement is the work of small, simple proteins. Krzysztof Kuczera, Gouri S. Jas, and Cheng Yang, all of the University of Kansas, are drawing a picture of one of these proteins, calmodulin, that is much clearer than any two-minute animated vignette. Calmodulin, which undergoes a dramatic structural change as it does its work, is a key part of calcium signaling in the cells of all eukaryotes. That is to say, it is at work in all plants, animals, and fungi—anything with cells that have nuclei. The team models calmodulin and its behavior on NCSA's Origin2000 supercomputer.

The three faces of calmodulin

Calmodulin is made up of only 148 amino acids and is highly conserved across species. Calmodulin in humans is identical to that found in many mammals, fish, amphibians, and birds. In other species, its makeup varies by only a dozen or so amino acids. Despite this seeming simplicity, calmodulin is a critical part of any well-oiled organism. It is central to cell functions such as gene regulation, protein synthesis, ion channel function, motility, nuclear division, and development of cells' cytoskeletons. It also mediates large-scale processes such as inflammation, metabolism, and muscle action, meaning that your weight lifting regimen, and even your regularly beating heart, relies on it.

Calmodulin comes in three flavors. When calcium is not attached to the protein, its two globular domains are relatively close together and the interdomain linker is relatively flexible, giving it a collapsed structure. During cellular signaling, two calcium ions bind to each of the domains. In this calcium-loaded form, the domains move farther apart as the linker becomes more elongated and rigid. To complete signal transmission, calcium-loaded calmodulin affixes itself to its target, typically another larger protein. This complex involves the third structural form of calmodulin, in which the two domains grip the target protein.

Traditional experimental studies of calmodulin have shown that calcium fitting itself into calmodulin's domains exposes hydrophobic patches on the protein. These patches are thought to serve as the ports by which calmodulin docks with its target.
Experimental studies alone are a less-than-ideal approach to understanding calmodulin for two reasons. First, the methods introduce their own bias. X-ray crystallography, for example, does not isolate a single protein, instead taking an average snapshot over numerous molecules involved in interactions with their crystal neighbors. NMR methods yield an image of the structure of a single molecule in solution, but the presence of the that solution is not directly accounted for when the image is refined using computational methods. Furthermore, simulations, and some fluorescence experiments, indicate that the protein cannot be well described by a single three-dimensional structure. Instead, these studies hold that the protein should be thought of as a family of structures, differing in domain positions and orientations.

"Everyone believes that traditional imaging methods like NMR and crystallography work, but there's a small chance that things may conceivably get obscured or lost in the shake and bake," Kuczera, an associate professor of chemistry, explains. "The known experimental structures of calmodulin do provide an invaluable starting point for our simulations, though."

**Just like the test tube**

The Kansas team's molecular dynamics simulations, in turn, complement the experimental approaches, giving researchers a new view of calmodulin to consider and compare to existing data. In 2001 and 2002, the team published three papers in the *Journal of Biomolecular Structure & Dynamics*, outlining their most recent findings.

The models discussed in these articles tracked calmodulin in all three of its common states as they shifted about over the course of 4 nanoseconds. Completed in 2001, these models ran for more than six months on eight Origin2000 processors.

"It's a pretty advanced treatment," Kuczera says. Calmodulin itself and a 20-amino-acid chunk of the target, an enzyme known as smooth muscle myosin light chain kinase, were included in the models. They also included the water that surrounds the protein and counterions that neutralize the solution. In total, about 35,000 atoms were simulated in a 92-angstrom domain.

Instead of simply truncating the calculation of the electrostatic energies governing the behavior of the models' elements at the boundaries of the domain, the models used what is known as Ewald summation to approximate the influence that atoms and molecules outside the calculation would have on those inside the boundaries. This approach is critical because both calmodulin and the target muscle kinase fragment are highly charged. The influence of such entities ranges far afield, well beyond the point at which many models truncate their interactions.

"We do our best to correspond to the physiological conditions that you find in your test tube," Kuczera says.

**What they've seen, where they're going**

To date, the structure and dynamics of the Kansas team's simulations have generally agreed with experimental results. The terminal domains of both calcium-free and calcium-loaded calmodulin slowly reorient themselves over time in both the models and the lab. Furthermore, the calcium-loaded calmodulin appears to be less compact and less flexible than the calcium-free form. The same seems to hold true in experimental imaging.
The 4-nanosecond runs have allowed the team to make some important observations, though. Most significant is the presence of large-scale relative domain motions, with translations of up to 10 angstroms and reorientations of up to 70 degrees. These movements imply that calmodulin is an example of a small protein that cannot be described by a single structure. The team also found that the hydrophobic patches thought to be exposed in calcium-loaded calmodulin, which are key to many theories about how the protein attaches to its target, don’t show themselves in the computational models.

“In our models most of the hydrophobic residues remain buried until the target is bound. After target binding, the protein-target interface is mostly hydrophobic, but many of the residues involved fall outside the classic patch region,” says Kuczera.

The dumbbell-shaped calmodulin wraps itself around the target like a bun around a hot dog. Calmodulin’s many negatively charged protein side chains do not appear to interact directly with the positively charged side chains of the target. The discrepancies between simulations and expected outcomes imply that experimentalists’ conventional wisdom on calmodulin’s structural changes and binding habits needs to be reexamined.

More computational study is required as well because, with a 4-nanosecond snippet of modeled motion, it’s impossible to know whether these motions and structures are a fluke or a common, predictable occurrence. Accordingly, the team is expanding and improving their models. They hope to complete 25-nanosecond models for each state within the next year. On eight Origin2000 processors, they complete about 50 picoseconds, or about one-twentieth of a nanosecond, per day. Over the course of the calculations, they expect to use about 150,000 hours of computing time.

“Longer runs allow you to get a sense of the timescale of the domain motions [how the ends of the dumbbell shift over time as they accept calcium ions or bind to the target],” says Kuczera. “Experiment tells us that the critical motions of calmodulin happen in 1 to 3 nanoseconds. Four nanoseconds doesn’t allow you a good statistical sample to prove that that is the case. Only seeing two movements isn’t really enough to say that this is how calmodulin moves in a particular situation or over a given amount of time.”

This research is supported by the American Heart Association, the National Institutes of Health’s Kansas Biomedical Research Infrastructure Network, and the Kansas Center for Advanced Scientific Computing.

Access Online  http://access.ncsa.uiuc.edu/CoverStories/calmodulin/

For further information:
http://oolung.chem.ukans.edu/~kuczerajhome.html

Team members
Gouri S. Jas
Krzysztof Kuczera
Cheng Yang
Nanoparticle research provides fuel for future catalyst designs.

Beneath the surface of pollution control

by

Jennifer Allerson

A false-color picture of the electron density in the surface layer of palladium above a second-layer vacancy. The color spectrum is assigned with red representing high density and blue representing low density. Comparing the hollow site above the vacancy (center) with ordinary FCC hollow sites (top, lower left, and lower right) shows the surprising result that removing an atom results in higher charge density.
The new year offers promising discoveries in a relatively new science, nanotechnology. While sci-fi writers predict nanorobots will be able to invade our cells in order to conquer the world for the forces of evil, physicists in the University of Pennsylvania's chemistry department and Laboratory for Research on the Structure of Matter apply complex nanotechnology simulations to pollution control for the forces of good. They predict they will be able to help clean our environment by continuing research on the properties of metal nanoparticles, also known as nanorods or clusters, which could someday soon lower emission rates in automobiles' catalytic converters. Metal nanoparticles might also prove useful in hydrogen fuel cells, systems proposed by many as a future alternative to combustion engines.

Andrew Rappe, who leads this research project, is no stranger to this realm. Since 1995, he and his team have been conducting computational experiments on nanosurfaces using NCSA resources. His team includes graduate students Valentino Cooper, Sara Mason, and Myung-Won Lee; postdocs Yashar Yourdshahyan, Na Sai, and Ilya Grinberg; and Russel Kauffman, a physics professor at Muhlenberg College in Allentown, Pennsylvania.

"Without NCSA resources, we would not be as effective," Rappe says.

Through their first-principles computations, Rappe's team explores the effects of materials modification on the properties of metal nanoparticles supported on oxide substrates. The main goal is to gain a fundamental understanding of how particle size and composition influence the structural, electronic, and chemical properties of the supported particle. Supported nanoparticles are an important subject because they can be used to gain insight into complex real-world systems, bridging the gap between fundamental research and future catalyst applications. Currently, the team can model the fundamental chemical reactions that a catalytic converter must perform in the presence of a simplified but fairly realistic model of a nanoparticle catalyst. This process allows the team to predict with confidence which particle sizes and compositions could lead to better catalysts than exist today.

**More than just empty space**

In studying supported nanoparticles, the Rappe team's work on vacancies below the surface of palladium has already provided significant results. Transition metals like palladium are the active ingredient in the catalytic converter. Pollutants adsorb, or stick, to the surface of the transition metal. The special properties of the transition metal also lower barriers to reactions that form gases that are not pollutants. The non-polluting gases desorb and leave the converter. In their research project, the team calculated what happens to a palladium surface when an atom is removed from one of the layers, creating a vacancy. The team wanted to know how the properties of the vacancy change with its depth and with the concentration of the vacancies in each layer.

They found that the energy required to create a vacancy increases with the depth of the vacancy. They also found that the atoms surrounding the vacancy relax toward it. Rappe described this phenomenon as being like teeth moving in toward the space in your mouth where a tooth has been lost or pulled. This relaxation is bigger for vacancies near the surface than for vacancies in the bulk of the material.

Subsequently, Rappe and Kauffman found that the bonds between the atoms surrounding the vacancy are strengthened as a result of this inward relaxation. These results are important because vacancies are fundamental defects in materials, and on surfaces, and this project studies how they affect each other. This work demonstrates how changing the structure of a metal changes its properties, opening the door to designing metal structures with the ability to process pollutants more effectively. Detailed results of these calculations have been accepted for publication in *Physical Review B*.

The main computational tool used in this research is a computer code that finds the configuration of electrons that minimizes the total energy of a system of atoms. The code then moves the atoms and minimizes the total energy with respect to their positions. Over the course of the team's studies, they consumed about 500,000 hours of computing time on NCSA's SGI Origin2000.

"These results," said Kauffman, "are both solid and interesting. They help to bridge the gap between the idealized surfaces that theorists study and the real surfaces with defects that are actually used." The results also open the door to further investigation. Questions yet to be answered include: How do vacancies change chemical properties? More specifically, how is the reactivity of a surface affected by the presence of surface and subsurface vacancies? What happens when the material is deformed or strained?

**Possible islands of paradise**

A challenging research project looking at platinum nanoclusters on aluminum oxide (alumina) by Rappe, Yourdshahyan, and Cooper is in a more nascent stage. The goal of this research is to understand the growth process of platinum clusters on supported oxide material and their catalytic behavior for the design of catalytic converters. By using a modified alumina surface for the manufacture of platinum nanoclusters, engineers might someday be able to create a more active catalyst that is less susceptible to poisoning by the sulfur often created by combustion.

Ceramic materials like alumina have many distinguishing properties. They are reliable and widely used as support material in many technologies, from catalytic converters in our cars to dental and bone implants. Another reason for the vast technological importance of alumina is its abundance—only oxygen and silicon are more abundant in the earth's crust. Finally, it's relatively cheap. The fact that it also appears in nature as sapphire and ruby may give it a pricey reputation; however, it is far less expensive than other materials such as platinum.
Studying the adsorption of platinum atoms and growth processes on alumina surfaces presents a challenging case, since the structure of the form of alumina used in automotive catalysts is still a matter of controversy. The Rappe group, therefore, uses the well-known, stable phase of alumina as a model for their study. By mapping the potential energy surface, they are able to identify the most energetically stable adsorption sites of platinum atoms on the alumina surface, knowing that atoms search for areas of lowest potential energy. They study surface modification of alumina to enhance the three-dimensional cluster formation of platinum atoms on the surface. This work demonstrates that each different platinum cluster size has different structural and chemical properties. The Rappe group plans to examine how particular nanoislands of platinum can be selected for desired pollution control reactions.

Yourdshayan uses the SGI Origin2000 for this project as well.

A new approach

In addition to these computational research projects, the Rappe group is experimenting with computational theory itself. Traditionally they have used, and will continue to use, density functional theory (DFT), a computational method that can model between 50 and 100 atoms. This technique has proven itself repeatedly as a powerful, efficient, and accurate method of analyzing and predicting phenomena in metallic, semiconducting, insulating, and molecular systems.

For certain fundamental systems, it is important to achieve higher accuracy than DFT can provide. Therefore, Rappe's team has begun using the Quantum Monte Carlo (QMC) technique, a relatively new computational method. QMC's near-perfect accuracy is impressive, but currently it is less efficient and more computationally expensive than DFT. Despite these drawbacks, the team finds the opportunity to use QMC in conjunction with DFT "exciting and promising."

There is little doubt nanotechnology will become an increasing part of industry and our everyday lives this year and in the future. The next time we read an article about nanotechnology, or more likely, the next time we find ourselves stuck in traffic, we might think not of evil. Instead, we might remember the team of computational scientists in Pennsylvania working for a better environment, one atom at a time.

This research is supported by the U.S. Air Force's Office of Scientific Research, Air Force Materiel Command.

For further information:
http://lorax.chem.upenn.edu

Team members
Valentina Cooper Sara Mason
Ilya Grinberg Andrew Rappe
Russel Kauffman No Sai
Myung-Won Lee Yashar Yourdshahyan
Left: The calculated potential energy surface for platinum atom adsorption on the alumina surface. The color range denotes high to low energy, shifting from red to blue. The most favorable sites are labeled with arrows. Right: Alumina surface top and side view. Possible adsorption sites for platinum atoms are shown with small gray circles, labeled A through N.
Checking the books

by

J. William Bell

University of Illinois scientists, along with collaborators around the world, draw a bead on the subatomic muon, hoping to better understand the Standard Model of physics.
Nature has a way of making a mess of subatomic particles’ balance sheets. “Particles chuck out other particles—sometimes even particles more massive than themselves,” says David Hertzog, a physics professor at the University of Illinois. “Then they grab them back [in an instant], before anyone notices. These particles would be good working at Enron.”

Hertzog is part of a large, international team known as the g-2 (gee minus two) collaboration. Made up of some 70 researchers at 11 institutions, the team is in the business of checking these particles’ books. The behavior of a subatomic particle known as the muon allows the g-2 collaboration to peer into the Standard Model of physics. With the Standard Model and a predilection for number crunching, researchers can predict the workings of three of nature’s fundamental forces—electromagnetism, the strong nuclear interaction that binds atomic nuclei, and the weak nuclear interaction that governs processes like nuclear decay—and the type and behavior of subatomic particles.

Despite an impressive, decades-long track record, the Standard Model is not perfect. Many researchers suspect that there’s more out there, that the physical universe is more complex than even the knotty Standard Model can explain. If researchers can show that the muon acts in a certain way, they can prove that it is being influenced by subatomic particles that the Standard Model does not account for.

“The process [of muons flinging off other particles and then snatch­ ing them back] alters properties of the muon that we can measure,” says Hertzog. “And if those properties don’t match the Standard Model, that implies something. We just don’t know what that is yet.”

After releasing a set of results in 1999, the g-2 collabora­tion caused a stir. Their experimental results were well outside what the Standard Model predicted should have been the case. Motivating theoreticians to go back and re-examine the Standard Model’s every tittle and jot, the results provoked about 230 citations in journal articles. Within a few months, scientists discovered a math error (a single negative sign was off) that made the theoretical numbers hew much more closely to the measured numbers.

Nonetheless, “it’s a very exciting time” in the field, according to Hertzog. The team published new data in August 2002. Relying in part on NCSA’s newest Linux clusters, these results confirm the quality of the team’s earlier experimental data. They differ from theoretical numbers, however, by between 3 and 1.5 standard deviations, depending upon which of the several, well-founded theoretical approaches is used. Needless to say, these data, and the mismatch, give theoreticians and experimentalists more to pore over in their probing of the Standard Model. The team intends to release further data in mid-2003.

Muon watching

The g-2 team runs its experiments at Brookhaven National Laboratory in Upton, New York. Protons fired at a nickel plate in a particle accelerator create billions of muons, which are captured in a storage ring 14 meters in diameter that contains a very uniform magnetic field. Muons, more massive cousins of the electron, orbit in the storage ring. Surviving for only 64 millionths of a second, they decay into two neutrinos and a positron. The positrons hit detectors, designed by researchers at the University of Illinois, on the inside walls of the storage ring.

The g-2 team members at the University of Illinois—including Hertzog, Professor Paul Debevec, Visiting Research Assistant Professor Gerco Onderwater, PhD candidate Fred Gray, and graduate students Chris Polly and Michael Sossong—parse and interpret data about the energy of the positrons hitting the detectors and the times when they strike. Using this information, they reconstruct muons’ properties such as their positions, their energies, and the directions in which they were moving at the moment of decay.
Of most interest to the team is a feature known as the anomalous magnetic moment. Muons are constantly spinning, and their anomalous magnetic moments cause them to precess. Precession is the angular velocity of the spin caused by any external torque. In other words, it's the wobble that develops when you nudge a child's top. With muons, undetected particles do the prodding.

Any time the measured anomalous magnetic moments don't match the theoretical physicists' predictions—and currently they don't—the scientific community can expect the kind of dust-up aroused by the g-2 collaboration's 1999 publication. Theoreticians check their math. Experimentalists analyze new datasets, just in case previous readings were a statistical fluke. It's an involved process, but it's currently the best way to uncover any yet-tinier dominion that might exist underneath our current conception of the subatomic realm.

Raining and swimming

Recording and parsing the detector data are more complicated than simple counting. Over the many months that the g-2 experiment ran at Brookhaven, about 10 billion positrons slammed into the detectors, and those collisions created fast and furious pulses of information.

"The hits often come so close together that they look like a single hit that's slightly too long," says Fred Gray.

It's like trying to count every raindrop that hits during a storm, looking at nothing but the ripples on a pond's surface. To overcome this confounding situation, the Illinois team developed a computer code that makes sense of the deluge of hits. After running the code on NCSA's Linux clusters and fine-tuning it over the last couple of years, Gray is now able to differentiate hits separated by 4 nanoseconds, or a scant 4 billionths of a second. The code further cleans up the mess of data with features that, among other things, discount the noise caused by other errant particles hitting the detectors.

While Gray separates and sorts positrons to discern the muons' anomalous magnetic moments, Michael Sossong studies the dynamics of the muons as they circle in the storage ring. He tries to better comprehend the properties of the storage ring and the influence that the ring's magnetic field has on particle beams that fly through it. "The beam breathes and swims," explains Sossong, changing in shape and trajectory as it moves through the ring. Changes like these can impact the measurement of the muons' precession and must be clearly understood.
To develop a clear picture of the beam, Sossong begins, like Gray, with the positrons born of muon decay. They hit another detector in the storage ring, which consists of 480 “straws.” Each straw contains a high-voltage wire surrounded by argon, carbon dioxide, and ethane gases. The detector records the time at which positrons pass through individual straws. With a piece of ingenious software and some time on an NCSA supercomputer, Sossong can figure the course and speed of the positrons—no small feat considering the variations in the fringe magnetic field where the detector he relies on sits. Also, because the detector does not resolve particles from different parts of the storage ring uniformly, intense calculations are required to adjust the raw data. Reconciling all this information, he can reconstruct the life and times of the muon beam, including its changes in shape and trajectory. By the time he’s finished, he’ll have looked at about 20 million detector events.

How can we help?

The Illinois team has been collecting data like these since 1997 and started out crunching the data on their own 20-machine Linux cluster. The work was slow going, but a bit of good press went a long way toward speeding things up.

“Radha Nandkumar [NCSA’s assistant director of campus and international relations] saw a story about us in the local newspaper. She called us up and asked ‘How can we help?’” says Gray.

NCSA offered massive processing power and storage space, and the Illinois team leapt at the opportunity. They began working on a prototype cluster at NCSA and then moved to the 1,024-processor Platinum cluster. They burned through 25,000 hours of computing time over two months. Individual runs often took weeks to complete. But, because of what Sossong calls “the comparatively enormous power and size of the NCSA Linux cluster,” the computation went almost 10 times faster than it would have on the team’s homegrown cluster.

Even with these intense sessions of calculation and the ever-increasing pile of data, debate about the Standard Model rages on. The g-2 collaboration’s view of the muon, however, just keeps getting more precise.
Solar-driven pumps, data-driven decisions

by Katherine A. Caponi

Argonne National Laboratory's phytoremediation program and a University of Illinois environmental engineering professor use NCSA's D2K to find patterns in large, inconsistent datasets.
Imagine you're walking around in a grove full of newly planted trees. What do you think you might see? Perhaps willowy trunks with promising foliage, signaling nature's eruption into spring? Or maybe you'll see birds and insects returning to an area once marred by landfill waste? Scientist working in the Environmental Assessment Division at Argonne National Laboratory outside Chicago, however, see an enormous network of solar-driven pumps, hard at work extracting groundwater pollutants.

What? Solar-driven pumps? You can't see any network of pumps here—this is just a grove of trees. But can you really see the forest for the trees? Well, look a little closer, and you may find there's more going on right underfoot than you ever could have imagined.

For many years, large amounts of waste were disposed of at Argonne. The materials included anything from nonhazardous solid waste deposited in Argonne's sanitary landfill to chemicals now recognized as harmful that were dumped in a drain in the ground. Eventually, much of that waste permeated the soil and polluted the groundwater.

Recently, Argonne has taken on the cleanup of wastes and contaminated soils at the site, as well as actively containing and getting rid of groundwater pollutants. In addition to a traditional pump-and-treat system, in which contaminated water is pumped out with mechanical wells and treated above ground, workers implemented a more innovative approach to cleanup. In 1999, they planted 800 hybrid poplar and willow trees in deep, plastic-lined boreholes called caissons. The shapes of the plastic caissons were designed to force the tree roots to grow very straight and deep into the ground.

The trees now go about their usual business, taking in water and nutrients from the soil. However, in addition to their normal intake, they also suck up the pollution in the groundwater. Once inside the tree, the chemicals may be stored in the roots, leaves, and stems; converted into harmless gases to be released into the air; or changed into less harmful substances and retained in the tree. This method of using trees or other plants as tools for soil and groundwater cleanup is called phytoremediation and is now practiced at many different sites across the country.

After they planted the trees, the workers no longer had to perform much physical maintenance at the phytoremediation site. Because the trees took over the job of sucking up contaminants, equipment and labor costs were lower. In addition, site workers were no longer forced to work with the dangerous chemicals, making site cleanup a hands-off experience with less potential for liability than more traditional methods of pumping and treating the groundwater.

However, since there is still some residual contamination left in the soil, the Argonne site stewards are obligated to perform continuous, long-term monitoring of the site. They collect and interpret data on water levels, contamination, and rainfall that allow them to see how well the trees are doing their intended job. By finding patterns in that data, they can answer some very basic questions that allow them to cost-effectively monitor the sites. What types of data are needed? From how many physical monitoring points must they take the readings? And how frequently must the readings be taken? Unfortunately, a large mass of unorganized data will not provide the answers to those questions.

To meet their data demands, the Argonne site stewards need a system for analyzing the data and getting answers to their questions. NCSA's Automated Learning Group and the Department of Civil and Environmental Engineering at the University of Illinois at Urbana-Champaign have been working with the Environmental Assessment Division at Argonne to take advantage of data mining and visualization software called Data to Knowledge (D2K), developed at NCSA. The goal is to take a load of different kinds of data and turn those numbers into a plan of action for the ongoing monitoring of groundwater cleanup.

Gathering groundwater data

So where do all these data come from? The scientists begin by collecting water level and contamination information from a series of monitoring wells at the remediation site. They also collect rainfall and other meteorological data from an onsite weather station. Plotting the wells' physical locations, the researchers build a map of groundwater fluctuations.

There are a couple of different ways that the data can be collected from the wells. One option is for scientists in the field to do this by hand, physically going out into the grove of trees and performing measurements...
and chemical tests. Another option is automated monitoring using sensors at the site that take numerous readings on a predetermined schedule. The sensors reduce the amount of human labor needed to take samples and regulate when samples are taken from each well.

The different methods of data collection, however, cause some common problems, not only at Argonne but also at many other waste management sites across the nation. Jack Ditmars of the Environmental Assessment Division at Argonne, says, “Environmental data tend to be kind of messy. Data are not always collected as regularly as you might think—that provides a real challenge in making sense of the readings.”

According to Ditmars, in the recorded data for any given remediation site, you may find continuous water-level readings for a couple of months that were taken by sensors and then very sporadic readings taken by hand for the rest of the year. The data collected for contaminants, rainfall, and various other factors are just as motley. In addition, there are many wells at each site, and readings can be taken from any combination of those wells at a variety of time intervals. Ditmars says, “While Argonne is a relatively small site with only tens of wells, there are other larger sites around the country with hundreds or thousands of wells. This can result in an unmanageable number of combinations of wells at each site from which you could potentially take readings.”

Without a strategic plan for monitoring, that kind of data collection can become extremely time consuming and incredibly costly. The long-term stewardship of multiple cleanup efforts going on across the country creates a driving need to find more economical means of handling the modeling and analysis of problems like groundwater cleanup.

**Data to decision-making**

Developing that plan can be quite the daunting task for site stewards and researchers. The good news is that there are researchers out there who specialize in optimizing monitoring schemes for environmental and civil engineering. Associate Professor Barbara Minsker, Professor Albert Valocchi, and their graduate students at the University of Illinois have been hard at work linking the real-life groundwater data provided by the Argonne Environmental Assessment Division with the tools supplied by NCSA’s D2K creators.

D2K provides a toolkit for researchers who have massive amounts of data that need to be organized and modeled. It supplies the user with data modules that can be plugged in and removed. That kind of toolkit allows the researcher to see how manipulating the variables changes the resulting model.

In the particular case of the groundwater monitoring problems, Minsker, Valocchi, and their students are working with NCSA’s Automated Learning Group to find out how Argonne’s historical groundwater data compares to the current data being taken both by hand and by sensors. Using D2K, they create models of groundwater levels throughout the site.

These models provide extremely useful information because they allow the Argonne site stewards to see exactly how closely they must monitor the site. The stewards can compare the D2K-generated models with incoming data taken directly from the field. Using those comparisons, they can determine how accurately they can estimate groundwater status using historical data. From there, they can make more informed decisions regarding how they will choose to monitor their wells.

Minsker says that using D2K works well with the phytoremediation project because “it’s such a flexible tool.” Because D2K provides the building block-like modules, the team is able to create models that show what happens when different kinds of data are combined in different ways, including innovative hierarchical combinations of models to accommodate data at vastly different scales. For example, the hand-collected data are sparse but more representative of different conditions across the site, while the sensor data are extensive but are collected at far fewer locations. If both types of data are combined in a single model, the sensor data overwhelm the hand-collected data and the model’s accuracy suffers.

The team’s work with D2K and the Automated Learning Group provides a way for Argonne to make informed decisions about the monitoring of their cleanup sites. On the larger scale, the project serves as a testbed for decreasing the cost of long-term environmental cleanup and stewardship at sites across the nation.

This research is supported by the Department of Energy’s Environmental Management Science Program, Argonne National Laboratory, NCSA, and the Environmental Engineering and Science Program at University of Illinois at Urbana-Champaign’s Walter E. Deuchler Fellowship.

### Access Online
http://access.ncsa.uiuc.edu/CoverStories/phytoremediation/

### For further information:
http://www.cee.uiuc.edu/emsa/
http://www.ncsa.uiuc.edu/Divisions/DMV/ALG/
http://www.ead.anl.gov/

### Team members
- Jack Ditmars
- John Quinn
- William Michael
- David Tcheng
- Barbara Minsker
- Albert Valocchi

A student from Minsker’s team, William Michael, at Argonne National Laboratory’s phytoremediation site.
A severe storm simulation, showcasing a prototype fault-tolerant meteorology code, debuted at SC2002, the annual supercomputing conference held in Baltimore, Maryland, in November 2002. The model system was based on the NCOMMAS storm code developed by Lou Wicker and a parallel model driver developed at the University of Minnesota’s Laboratory for Computational Science and Engineering (LCSE). Collaborators on the project include Sarah Anderson, David Porter, and Paul Woodward, of LCSE; Lou Wicker, of the National Oceanic and Atmospheric Administrations’ National Severe Storm Laboratory; Crystal Shaw, Mark Straka, and Bob Wilhelmson, of NCSA; and Rob Fowler, Guohua Jin, and John Mellow-Crummey, of Rice University.

The simulation ran on a 32-processor Itanium 2 cluster built by NCSA for the SC2002 conference. The visualizations shown above and at the conference are from a similar model system run and show a storm 2,860 seconds after cloud initialization. They were created with LCSE’s Hierarchical Volume Renderer.

The left-hand upper image shows the storm clouds. The central upper image shows the storm’s vorticity. The right-hand upper image shows the storm’s vertical velocity. And the large, lower image shows the vertical component of the storm’s vorticity.