From Supercomputers To The Grid
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>From Supercomputing to the Grid</td>
<td>01</td>
</tr>
<tr>
<td>Taking the Pulse of a Red Giant</td>
<td>02</td>
</tr>
<tr>
<td>Modeling the turbulent interior of these massive stars is staggeringly complex. But these researchers have generated the largest simulation to date. Paul Woodward and David Porter, University of Minnesota</td>
<td></td>
</tr>
<tr>
<td>Star Light, Star Bright</td>
<td>06</td>
</tr>
<tr>
<td>Before there was a Milky Way galaxy, planet Earth, or solar system, out of cosmic darkness came the first star in the universe. Jeremiah Ostriker, Princeton University, and Nickolay Gnedin, University of Colorado at Boulder</td>
<td></td>
</tr>
<tr>
<td>Colliding Neutron Stars</td>
<td>10</td>
</tr>
<tr>
<td>Newtonian physics are sufficient for modeling the first few million years of the spiraling infall of two orbiting neutron stars, but for the last milliseconds, scientists need relativity. Doug Swesty, NCSS, Paul Saylor, University of Illinois at Urbana-Champaign, and Wei-Mo Suen, Washington University in St. Louis</td>
<td></td>
</tr>
<tr>
<td>Quarks Come un-GLUon-ed</td>
<td>14</td>
</tr>
<tr>
<td>Scientists believe that in the first milliseconds of history, the universe was nothing more than a formless plasma of quarks and gluons. Computer simulations may help prove them right. Robert L. Sugar, University of California, Santa Barbara</td>
<td></td>
</tr>
<tr>
<td>A New View on Atoms</td>
<td>17</td>
</tr>
<tr>
<td>Understanding why materials crack is easier when the fractures are simulated in 3D. Diana Farkas, Virginia Tech</td>
<td></td>
</tr>
<tr>
<td>Stressed Out Metals</td>
<td>20</td>
</tr>
<tr>
<td>By adding just the right amount of impurity to a metal alloy, mechanical engineers can design materials that will bend but not crack under stress. Arthur Freeman, Northwestern University</td>
<td></td>
</tr>
<tr>
<td>New Matter: Friction</td>
<td>24</td>
</tr>
<tr>
<td>Free Hydrogen</td>
<td></td>
</tr>
<tr>
<td>Helium isn't the only superfluid around. With the addition of a few impurities, liquid hydrogen also will forgo its natural tendency to freeze solid at low temperatures and enter the strange, frictionless state of matter called superfluidity. David Ceperley, University of Illinois at Urbana-Champaign</td>
<td></td>
</tr>
<tr>
<td>Stretchy Molecules in Low Drag Solutions</td>
<td>28</td>
</tr>
<tr>
<td>Common sense tells us that thick fluids should flow more slowly than thin ones. Experience with polymer additives tells us otherwise. Antony Beis, University of Delaware</td>
<td></td>
</tr>
<tr>
<td>Code that Keeps Blood Flowing</td>
<td>32</td>
</tr>
<tr>
<td>Half of all arterial grafts fail within 10 years, often because they've become as blocked as the arteries they replaced. The problem may be the angle at which the grafts are attached. George Karniadakis, Brown University</td>
<td></td>
</tr>
<tr>
<td>City of Light</td>
<td>36</td>
</tr>
<tr>
<td>Tiny structures called photonic crystals, which trap and control light with extraordinary efficiency, may herald a new generation of miniature optical devices. John Joannopoulos, Massachusetts Institute of Technology</td>
<td></td>
</tr>
<tr>
<td>Giving Electrons a Nudge — One at a Time</td>
<td>40</td>
</tr>
<tr>
<td>Plants convert sunlight to energy by moving single electrons from one point to another. Scientists would like to replicate this simple, yet amazing feat. Ralph A. Wheeler, University of Oklahoma</td>
<td></td>
</tr>
<tr>
<td>Industrial Strength</td>
<td>44</td>
</tr>
<tr>
<td>Antibodies</td>
<td></td>
</tr>
<tr>
<td>Antibodies can bind to just about anything. Surprisingly, that makes them a potential source of new catalysts. Kendall Houk, University of California, Los Angeles</td>
<td></td>
</tr>
<tr>
<td>Blueprints of the Protein Factory</td>
<td>48</td>
</tr>
<tr>
<td>As in all organisms, our proteins are assembled in intracellular factories called ribosomes. A new imaging method is producing some of the first interior views of this factory as well as details of how the nascent proteins are exported. Joachim Frank, Wadsworth Center</td>
<td></td>
</tr>
<tr>
<td>Striking Gold in Mountains of Data</td>
<td>52</td>
</tr>
<tr>
<td>The rush is on. With data mining techniques developed at NCSS, companies are discovering valuable nuggets of information buried in mountains of data. Michael Wege, NCSS</td>
<td></td>
</tr>
<tr>
<td>Expanding Alliance resources</td>
<td>58</td>
</tr>
<tr>
<td>to keep pace with user demand</td>
<td></td>
</tr>
<tr>
<td>Science on the Grid</td>
<td>60</td>
</tr>
<tr>
<td>What, exactly, is the Grid?</td>
<td></td>
</tr>
<tr>
<td>Alliance Academic Research Contacts</td>
<td>63</td>
</tr>
</tbody>
</table>
| Access (ISSN 1064-9409) is published by the National Computational Science Alliance (Alliance), a partnership among more than 50 academic, government, and industrial organizations from across the United States to prototype an advanced computational infrastructure for the 21st Century. The Alliance is led by National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign with major support from 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. Copyright © 1998 Board of Trustees of the University of Illinois. All brand and product names are trademarks or registered trademarks of their respective holders. Printed with soybean ink on recycled paper.
15 years ago, access to supercomputers was as rare as comet sightings. Only a handful of supercomputers were available to academic researchers. Of these, most of the machines were secluded in laboratories operated by the Department of Energy or National Weather Service. None existed at U.S. academic institutions.

Although academic researchers still cannot obtain as much time on supercomputers as they need, the availability of supercomputing cycles has changed dramatically during this past decade. Due in large part to a concerted effort by the National Science Foundation to provide academicians with access to supercomputers, today thousands of U.S. researchers are regularly computing on these high-powered machines. And, consequently, they are examining phenomena at a degree of detail that seemed impossible only a few years ago.

Sometimes these researchers' accomplishments read like the sports pages: the largest simulation, the most computing cycles, the longest run. But behind these statistics are true breakthroughs in science and engineering — often with profound implications for helping the nation maintain a global competitive edge. Researchers are designing new materials, atom by atom, and deciphering the mechanisms by which cells produce proteins. They are refining catalytic reactions that are the core of many industrial processes and are designing medical devices that will extend lives. They are discovering subtle patterns in data that are helping companies save millions of dollars by marketing their products more effectively, detecting fraud, and building safer, longer-lasting machines.

Profiled here are examples of advanced scientific and engineering projects conducted using the Alliance's supercomputers. These projects are important in and of themselves, yet they have a dual purpose for the Alliance. Besides furthering knowledge, these projects are serving as crucial application drivers for the National Technology Grid (Grid) — the computational infrastructure the Alliance is prototyping. The Grid will integrate high-performance computers, advanced visualization environments, mass storage devices, and massive databases via high-speed networks to form the most powerful problem-solving environment ever assembled. This distributed computing environment will be accessible anywhere, any time by ubiquitous Web browsers.

Research projects, like those described in this publication, are establishing the performance requirements for this Grid because the problems they address are of such complexity that in trying to solve them, the researchers are pushing the limits of existing computing technology. By attaining the performance requirements that these kinds of applications are seeking today, the Grid will provide a glimpse of the capabilities of the Internet of tomorrow.

For now, research applications will continue to rely almost exclusively on the capabilities of individual supercomputers. But in coming years, as new Grid technologies emerge, more applications will employ distributed computing systems. As they do, the potential for extraordinary research will take another leap forward.
Taking the PULSE of a Red Giant

Modeling the turbulent interior of these massive stars is staggeringly complex. But these researchers have generated the largest simulation to date.

by Holly Korab
With hundreds of billions of stars in each of the universe's hundreds of billions of galaxies, astronomers have plenty to study. Using telescopes and satellites they can measure the color, location, and changes in luminosity of many of the 200 billion stars within the Milky Way galaxy alone. However, they can see only one of these — the Sun — up close. Even then, they can't look below its outermost layer.

Since astronomers can neither zoom in on more distant stars nor peek under the cover of even the nearest one, how do they learn what powers them? If they are Paul Woodward and David Porter, they create their own.

These Alliance astrophysicists, nearly as well known for setting computing records as for analyzing stellar fluid dynamics, practice an artful brand of computing that is extending the limits at which stars can be accurately simulated for ever greater realism. Recently they peeked inside a red giant — a dying star. Along with other members of the Laboratory for Computational Science and Engineering at the University of Minnesota, they generated a 3D simulation with such detail that they could watch the star pulsate.

One of the things they saw was a region of superhot gases, encompassing nearly the entire star, flowing asymmetrically across the surface. Gas flowed outward from the hotter side of the star to the cooler — giving off heat along the way — then sank, forming a funnel that reheated the gas as it passed the hot, stellar core. If this pattern is confirmed, it may explain differences in illumination within these stars — a finding that is important to astronomers who rely on these "standard candles" for mapping distances in the universe. It will also help scientists know what to expect from our own Sun when it, too, becomes a red giant.

**Shrinking and expanding with age**

Stars, like people, are born and die. They begin life as a gleam — interstellar clouds of dust and gas that collapse under the pull of gravity. The extreme temperatures and pressures of these coalescing clouds fuse hydrogen nuclei into helium, releasing energy — in the form of photons — that slowly works its way to the surface where it escapes as light. These nuclear reactions power the stars at the same time that they halt the stars' collapse by counterbalancing gravity. In this state of quasi-equilibrium, stars burn hydrogen contentedly for millions to billions of years. The Sun, which is at the midpoint of its 10-billion-year life, will exhaust its hydrogen fuel in another 2 to 5 billion years. When it does, it will begin its dramatic transformation into a red giant. With gravity no longer held at bay, the core will contract violently — shrinking from the volume of 300,000 Earths to a few radii of Earth. At the same time, its envelope — the cooler, less dense layers of gas that surround the core — will balloon to 100 times its original size.
Woodward and Porter’s interest in these aging stars is their vigorous convective envelopes. In younger stars, heat from nuclear fusion finds its way to the surface primarily through radiative diffusion. Only during the cooler last third of their journey is this heat transported by convective motions of stellar gas. In red giants, almost the entire envelope is convective — its layers roiling like water boiling on a stove.

Modeling the convective movement of energy through this system is staggeringly complex. Slight changes in temperature, pressure, and density in one part of the flow may lead to turbulent vortices elsewhere; or, as in Woodward and Porter’s model, small surface eddies may lead to large convective flows. Thus, the model must capture large-scale effects and great detail — without the computations overwhelming the computing system. To make their computation possible — the largest, to date, for a red giant — the team simplified their model through imaginative code work, some convenient physical approximations, and close attention to the capabilities of NCSA’s SGI CRAY Origin2000.

**Penciling in the details**

The numerical approach adopted by Woodward and Porter models the simplest equations of gas dynamics — satisfying the laws of conservation of mass, momentum, and energy while ignoring viscosity, magnetic fields, and ionization. Their model also approximates the core — pumping heat steadily into the envelope rather than computing the internal nuclear fusion.

By approximating these detailed physical effects, Woodward and Porter are able to focus on the hydrodynamics, explicitly following the movement of gas and sound through the star, down to the smallest scale represented on their grid. They chose sound to follow because it is the fastest signal in red giants — beating out photons, which bounce randomly from one atom to another, scattering many times before they make much progress. Also, says Porter, “resolving sound in detail resolves convection because in red giants the velocities of the two are nearly equivalent.”

Following sound explicitly requires small time-steps — so small that a sound wave cannot cross more than one grid cell per time-step. Small time-steps require many computational cells — in this case, 134 million (the mesh cube has 512 grids on a side, or 512^3, to advance the system by 7,000 time-steps. Such a large number of computational cells is manageable, though, because the calculations associated with each time-step are completely local. “Our model can update the state of the fluid at one grid point without needing to know what is happening at other, distant grid points,” says Woodward.
Exploiting this localized model, the code divides the model into "pencils," or groupings of grid cells — eight on a side and 512 long — that can be updated independently of each other and computed in parallel on the 512-node Origin2000. A given processor retrieves into its cache memory a pencil from shared memory, updates the pencil, returns it to global shared memory, then moves on to the next pencil in the queue. If one chunk of grid takes longer to update than another, the next processor in line simply moves on to the next pencil, no longer tied to a particular computation because of its access to a global distributed shared memory. All the updates are overseen by a self-scheduling master loop that ensures each 1D sweep is completed before it executes the next.

The ability of their code to appropriate computing time where needed is what enabled Woodward and Porter to treat the star's hydrodynamically active surface more realistically without compromising the size of the computation nor resorting to irregular meshes or curvilinear coordinates — the very complicated means by which irregular surfaces are usually computed. Their simulation is the first to model irregular problems — like the red giant's pulsating surface — on regular 3D Cartesian grids.

"We got the best of both worlds — irregular boundaries but a regular internal mesh," says Porter. "And these spherical systems are the first examples of what we can do with this. The surface is actually free to move through the mesh. It can deform and change its topology. It can splash. It allows for a wider range of fluid dynamical behavior."

Woodward concurs. "With the development of distributed shared memory machines like the Origin2000, we saw there was an opportunity to go beyond regular calculations and do irregular types, which opens up a whole new class of applications, of which red giants is one example."

Although these new applications won't be limited to stellar studies, that's what Woodward and Porter are pursuing. Not satisfied with a partial simulation of a red giant, their next goal is to peer into its very core. ■
Star Light, Star Bright

Before there was a Milky Way galaxy, planet Earth, or solar system, out of cosmic darkness came the first star in the universe.

by Michael Schneider
Evening

in the country, a pale glow lingers overhead as darkness creeps in from the east. A gleam of light catches the corner of your eye. First star. Make a wish. The daily drama of this moment hints at a similar moment billions of years ago — the first star in the universe.

To people such as Princeton cosmologist Jeremiah Ostriker and Nickolay Gnedin, his collaborator at the University of Colorado at Boulder, this moment marks a singular phase in the evolution of the cosmos. “There was a period we can characterize as the end of the dark ages,” says Ostriker who is also the team leader of the Alliance’s Cosmology Application Technologies Team. “when all the activity we now take for granted started — the first stars, the first supernova, the beginning of galaxy formation.”

Using NCSA’s SGI CRAY Origin2000, Ostriker and Gnedin have produced the first detailed understanding of this stage in the growth of the universe. Unlike prior studies, Ostriker and Gnedin’s simulation is a fully nonlinear, 3D treatment that includes hydrodynamics, the physics that describes how gas in the early universe cools and collapses under the force of gravity into clumps that become the first stars. Their results provide illuminating explanations for astronomical puzzles, such as what the first galaxies look like and how elements from the upper rungs of the periodic table got scattered like dust between the galaxies.

The big bang itself was the universe’s original burst of light. As the early universe rapidly expanded, however, matter cooled from its beginnings as a seething stew of disassociated particles, and it “deionized” into atoms of neutral hydrogen. Because neutral hydrogen absorbs light energy — unlike most of the hydrogen in our present universe, which is ionized — the universe entered the phase of relative darkness astrophysicists refer to as “the dark ages.”

At this stage, there were no stars, and if there had been, their light wouldn’t have traveled far. At some point stars formed, radiating energy that ionized hydrogen, and the universe became transparent. When and how did it happen? Existing astronomical observations give no answers. “Computer simulations give us a way to explore that intervening period,” says Ostriker, “and find out when all the activity started — when the universe reionized.”
These three frames represent formation of the first stars in the universe. Evolution during this reionization period leads to large-scale structure that resembles the current universe. Color (blue through red) corresponds to the age of the universe (.21 to 1.35 billion years) when the stars form.

"Galaxies are all in these filaments," says Ostriker, "and we can see in the physics exactly why this happens. It's basically that shocks form, and then gas cools in those shocks and condenses into stars."
Reionizing the cosmos

The simulations reveal a panoply of new information, much of it unexpected. For starters, contrary to prior less-detailed studies, Ostriker and Gnedin found that reionization happens suddenly, about 500 million years after the big bang, preceded by slow reheating as matter coalesces. “One moment it's the dark ages,” says Gnedin, “then shortly after it's bright. It takes about 50 million years, which from the point of view of the age of the universe is almost instantaneous.” If the universe is a 45-year-old person, says Gnedin, reionization happened when she was two-years-old and took about two months.

Another important new finding adds to what we know about how elements scatter through the universe. As humans we have in common that we're made from elements heavier than hydrogen and helium - carbon, oxygen, and others — formed only in supernovas and cast into interstellar space by supernova explosions. But there are also heavy elements in intergalactic space, vast expanses where there are no stars, galaxies, or supernovas. How did they get there?

Until Ostriker and Gnedin’s simulations, it was an unanswered question. Physicists postulated that supernovas somehow blasted matter the enormous distances required to deposit it between the galaxies. The simulations show another way. Almost immediately after reionization, stars begin to gather in small groups, forerunners of present-day galaxies. Drawn by each other's gravity, some of these protogalaxies merge, and in the swirling tango of merger, some of their gas breaks off, thrown into space at high velocities. “This is new physics,” says Gnedin. “This mechanism wasn't noticed till now. It's a good example of how large-scale computation can give a qualitatively new result.”

Matching signatures

A number of features from the simulations match well with the current observed universe, indicating that the model accurately reflects reality. One of the more convincing of these features is clouds of gas — called Lyman-alpha clouds — in starless space. Because light from quasars, the most distant objects in the universe, travels through these clouds to reach Earth, scientists have learned a good deal about their composition — density, temperature, and other properties. “We see,” says Ostriker, “that the model we've computed has produced clouds in very good agreement with those observed in the real world.”

By listening to the cosmic microwave background with greater sensitivity than has so far been possible and by adjusting the reception frequency, astronomers of the future should be able to detect a frequency (100-300 MHz) when the background signal goes silent. That frequency, if found, will correspond to the moment, predicted by Ostriker and Gnedin, when the fog of neutral hydrogen suddenly clears. Because of their work, we will know that this frequency pinpoints when the first stars shone.

This research was funded by the National Science Foundation with a grant awarded to the eight-institution Grand Challenge Cosmology Consortium.

Although astronomers cannot observe reionization by the first stars, they can see traces of this process in the structure and elemental composition of the so-called Lyman-alpha forest. The forest reveals itself as a thicket of absorption lines in the spectra of distant quasars, with each line corresponding to an individual gas cloud in intergalactic space. Some of these clouds contain heavy elements produced by the stars responsible for reionization. This simulation of the Lyman-alpha forest shows how these clouds form the network of filaments, sheets, and voids thought to exist in the early universe. Gas temperature ranges from 10^6 K (blue) to 10^9 K (red). The image was computed at NCSA by Yu Zhang, Avery Meiksin, Peter Anninos, and Michael Norman using a sophisticated Eulerian multi-species hydrodynamic cosmology code. Norman and Ostriker are members of the National Science Foundation’s eight-institution Grand Challenge Cosmology Consortium as well as the Alliance’s Cosmology Application Technologies Team.

The image was rendered by John Shalf.
Colliding Neutron Stars

Newtonian physics are sufficient for modeling the first few million years of the spiraling infall of two orbiting neutron stars, but for the last milliseconds, scientists need relativity.

by Holly Korab
In each of the universe's billions of galaxies are two neutron stars spiraling towards each other at ever-accelerating speeds. These inspirals, begun millions of years before, will reach nearly the speed of light just before the stars collide, producing crashes so violent that they warp the fabric of space-time. Using instruments that will be available early in the next century, the gravitational waves from these cosmic splashes will be detectable from Earth.

"It's an amazing phenomenon," says Doug Swesty, a research scientist at NCSA. "You take these things with incredible densities — they are one-and-a-half times as massive as the Sun, but packed down to a radius of 10 kilometers — then have them going around each other at 60,000 RPMs. When they collide, it is violent beyond belief."

In the 80 years since Einstein predicted these collisions, scientists have accumulated impressive indirect evidence of their occurrence. Nevertheless, a multitude of questions remain, from what factors initiate the inspiral to their ultimate fate. Do they meld into a larger neutron star or does matter coalesce further, yielding a black hole? Does their merger produce mysterious gamma ray bursts, and is the merger the source of scarce heavy elements in the universe?

Assembling a celestial orchestra

Swesty is a member of the Alliance's Application Technologies Cosmology Team and part of a five-institution NASA Colliding Neutron Star Grand Challenge Team that believes the answers to these and other questions may lie in the dynamics of the collision — dynamics that have long eluded simulation. They are optimistic that they will accomplish what others have not for a couple of reasons. First, they are among only a handful of teams developing a relativistic code for simulating the collision; that is, one that employs Einstein's complex set of gravitational equations for explaining how a star moves through the universe. Most computational approaches rely on Newtonian physics, which are simpler to compute but provide poor approximations of space-time during the collision.

"Think of Beethoven's Ninth Symphony," says Paul Saylor, a computer scientist at the University of Illinois at Urbana-Champaign who is crafting many of the team's algorithms. "You can play it on the piano, but if you've ever heard it in concert you know something is missing. Newton's equations are not wrong, but they're like the piano version whereas Einstein's are the full symphony."

Adding instruments to the celestial orchestra, though, increases computing demands astronomically — doubling them for post-Newtonian versus Newtonian physics, and increasing 10-fold for a fully relativistic computation. Consequently, another reason the neutron star team is optimistic is that recent tests on the SGI CRAY Origin2000 at NCSA indicate the extra computational complexity is worth the effort. By adding one relativistic effect to an otherwise Newtonian simulation, Swesty and his research group — graduate student Edward Wang and postdoctoral student Alan Calder — saw a significant difference in the speed of the merger. Greater speed influences the amount of matter cast into space and the likelihood that the system will form a black hole. "We are still checking our calculations so we can say how much of a difference it made," says Swesty, "but it was clearly a significant force in the last few orbits."
**Stellar friction**

Neutron stars are the remnants of supernovas — stellar explosions that spew the star’s outer layers into space but compress the core until it is no larger than a small city. Round and smooth, these neutron stars are so dense that even the space within atoms is nearly squeezed out. Some of these neutron stars orbit in pairs, which gradually inspiral and collide.

Most astronomers believe that this inspiral is initiated by the stellar equivalent of friction on the orbits of the binary stars. As the stars lose energy through gravitational radiation, their mutual attraction pulls them closer and closer.

Despite the influence of gravitational radiation on stellar orbits, it is not modeled numerically — that is, relativistically — during the early stages of the stars’ inspiral. Because it is weak, it can be accounted for by simpler, analytic mathematical formulae. Gravitational radiation is often simplified in the latter stages of the inspiral, too, because another phenomenon is known to drive the merger, a purely Newtonian one called tidal distortion.

“Most people think of orbiting binary systems as circular neutron stars moving around each other in nice round orbits,” says Swesty. “But in reality, as these stars get closer, they distort a little bit and form shapes like hour glasses. The forces that cause this are the tides.”

**Rip tides**

Tides in neutron stars like those in the ocean are caused by gravity. Although neutron stars are extremely dense, they are still gaseous and, consequently, behave like fluids. Gravity pulls on their fluid matter, elongating it. This tidal distortion, though slight, can have a momentous effect on the stars’ orbits. In 1994, Stuart Shapiro and Fred Rasio, then physicists at Cornell University, predicted that when the separation distance between neutron stars is approximately three times their radii, the stars are unable to maintain separate orbits. The stars are ripped apart by the tides, forcing them to merge.

Astronomers already knew that tidal instabilities alone were enough to cause a merger. But Swesty’s group suspected that gravitational radiation was underrated as a factor during the stars’ final orbits. “When stars are far apart they act like point masses — that is, as if all their mass were concentrated at a single point in space — and it is easy to approximate their behavior with Newtonian physics,” says Swesty. “Where gravitational radiation becomes a factor is when the stars are close enough for their actual size to matter.”

Swesty’s simulations indicated that this critical distance for neutron star collisions is a few tens of kilometers, when the stars are in their last few orbits and the timespan only a couple of milliseconds. Gravitational radiation dominated the last milliseconds before the merger, proving that it is significant in the dynamics of the merger and providing the Grand Challenge team with a cut-off point for Newtonian physics.
How will they know if their simulation is right?

They'll check their math. And they'll check the skies.

Another impetus behind the work of the Colliding Neutron Star Grand Challenge Team is the completion, in 2000, of NSF's Laser Interferometric Gravitational Observatory (LIGO). Neutron star collisions are rare — occurring maybe once every 10,000 years within a single galaxy. But when LIGO comes online, the scientists will be able to observe between 2 and 300 neutron star collisions a year, given the millions of galaxies it will scan. If measurements taken by LIGO match the team's simulations, they'll have confirmation. "Then we can sigh in relief," says Swesty, referring to the lack of direct evidence of neutron star collisions or gravitational waves. The underlying theories also await confirmation. "The indirect evidence in support of both general relativity and these collisions is overwhelming," says Swesty, "but it will be nice to know."

"No one really knew at what point of the coalescence process the physics of Einstein takes over that of Newton," says Wai-Mo Suen, an astrophysics professor at Washington University in St. Louis, whose research group is developing fully relativistic treatment. "Our goal is to do fully relativistic coalescence with relativistic physics at the end, but we still need to join it to the early part of the evolution, which is well described by Newtonian physics. The investigation that Swesty is carrying out will help determine when we need to replace Newton with Einstein."

Having narrowed the critical time period for relativistic simulation from millions of years to milliseconds, Swesty is running additional comparison tests to determine how other relativistic effects would factor into this sliver of time. Within months he will be able to supply Suen with the information he needs for the starting point of a fully relativistic simulation.

At that point, this team may have their symphony.

This research is funded by NASA's High Performance Computing and Communications program/Earth and Space Sciences.
Quarks Come un-GLUon-ed

Scientists believe that in the first milliseconds of history, the universe was nothing more than a formless plasma of quarks and gluons. Computer simulations may help prove them right.
With billions of years of history to choose from, Robert L. Sugar is interested in just a few fractions of a second — the very first ones.

This physicist from the University of California at Santa Barbara (UCSB) is part of a nine-member, multi-institutional team of high-energy physicists that is recreating the first billionths of a second following the big bang — the explosion of energy that gave birth to the universe. In those microseconds, when temperatures are thought to have exceeded one trillion degrees Kelvin, physicists believe that matter was a formless plasma of quarks and gluons — the infinitesimally small particles from which all nuclei of matter are built.

Sugar and his colleagues are attempting to simulate these extreme conditions to address fundamental questions about the origins of the universe. The team — physicists from the University of Arizona, UCSB, University of Colorado, Florida State University, Indiana University, University of the Pacific, University of Utah, Washington University, and NORDITA — comprise the MIMD Lattice Calculation (MILC) Collaboration, a Department of Energy Grand Challenge initiative that is using parallel computing to model quantum chromodynamics (QCD). QCD is an intriguing and relatively new theory, having arisen within the past quarter century, that describes the "strong" nuclear force that binds quarks and gluons. The model being created by the MILC Collaboration, and simulated using Alliance computers, may help confirm this early plasma state of matter. It will deepen physicists' understanding of how the elementary particles unite and, in particular, how they decay.

**Quarks and quantum chromodynamics**

Although no one has observed isolated quarks and gluons, the theory governing their behavior — quantum chromodynamics — is well-tested and, therefore, accepted as one of four theories that govern all of nature, with gravity, electromagnetism, and the weak nuclear force being the other three theories. QCD tells us that quarks are the subatomic particles that, in addition to forming protons and neutrons, are the building blocks for a host of short-lived particles observed in high-energy physics experiments. These quarks come in six varieties, each with successively greater mass: up, down, strange, charm, bottom, and top. Quarks typically are found in pairs or triplets. Although these particles are continually in motion, physicists can approximate the forms of matter that various combinations of quarks produce. For example, it is thought that a proton contains two up quarks and one down quark, and a neutron, two down quarks and one up quark. Other particles, called mesons, consist of a quark and an antiquark, its antimatter opposite.

Binding, or "gluing," the quarks together are gluons, the carriers of the strong nuclear force. This force is the strongest in nature and behaves unlike any other. Whereas other forces decrease in strength with distance, the strong force increases indefinitely as quarks move apart. As a result, the attempts of physicists to replicate the plasma state by removing a quark from a proton have been something of a Catch 22. Rather than separating the quark, the energy exerted to pull it free creates a bigger and bigger energy field. Eventually the energy field is so large that it spontaneously produces a quark-antiquark pair. Instead of generating a free quark, they produce two hadrons (particles made of two or more quarks). Although physicists have yet to free single quarks from the bundles that contain them, they believe it can be done. At extremely high temperatures, such as those that existed at the birth of the universe, the energy may be sufficient for the familiar particles of protons and neutrons to come unglued, breaking into quarks and gluons.
The MIMD codes are written in the highly portable C language, with only the communications routines being machine-dependent. Sugar says that moving from one machine to another simply requires linking the appropriate communications file. The MIMD code and lattices are available to other researchers. They may be found online at http://cliodhna.cop.uop.edu/~hetrick/milc.

Transition to the quark-gluon plasma

The calculations by the MILC Collaboration are focusing on how quarks and gluons behave when they are heated by taking into account the influence of the strange quark. Previous numerical studies have considered only the up and down quarks because physicists believe that lighter quarks spur the transition to a quark plasma. The strange quark, though significantly heavier than either the up or down quark (which have almost identical mass), is the only other quark light enough to have an appreciable impact on the transition. Preliminary calculations have identified the range of thermodynamic properties where the transition will occur. The current calculations are aimed at determining the precise nature of this transition. Says Sugar: "It might be a rapid crossover or a bona fide phase transition between ordinary matter and the quark-gluon plasma, depending on the effects of the strange quarks."

Recreating this transition, though, is no small undertaking, which is where the Alliance comes in. QCD calculations are among the most challenging numerical calculations in science. For example, one of the MILC's recent studies of the decay constants of heavy-light mesons required hundreds of thousands of processor-hours.

The MILC Collaboration has made significant progress, in large part, by not restricting their computations to single processors, or even to single computing systems with many processors. They have developed a highly portable MIMD (multiple instruction, multiple data) code with which they can run QCD calculations on many different high-performance parallel machines — without having to modify the codes extensively for each machine. They recently ported their MIMD codes to two of the Alliance's SGI CRAY Origin2000s — at NCSA and Boston University — to simulate the complicated QCD phenomena. Because runs with different parameters do not depend on each other directly, Sugar and his colleagues can take advantage of the computers at these separate locations to run several jobs simultaneously.

Moving code between the Alliance computers was especially easy, says Sugar, because of the uniform computing environment. "Basically, we didn't have to alter our code at all moving from NCSA to BU," he says. "The Alliance is clearly providing a very important set of resources to the national community."

Already the MILC Collaboration is generating some of the most promising new knowledge in particle physics. With more computing power and scientific insight, they may soon know what happened during those first split seconds of history.

Nuclear physicists are now planning experiments that will replicate these extreme conditions. If the MILC Collaboration can simulate this quark-gluon plasma, their work will assist these other physicists in establishing the experimental conditions necessary for verifying this plasma state.

This research is funded by the National Science Foundation and the Department of Energy.
A New View on Atoms
Understanding why materials crack is easier when the fractures are simulated in 3D.

by
Karen Green
Sometimes big problems in science and engineering are best examined on the minutest scale. For Diana Farkas, a researcher at the Center for Modeling and Simulation in Materials Science at Virginia Tech, that means studying the development of fractures in intermetallic alloys by examining the alloys, especially their defects, atom by atom.

For more than 12 years, Farkas and her research team have studied new and promising alloys at the atomic level, looking at the bonds between atoms in defective alloy crystals and how they behave under different levels of externally applied stress. The goal of the research is to develop new structural materials that hold up to extreme heat, resist oxidation, and have a high degree of ductility in combination with strength. The work has important implications in the field of materials design, including aircraft and aerospace design.

"We are looking at the characteristics of experimental alloys, such as whether they are brittle or ductile, and whether they maintain their strength at high temperatures," explains Farkas. "We look at what happens in the propagation of fractures at the atomic level. We need to understand the mechanisms of the fracture process in order to design materials with improved ductility and the ability to maintain strength at high temperatures."

Farkas gets her data from computer simulations, which are based on a description of the atomic interaction matched to the equilibrium properties of defective atomic crystals. She then uses a FORTRAN code to calculate the atomistic equilibrium configuration of the defects — either cracks or plastic deformation in the alloys. Until recently, the research team could only view the data on a desktop workstation as 2D slices. These 2D visualization packages are far from perfect; for one, they can't handle the large number of atoms involved in the massive simulations necessary to study the propagation of cracks. Furthermore, 2D renderings can't give a complete picture of a problem that is truly 3D in nature. But in the last few months, Farkas has been able to "see" her data in a new and much more realistic light. Using an application called AtomView, Farkas can now study the behavior of atoms in propagating cracks in the immersive, 3D environment of the CAVE.

**With help from the Alliance**

The development of AtomView is a direct result of Virginia Tech's involvement in the National Computational Science Alliance. John Shalf, a visualization research programmer at NCSA, the leading-edge site for the Alliance, worked with Ron Kriz of Virginia Tech's Advanced Communication and Information Technology Center during the summer of 1997 and helped him write an application that took Farkas's FORTRAN code output — numerical representations of the positions of every atom in a fracture being studied — and put it into the visual 3D form that became AtomView. Shalf's help in creating AtomView was part of the training offered during NCSA's Visual Supercomputing Institute, a program held each year to help scientists utilize the tools of the CAVE and ImmersaDesk virtual environments.

"They came to the institute with their simulation data in hand," Shalf recalls of the Virginia Tech research team. "Because of that, it was possible for me to get them started by helping them get a very simple application running that could read their data." Since then the Virginia Tech researchers have taken over the development of AtomView and are busy extending its capabilities in ways that are useful to Farkas's research team. As a partner in the Alliance, the researchers will continue to develop and evaluate virtual reality tools and applications such as AtomView in their newly installed CAVE. When appropriate, these tools and applications will be disseminated to other Alliance partners and the larger scientific community.
Atom View, says Kriz, is one of the first success stories to emerge from Virginia Tech's collaboration with the Alliance to develop and disseminate cutting-edge visualization tools and applications. "This project was designed in the spirit of developing CAVE applications remotely for viewing animations of supercomputing simulations," Kriz says. "I anticipate that this CAVE application will be further developed and shared by Virginia Tech and NCSA."

For researchers such as Farkas, Atom View means the opportunity to view bigger blocks of data from any angle, and the ability to view 3D simulations of the fracture process for the first time. "Imagine walking between the atoms in an alloy and you get the idea of what it is like to view this information in Atom View in the CAVE," says Farkas. "The beauty is you are able to see something that is very close to a real fracture, only you are seeing it at the atomic level."

### Fractures in 3D

According to Shalf, Atom View is a better tool for looking at alloy fractures and defects at the atomic level because the researcher can see many more levels of atoms at once. "They want to see what's happening at the crack tip, and that's a phenomenon that happens in three dimensions, not on a flat surface," he says. "You need a tool that can handle the large number of atoms involved in 3D simulations. Using a tool like this, you can look in three dimensions instead of one or two, and see what's happening at all levels."

Simulating a fracture, for example, requires data on at least 100,000 atoms, according to Farkas, and needs the power of 128 processors on NCSA's SGI CRAY Origin2000. Utilizing the speed and the performance of the Origin2000, Farkas hopes to soon run Atom View simulations that examine as many as 1 million atoms.

"We are already seeing an increase of three orders of magnitude in the number of atoms we are looking at," she says. "It used to be that we barely had enough memory on computers to handle 2D data. Now we have enough memory to simulate materials in full 3D. We need tools to understand this 3D data and the CAVE and Atom View are a very important part of this."

Bigger simulations that look at the mechanisms of fractures in a realistic manner mean more information about what happens in the fracture process, especially at the critical region of the crack tip. And understanding what happens to cause fractures and other defects is an important step in developing stronger, more ductile materials. It's a way of tackling the big problems of engineering by looking at the tiniest pieces of matter — but in ever more massive chunks.

With the bigger simulations "there is time saved and there is more information to be studied," says Farkas. "And with the Alliance and NCSA, there is an emphasis on collaboration instead of competition. In the long run, you get a lot more done."

This research is funded by the Office of Naval Research, Division of Materials Science, and the National Science Foundation.
By adding just the right amount of impurity to a metal alloy, mechanical engineers can design materials that will bend but not crack under stress.

Stressed Out Metals

by Daniel Pendick
The British poet William Blake saw a universe in a grain of sand and eternity in an hour. Physicist Arthur Freeman also sees a universe in tiny things — a universe of opportunity for materials science. Using modern quantum physics, Freeman explores the interactions among electrons and nuclei that give rise to the fundamental properties of matter. And fortunately for him, access to the supercomputing resources of the Alliance ensures that it won’t take an eternity to make the calculations.

Freeman is the Morrison Professor of Physics at Northwestern University in Evanston, Illinois. He’s a leader in the emerging field of computational materials science, which rests on the principle that the properties of matter arise from interactions among atoms that can be rendered mathematically. By calculating the distribution of electrons in a material across a spectrum of energies — its electronic structure — it is possible to predict how well the material conducts heat or electricity, how ductile or brittle it is, how it responds to magnetic fields, and other fundamental properties.

Freeman is best known for his theoretical advances in understanding the magnetic properties of thin films and surfaces. His calculations revealed, for example, that thin films — those less than one micron (one millionth of a meter) thick — can actually have stronger magnetic properties than the material in bulk form. Other researchers are working to parlay some of his discoveries into higher-density data storage technologies such as compact disks. For his work in magnetic materials, Freeman received the first Materials Research Society Medal in 1990 as well as the first International Union of Pure and Applied Physics (IUPAP) Award in Magnetism in 1991. Freeman is also the founding editor and editor-in-chief of the Journal of Magnetism and Magnetic Materials.

Recently, Freeman and two post-doctoral students from his group at Northwestern — Oleg Kontsevoi and Oleg Mryasov — used the SGI CRAY Origin2000 at NCSA to study atomic-scale defects that develop in all solid materials. The defects are important because they determine the strength of a material — its ability to resist mechanical stress. When materials undergo stress, dislocations — that is, shifts in the position of atoms in the crystal — allow the material to deform plastically instead of break. The Northwestern team simulated different types of dislocations. One type, called edge dislocations, can be imagined as extra rows of atoms squeezed between adjacent rows. The other type that they studied, called screw dislocations, happens when rows of atoms are twisted with respect to each other into a pattern resembling a spiral.
**Bending without breaking**

The study by Northwestern physicists focused on an alloy of nickel and aluminum. Mechanical engineers have their eye on these alloys for a wide range of aerospace and power generation applications, including turbine blades for jet engines and electrical generators. The alloys are useful in that they are light and highly resistant to corrosion and fracture — at high temperatures, that is. At low temperatures, the material becomes more brittle, and in aerospace applications that is potentially catastrophic. "You want to be able to control brittleness," Freeman says. "You don't want something in a jet engine to break apart."

One way to control brittleness is to add impurities. The impurities change the strength of the bonds between atoms in the alloy. The magnitude of the change determines whether dislocations or cracks will be more likely to occur when the material is stressed. Dislocations are potentially useful because they enable rows of atoms to shift ever so slightly with respect to each other without the material breaking, a quality known as ductility. If a sufficient number of the bonds break, cracks develop. But a balance must be struck between ductility and toughness. If the impurity makes the material too ductile, it will lose its shape over time. A fundamental understanding of the effects of dislocations on electron structure, Freeman says, can guide the choice of the right impurity to strike that critical balance.

To do these calculations, the researchers did not actually use the highly precise Full-potential Linearized Augmented Plane Wave (FLAPW) method that Freeman and several colleagues developed in the 1980s. The reason was the sheer magnitude of the simulation they decided to attempt. Dislocations have far-ranging effects on the atoms surrounding them. The researchers simulated dislocations in a central core of 500 atoms, in itself beyond the scope of the FLAPW method because of the enormous amount of computer time and memory the calculations would require. To include the long-range effects of dislocations, the Northwestern team built a shell of 20,000 additional atoms around the core. To make the calculations, they developed an "effective real-space technique" based on the so-called Linear Muffin-Tin Orbital method. This newer approach introduces several approximations compared to the FLAPW method but allows electronic structure calculations involving large numbers of atoms. Still, the calculations of dislocations remain daunting and were made possible by the development of these new parallel computational algorithms, which Freeman's team modified to take advantage of the full power of NCSA's Origin2000 — so far with up to 128 processors.
Confirmation from first principles

Among the results of the new calculations is the first confirmation using a first-principles approach of the appearance of quasi-localized states in metallic materials. They occur when the shift of atoms caused by edge dislocations leaves some atoms in the core of the crystal with fewer neighbors than normal. This results in the appearance of “broken” bonds in the core. When impurity atoms interact with these bonds, it can have a drastic effect on the material and possibly shift the balance between ductility and toughness in a particular material. These findings, Kontsevoi says, will guide engineers in their attempts to modify the mechanical properties of materials in desirable ways.

For now, Kontsevoi and Mryasov have completed the first step, which is a calculation of the electronic structure for a crystal of nickel and aluminum alloy with dislocations. This provides a foundation to build on in the next phase of the research; adding different impurity atoms to the crystal and calculating their effects on electronic structure. The Northwestern researchers are already collaborating with a team of experimentalists at Johns Hopkins University led by mechanical engineer Kevin Hemker. They will try to confirm the predictions of the electronic structure calculations with measurements on real alloys.

That journey from theoretical insight to real-world application, says Freeman, is what makes computational materials science so satisfying — and so much fun. Although it is true that his papers have had a wide impact in the academic world — a tally of citations by the Institute for Scientific Information in papers published from 1981 to 1997 ranks him as fifteenth in the international community of more than 517,000 publishing physicists — Freeman has particularly enjoyed being able to see the theoretical insights put to the test on the laboratory bench. "What I've enjoyed so much is the interactions with experimentalists and the fact that a lot of what we do leads to applications — that it has impact," Freeman says. "The fun is in seeing new understanding emerge about the physics of materials.”

From single atoms to many thousands

Northwestern University's Arthur Freeman is an heir to the pioneering work of John C. Slater, one of the architects of the field of computational physics. In the early 1950s, Freeman was a doctoral student in Slater's Solid State and Molecular Theory Group at the Massachusetts Institute of Technology. Because of the large number of calculations involved, armies of slide rule-wielding graduate students couldn’t have solved problems involving more than a hydrogen atom with its one proton and single electron. Even this was considered a hopelessly difficult many-body problem, akin to manually calculating the trajectory of a space probe.

But Slater's group had access to Whirlwind I, a supercomputer of its day, and used it to perform what were at the time formidable challenging tasks like calculating the electronic structure of OH, a simple molecule of oxygen and hydrogen. "We were allowed to run on the graveyard shift; that is, from midnight to 8:00 a.m.,” Freeman recounted in a recent essay in the Annual Review of Materials Science. "Still, we considered ourselves lucky to be able to harness the power of this machine.” Even so, the work on the OH molecule took "only" a year and a half, Freeman recalls.

In the 1980s, access to more powerful supercomputers enabled Freeman and others to speed up computations considerably and also develop more sophisticated computational methods. Working with his former post-docs Erich Wimmer, Mike Weinert, and Henry Krakauer, Freeman developed the FLAPW method. It was an extension of the Augmented Plane Wave technique formulated by Slater and first published in 1937. It remains the most precise method to date for electronic structure calculations.

In stark contrast to those early days with Whirlwind I, Freeman and his colleagues at Northwestern are now doing simulations involving tens of thousands of atoms. And in large part because of parallel computing, Freeman says, they don’t have to wait years for the results. But to Freeman, access to the latest computational horsepower is about much more than just getting things done in time for dinner; it has enabled him to apply more sophisticated mathematical tools — like the FLAPW — to more and more complex problems. "The key to computational materials science is complexity,” he says. "The more we can closely simulate a real system, the more of the complexity of the material we can predict.”
New Matter: Friction Free Hydrogen

Helium isn’t the only superfluid around. With the addition of a few impurities, liquid hydrogen also will forgo its natural tendency to freeze solid at low temperatures and enter the strange, frictionless state of matter called superfluidity.

by Michael Schneider
Add enough antifreeze and the water in your car's cooling system will keep on flowing through the iciest winter. Applying basically the same idea, say University of Illinois at Urbana-Champaign physicists David Ceperley and M. Carmen Gordillo, you can prevent hydrogen from freezing into a solid as it usually does at low temperatures (below 14°K). Their simulations show, furthermore, that when it gets even chillier (1.2°K) this unfreezeable liquid hydrogen shifts to the strange, frictionless state of matter known as superfluidity.

Their work (reported in Physical Review Letters, October 20, 1997) opens new avenues for research in condensed matter physics by adding hydrogen to the exclusive list of substances that become superfluid. At the top of this list is helium, which has been a focal point of research into low-temperature properties of matter since it was first liquefied in 1908.

Since hydrogen is in many fundamental respects similar to helium, physicists speculated for years that hydrogen should be a candidate for superfluidity. Hydrogen molecules, however, attract each other more strongly than do helium atoms and normally combine at low temperatures to form a solid. Ceperley and Gordillo surmised that adding impurities to liquid hydrogen would suppress this molecular urge to get together and solidify. "It's so close to becoming a liquid," says Ceperley. "If we could just make the liquid more stable, then maybe it could become a superfluid."

Simulations on NCSCA's HP-Convex Exemplar (20,000 processor-hours) confirmed this insight. With an exact quantum approach that Ceperley derived from 1950s work by Nobel Prize-winning American physicist Richard Feynman, the researchers modeled hydrogen molecules on a flat surface salted with potassium atoms. The repulsive effect of the alkali-metal stops the hydrogen molecules from combining to freeze into a solid, explains Ceperley, and the liquid hydrogen undergoes a transition to superfluidity just above absolute zero.

Liquids that flow without friction

Although a number of this century's outstanding physicists have contributed to the understanding of superfluidity, it remains one of the fascinating phenomena of physics. No one knew there was such a thing until the early experiments with liquefied helium by the Dutch physicist J. Kamerlingh Onnes. In 1910 he found that helium wouldn't freeze even at one degree above absolute zero — as cold as he could go — and measurements showed, strangely, that as liquid helium declined in temperature its density peaked at 2.2°K. Later studies by Dutch, Canadian, and Russian physicists, including the noted theorist Lev Landau, identified this density maximum as a transition to a special state, which by the late 1930s was called superfluidity.

Experiments show that superfluid helium has no viscosity and flows through even the tiniest capillary tubes with no resistance. Perhaps the most striking demonstration of superfluid defiance of the ordinary is its behavior in a rotating container. Put a cup of coffee on a turntable and the coffee rotates with the cup. Place liquid helium above superfluid temperature inside a rotating container; explains Ceperley, and it acts like the cup of coffee. As you cool it through the superfluid transition, however, the liquid stops rotating and comes to rest. That's the lower-energy state that superfluids get into. They completely insulate themselves from what's going on around them.

Landau and others recognized these strange behaviors as quantum phenomena occurring at an observable scale. "It's a manifestation at the macroscopic level of the fact that the quantum particles are indistinguishable," says Ceperley. Instead of moving independently, like the atoms of a liquid normally do, atoms of superfluid helium cohere in a collective state. The effect is similar to a laser beam, notes Ceperley, in that in both cases — photons from the laser and helium atoms or hydrogen molecules — the quantum particles are bosons, particles with integral spin. "The physical effect is that the more there are in a particular state, the more want to go into that state."
Superfluidity also bears similarities to its better known first cousin, superconductivity, another macroscopic realization of quantum effects. "If you try to put a magnetic field on a superconductor, you can't," says Ceperley. "It pushes out the magnetic field. It's like rotation to a superfluid." Similarly, as it's possible to establish current in a superconducting circuit that will flow forever, superfluid helium can be made to flow in a loop, a "superflow" that continues as long as superfluid conditions are maintained.

**Line dancing in imaginary time**

To accurately simulate the exotic properties of superfluid helium and to test their hypothesis about hydrogen, Ceperley and Gordillo used a computational approach, path-integral Monte Carlo (PIMC) calculations, which Ceperley developed as an extension of work by Richard Feynman. "Feynman provided the first satisfactory theoretical explanation of superfluids," says Ceperley, "and he introduced the idea of imaginary time path-integrals."

In the 1950s, Feynman's ideas were limited by computational technology. Ceperley developed PIMC into a practical numerical algorithm for calculating superfluid properties: "Essentially, Feynman showed that this quantum system is equivalent mathematically to a classical system of exchanging polymers. The thermodynamic properties - energy, pressure, and superfluid properties - have exact equivalents in the classical domains, and the PIMC simulation translates from one domain to the other. There are many refinements needed to make it effective on computers, and that's what we developed."

The method relies on a simple mathematical transformation - changing time in Schrödinger's equation, the fundamental equation of quantum mechanics, to imaginary time. "You can only do these many-body quantum simulations in imaginary time so we have to make a virtue of necessity," says Ceperley. The calculations trace the particles as they move in imaginary time and, like a game of musical chairs, they must either return to their starting place or the starting place of another particle.

In this formalism, says Ceperley, it turns out that superfluidity is equivalent to what happens if you progress from one position to the next in a line dance. "It takes a long time to get back to where you started. If the line is 20 couples long, it takes 20 iterations. That's the property we look for - it's called a permutation of labels - an exchange that extends across the system. If you look at those paths, you'll see they're all linked together, like one huge molecule, and that's superfluidity, in this imaginary time."

**Looking ahead**

Ceperley and Gordillo's result with simulated hydrogen and alkali-metal atoms suggests an obvious direction for experimental work, and several research teams are expected to try to confirm their finding of superfluid hydrogen. In future simulations, they'll refine their calculations of hydrogen superfluidity and address the possibility of hybrid superfluids such as mixtures of helium and hydrogen.

As part of the Alliance's Nanomaterials Application Technologies Team, they will also work to forge PIMC into a generalized computational tool. "We hope that eventually PIMC can be a 'black box,'" says Ceperley. "A nonexpert could specify temperature, particle masses, spins, interactions, chemical potentials, and boundary conditions of the quantum system, and the computer could return estimates of various observables, complete with error bars. It should be a major goal of computational many-body physics to show how this can be done."

This research was supported by the National Science Foundation and the Spanish Ministry of Education and Culture.
Making a **Metal Out of Hydrogen** by Holly Korab

**Extreme** cold may bring out the superfluid nature of hydrogen, but David Ceperley’s group is also narrowing in on another exotic behavior of this molecule. Squeeze hydrogen at pressures 2 to 10 million times greater than normally found on Earth and this colorless, odorless, tasteless gas transforms into a metal.

“Eugene Wigner, the Hungarian physicist, predicted this metallic state more than 60 years ago,” says Ceperley, “but he lacked the ability to make quantitative predictions of this state. With new computing algorithms we are tantalizingly close to doing so.”

Recent simulations by Ceperley and graduate student Burkhard Militzer — again using their path integral Monte Carlo calculations — indicate that hydrogen’s transition into this metallic state is abrupt. Just as water skips the slush phase when it freezes into ice, their simulation shows hydrogen undergoing a similar first-order transition, in this case from a metallic molecular fluid into a metallic atomic fluid, when it reaches 2 million atmospheres. This kind of information will help physicists better understand hydrogen’s role in the evolution of stars and planets — whose interiors do sometimes reach these pressures — and to manipulate these conditions to produce exotic and potentially useful materials. These extremes of pressure can exist on Earth as well — during thermonuclear fusion.

How does hydrogen convert from a freewheeling gaseous state to the orderly, compact lattice that is characteristic of cold hydrogen? As hydrogen is compressed, its electrons abandon their usual spherical orbits and lock into a herringbone lattice, with the alternating rows pointing in different directions. Greater compression is thought to force the electrons into a simple metallic lattice. Ceperley and Militzer are studying what happens when both density and temperature are increased a few thousand degrees higher. Says Ceperley, “We’re interested in those temperatures and pressures at which hydrogen is a real metal — the molecules are gone and you have just hydrogen atoms and free electrons.”

As density increases, hydrogen acquires some metallic properties. Electrons that were localized around a bond are now becoming delocalized and are exchanging; that is, they are forming connected-conductive paths (denoted in yellow).

Here hydrogen is a true metal — with all the bonds broken and the electrons delocalized. Ceperley and Militzer’s work indicates that this transition is abrupt.

This phase diagram plots the state of hydrogen — molecular, atomic, or plasma — at given temperatures and densities (a function of pressure), based on the findings of Ceperley and Militzer. The blue line indicates the halfway point in a gradual transition from one physical state to another, here from a molecular fluid to plasma. Red indicates an abrupt, first-order phase transition, in this case from a molecular fluid to and ionized metallic one.
Stretchy Molecules in Low Drag Solutions

Common sense tells us that thick fluids should flow more slowly than thin ones. Experience with polymer additives tells us otherwise.

by Michael Schneider
During World War II, researchers at Edgewood Arsenal in Maryland found something they didn't expect to find. They were curious about gelled gasoline, the thick, jelly-like substance used in flame throwers—formed by adding a chemical thickener, such as aluminum soap, to gasoline. The researchers noticed that at high flow rates the thickened gasoline flowed with less pressure, as much as 50 percent lower drag, than the unthickened gasoline.

At the time, due to wartime security, this intriguing finding went unreported. In 1949, a British chemist, B. A. Toms, announced his similar, independent finding that dissolving small quantities of heavy, long-chain molecules—polymers—in solution reduced drag during turbulent flow through a tube by up to 70 percent. Since the 1960s, a large body of experimental knowledge has built up on this phenomenon—sometimes called "the Toms effect." It has been applied in the oil industry: in fluids pumped at high pressure to fracture oil-bearing rock formations, and in the Alaskan pipeline, where mixing polymers with crude oil cuts pumping costs in half.

What, exactly, is going on here? Centuries of engineering experience, not to mention common sense, tell us that thickened "heavy" fluids should flow with more resistance than "thin" liquids. The experimental work has fostered a good deal of theorizing about this quirky phenomenon, but in many respects it remains an enigma. What characteristics of the polymer additives govern the effect? How can it be controlled and used in other applications? Despite almost half a century of research, says Antony Beris, "details are sketchy and direct evidence for any specific mechanism is still lacking"—which is where supercomputing enters the picture.

A professor of chemical engineering at the University of Delaware, Beris has taken on the challenge of building an accurate computational model of this phenomenon. He appears to have solved some vexing problems (numerical instabilities) that defeated prior attempts at modeling these flows. His recent work—in collaboration with former graduate student R. Sureshkumar (now on the faculty at Washington University in St. Louis)—is the first successful simulation of polymer-induced drag reduction "from first principles." Starting from scratch, without experimental data—with only the governing mathematical expressions and a theoretical model of polymers, Beris obtained results that match observed data about these thickened fluids. More importantly, his results contribute at the theoretical level—by offering new support for an explanation of the reduced drag that relates the molecular characteristics of polymers and the macroscopic behavior of turbulent flow.

**Stress memory and relaxation time**

Most fluids are viscous (from the Latin word for the sticky deposits left by birds on cars parked under trees), which means they have internal friction, where the molecules tend to adhere to each other and resist an external force. Water has low viscosity, for instance, and molasses much higher, especially in January. In most fluids, as Newton showed in the 17th century, viscosity is measurable as a simple relation between how much force is applied and how fast the fluid flows in response.

But add even small amounts of polymer and you transform a normal, well-behaved viscous fluid into a miscreant. The resulting solutions are viscoelastic; they defy Newton's orderly law of viscosity. The challenges of simulating polymer-induced drag reduction stem from this difference.
Viscoelastic fluids are partly viscous and partly elastic, having properties like a rubber band, which when stretched has memory and wants to get back to its unstretched state. A viscoelastic fluid flows in response to force, like a viscous fluid, but it also "remembers" its previous state and resists extension of the flow.

To simulate these complex effects on a CRAY T3E at the Pittsburgh Supercomputing Center (and later at NCSA), Beris and Sureshkumar adapted a numerical method (a well-established spectral code) with proven reliability in simulations of viscous turbulence. They used fully turbulent viscous flow data (provided by their collaborator Robert Handler of the Naval Research Lab) to initialize the viscoelastic simulation. To address the added complexities of viscoelasticity, they implemented a model used extensively to correlate the molecular characteristics of long-chain polymers with experiments on polymer flow. This model, known as the FENE-P (finitely extensible nonlinear elastic-Peterlin) "dumbbell" model, represents a polymer as two spheres with a spring between them. In strong flows, the dumbbell orients with the flow, and the spring stretches with an elastic memory that opposes the flow.

Earlier attempts to incorporate viscoelasticity with turbulence ran up against problems in modeling the "relaxation time" of the polymer. A stretched polymer's memory of its unstretched state fades with time, like a worn-out elastic waistband. How rapidly it fades depends on its molecular characteristics. In work preliminary to their recent simulations, Beris and Sureshkumar focused on resolving the problems of accurately modeling this relaxation time.

"Our simulations," says Beris, "solve a set of nonlinear partial differential equations that describe flow within a parallel channel geometry. The geometry is simple, but the governing equations and flow conditions are highly complicated, corresponding to a fully developed time-dependent, 3D turbulent flow modified by the presence of macromolecules. This complexity requires a highly accurate simulation, which in practice can only be achieved through a multimode spectral representation. The size of the problems solved requires supercomputing, and the complexity of the flow prohibits lower resolution. In this application, therefore, the supercomputer makes the difference, not only in terms of speeding up the calculations, but also — because of the memory requirements — making the simulations possible."

**Streaky structures and low drag**

The signature imprint that differentiates turbulence from the smooth streamlines of nonturbulent flow is whirling eddies, big and little whirlpools, known in fluid dynamics as vortices. An important technique for observing and analyzing the effect of these eddies is a laser snapshot through a 2D slice of the flow. The "streaky structures" observed with this technique, which depict fluctuations in the flow velocity, change noticeably depending on whether the flow is viscous or viscoelastic, and they give a visual indication of the drag-reduction phenomenon.

Visualizations from Beris's simulations correspond to the observed streaky structures, and they show the same features seen experimentally. "We're able to see all the complicated structure exemplified by the streaks," says Beris. "By following how the colors merge and fade, you follow the correlation among velocity values, which represents the underlying large-scale structure that feeds the turbulence. Each structure represents the core of an eddy."
Streaky structures from the simulations and other quantitative results show — for the first time numerically — a decrease in vorticity that corresponds to lowered drag. Introducing the polymer not only reduces the intensity of the velocity fluctuations, says Beris, but also shifts the most intense fluctuations away from the wall and toward the bulk of the flow.

This result tends to confirm a prominent explanation for the lowered drag. Eddy formation along the wall is what sustains turbulence throughout the flow, according to theory, and several researchers — including Beris's University of Delaware colleague Arthur Metzner — proposed as far back as the 1960s that the elastic memory of the polymers, which pulls back against the flow, reduces turbulence by making it harder for these near-wall eddies to get started. "There were other ways to explain the drag-reduction phenomenon," says Beris. "By providing these data independently, we've given credence to the eddy assumption."

The simulations also provide quantitative findings that relate the onset of drag reduction to characteristics of the polymer. The potential for the polymer to resist flow extension is represented by a parameter, \( L \), that corresponds to how much the long-chain molecules can stretch. Simulations show no drag reduction at \( L = 2 \) and increasing drag reduction as \( L \) increases from 10 to 30. "It's strong evidence," says Beris, "that extensional behavior is critical."

Another important parameter, known as the Weissenberg or Deborah number, relates the relaxation time of the polymer to how fast the flow can stretch it. As this parameter increases, the solution has a stronger tendency to resist extension. "It's critical," says Beris, "that the characteristic relaxation time of the material is higher than the characteristic flow time." The simulations show that a high Weissenberg number, 10 or above, is required for drag reduction.

This research was funded by the National Science Foundation, Division of Fluid Mechanics and Hydraulics, Ocean Engineering and Marine Systems, and a NAVY/ASEE sabbatical fellowship at the Navy Research Laboratory.

Looking ahead to surfactants

In collaboration with graduate student Costas Dimitropoulos (and Mark Straka of NCSA), Beris has ported his code from the CRAY T3E at Pittsburgh to NCSA's SGI CRAY Origin2000. Initial runs show improved performance on the Origin, and Beris expects further improvement with code optimization. In upcoming simulations, he expects to simulate polymer drag reduction at higher turbulence, and to investigate the impressive drag reduction observed with soaplike materials called surfactants.

Interest is growing in this field, says Beris, because of progress in the ability to do simulations, and also due to the potential of surfactants. Experiments have shown up to 300 percent lowered drag, but the result is very sensitive to specific conditions. "There are many parameters involved, and it's very difficult to study experimentally. Once you have the suitable models, it's much easier to do it with simulations."

In the past, experiments have looked at polymer-induced drag reduction for naval vessels and submarines. Studies with fire department pumpers show a 50 percent longer throw of the water stream, with a more coherent jet. These and other applications can now be re-examined in light of the improved results available from simulations.

"For a long time," says Beris, "it wasn't clear you could predict this phenomenon from first principles, and I suppose that's what's most exciting about this work. To see these structures emerging, which people have observed experimentally, in a phenomenological fashion in the simulation — that's been very encouraging."
Half of all arterial grafts fail within 10 years, often because they've become as blocked as the arteries they replaced. The problem may be the angle at which the grafts are attached.
for the millions of Americans who suffer from atherosclerosis is one of uncertainty and fear. A fatty blockage of the arteries gradually obstructs blood flow and, ultimately, causes the heart to stop beating. Their only warning may be a short-lived but binding pain that feels as if someone is tightening a belt around their chest. Seemingly defying medicines and surgical procedures, atherosclerosis remains one of the leading causes of heart attacks in the United States and around the world.

Now a new set of computer codes for modeling blood and other complex fluid flows may lead to changes in accepted surgical practices that will extend the life expectancy of those suffering from arterial diseases such as atherosclerosis. These codes can faithfully produce simulations of flows that once could only be captured accurately in "still" formats such as magnetic resonance imaging (MRI).

The researcher behind this dynamic technology is George Karniadakis, a professor of applied mathematics at Brown University. The codes he is developing with his students at the university's Center for Fluid Dynamics and simulating at NCSA, will allow scientists to more accurately model all types of fluid flows. Moreover the codes, collectively known as NekTar, allow scientists to zero in on specific areas of a calculation while the computation is running, making corrections or alterations on the fly. Developing new applications of this kind of computational steering is a major emphasis of the Alliance.

Atherosclerosis has defied scientists in part because of the sheer complexity of the human circulatory system. The circulatory system, with its miles of arteries, veins, and vessels, not only carries oxygen but also transports wastes and aids in heating and cooling the body. For researchers, trying to study blood flow with MRIs and other "still" media has been like trying to guess the plot of a film from its publicity stills. Pinpointing the body's natural defenses against atherosclerosis has been all but impossible. Recognizing this, it shouldn't come as a surprise that one of the leading surgical procedures to correct atherosclerosis has a high failure rate. Some 400,000 people receive arterial grafts each year to correct blockages or replace damaged arterial tissue. Half of these grafts, whether polyester prosthetics or constructed from the patient's own veins, will fail within ten years, often becoming blocked themselves.

The most common area for blockage is at the graft junction. Until now, most grafts were joined at right angles and in the same plane as the existing artery. This creates a U-shape around the blockage. However, this may not be the proper shape for arterial grafts, says Karniadakis's former student Spencer Sherwin. Sherwin and his colleagues at London's Imperial College of Science, Technology and Medicine initially tested this theory on NCSA computers, using an adapted version of one of the NekTar codes. Using the code in London, they recently discovered that blood does not flow symmetrically, as scientists originally believed. Instead, it flows in two opposing corkscrews surrounded by a third corkscrew-like flow. This combination creates an asymmetrical, spiraling flow that is encouraged by the natural curves and bends of human arteries. The researchers believe this spiraling may remove the plaque that
can lead to atherosclerosis. Further, attaching grafts so they curve in to the blocked artery like freeway on-off ramps, instead of flat Us, may better promote this plaque-scrubbing effect.

Sherwin’s team is now running arterial blood flow simulations with NekTar on NCSA’s SGI CRAY Origin2000 to determine the optimum shape for arterial grafts. They hope to determine how much surgeons will need to angle the grafts to maintain optimum friction between blood and the arterial wall: too little will result in blockage; too much could damage the graft.

For their NCSA runs, the researchers constructed a 5,000-element prototype of an arterial graft from hundreds of MRI images. In their simulations, they will vary the angle of the graft, gradually lifting it from the plane of the blocked artery up to 90 degrees until they find the optimum level for swirling flow. At the same time, they will incorporate factors to account for changes in blood flow resulting from changes in the body’s physical position and poor posture. Switching from standing to lying down affects blood flow velocity as does slumping one’s shoulders. The researchers also will incorporate data to simulate the pulsing nature of blood as it is pumped by the heart. Each simulation will take more than 3,000 computing hours.

Despite the complexity of the calculations, Karniadakis has faith in his method and the codes, and with good reason. Last fall, comparisons of MRI blood velocity in actual arterial grafts, both in the same plane as the blocked artery and inclined, corresponded closely with Sherwin’s computations.

Flexible code for a variety of flows

Karniadakis began working on NekTar in 1983 while a graduate student at the Massachusetts Institute of Technology. He was inspired by the work of Princeton University’s Steve Orszag, who was using spectral methods to study fluid flows with simple geometries.

Spectral methods rely on Fourier analysis, a process of approximating data for continuums, such as the range of sound wave frequencies, by using points calculated from complex, exact mathematical functions. This process creates the grids that computational scientists use to study objects and the forces acting upon them.

Building on Orszag’s work, Karniadakis theorized that it should be possible to develop a spectral method for examining flows with more complex geometries that may have broader applications in physics and engineering. For example, instead of studying flows past a cylinder, Karniadakis wanted to study flows around a cylinder in a walled box: “I wanted to be able to reproduce not only the physical context of the experiment, but the surrounding details,” he says.
Karniadakis's method and the resulting NekTar fluid dynamics codes can simulate flows on a variety of grids whether the grids are affixed to the object they represent (structured), fixed in space with material flowing across them (unstructured), or a hybrid of the two. This makes the codes highly versatile. Scientists can use NekTar to study large-scale, microscopic, and atomic flows. At NCSA, Karniadakis and his students are using NekTar to verify a new theory of turbulence. The codes already have found their way into other universities and private research groups. Companies, including Boeing and Ford, also have requested the codes from Karniadakis.

NekTar's accuracy is due, in part, to its central algorithm, which is based on a new class of computational calculation methods. "It is a superset of the three methods used in computational fluid dynamics," says Karniadakis. The codes' accuracy also are due to the fact that NekTar uses complex equations within its grid units to provide increased resolution. Increasing the complexity of these equations expands this resolution by raising the likelihood that points will be accurately plotted within the grid units.

The major difference between NekTar and other fluid codes lies in how it handles increased resolution. By their nature, fluid simulations require enormous amounts of memory and computing power. Because NekTar employs computational steering, it allows researchers to discreetly increase resolution in specific areas, for example, the junction of an arterial graft, while the simulation is running. This keeps memory requirements, computation time, and costs in check.

Karniadakis's work on NekTar is far from over, though. He and his students are refining the codes to accurately compute the continuum of fluid flows between microscopic and atomic levels. This work, as with what has preceded it, may have wide-ranging implications, even improving the performance of the cars we drive and the jets we fly in. The most important impact of the NekTar codes, though, may well lie in helping to extend the length of time we have to enjoy these things.

This research was funded by the National Science Foundation, Department of Energy, Air Force Office of Scientific Research, and Office of Naval Research.
Tiny structures called photonic crystals, which trap and control light with extraordinary efficiency, may herald a new generation of miniature optical devices.

City of Light
Physicist John Joannopoulos occasionally entertains such fanciful visions; mostly, though, he spends his time bringing them to fruition. Joannopoulos’s team at the Massachusetts Institute of Technology (MIT) is using an SGI CRAY Origin2000 at NCSA to simulate photonic crystals — structures that, like imaginary cities of light, guide photons along minuscule “streets” (waveguides), store it in tiny “buildings” (microcavities), and control its symmetry and frequency. This line of research, says Joannopoulos, “will eventually enable scientists to do things with light that they haven’t been able to do before.” At the very least, he says, these crystals will lead to much-improved designs of optical devices such as light filters, lasers, and light-emitting diodes, which send out the pulses of light that are used widely in communication networks.

Already materials fabricated with photonic crystals based on the simulations by Joannopoulos’s group are reducing the amount of light lost when it is squeezed into a small space or when a cable makes a 90-degree turn. In most existing fiber optic cable, for example, only about 30 percent of a light wave successfully navigates sharp bends — the remainder scatters off the corner. In the material being tested by Joannopoulos’s collaborators, light is transmitting around these corners with 98 percent efficiency. This tighter control, says Joannopoulos, is what could make structures like those he is simulating mainstays of the miniature optical devices of the future.

Controlling the flow

Controlling light amounts to trapping it preferentially. Light tends to travel through materials that have a high index of refraction — a measure of how much light bends when it travels through a material. Fiber optic cables take advantage of this fact by forcing light to travel through a high-index cylinder. The cylinder, a fiber optic waveguide, is surrounded by a lower-index material that confines the light to its path. But this layering approach yields control only in the direction of the cylinder... and it is inefficient.

It was nearly a century ago that physicists learned that they could control the flow of light — in one direction — by alternately stacking a material with a high refractive index with a lower index one. This “sandwich” arrangement prevented light waves with specific frequencies from passing through the stack.
To control light in all directions, researchers turned to semiconductors — unusual materials that conduct electricity without resistance — as models. They knew that electrons flowed easily through metals because only a small amount of energy was required to boost electrons into a metal’s many unfilled low-energy orbitals. Semiconductors impede this flow because, in this material, the low-energy orbitals are mostly filled. The only available orbits are of higher energy, requiring a considerably larger input of energy to kick an electron into this region. Also, separating the low-energy orbitals from the higher ones is a range of energy levels, called band gaps, in which electrons cannot exist.

A few years ago, physicists discovered that by introducing impurities or defects into these band gaps — say a phosphorus atom in a matrix of silicon — they could create a new energy level in the midst of the band gaps — and only in the vicinity of the impurity or defect. The photonic crystals being designed by Joannopoulos and his collaborators are attempts to take advantage of these properties to control specific frequencies of light. Determining what materials to use in creating these photonic gaps and identifying the impurities or defects that will create a desired effect is where NCSA’s Origin2000 comes in.

Carving out traffic corridors

Using the Origin2000, Joannopoulos’s team varies the 3D structure and the refractive index values of the constituent materials to identify which structures — in theory — can best control the light. Their choice of structures is guided by the Maxwell equations, written by James Clerk Maxwell in 1865, which describe how the magnetic and electrical portions of a photon respond to the 3D structure and properties of the atoms in their vicinity. “Suppose I have this (proposed) photonic crystal system and I want to predict the magnetic and electric field at a place in time and space. Maxwell’s equations give that information,” says Joannopoulos. “Then we tune the structural properties and materials to make the photons do what we want.”

Once his team has zeroed in on the most promising candidates, they turn the model over to Leslie Kolodziejski, Erich Ippen, and Henry Smith in MIT’s electrical engineering department and Kim Kimmerling in the materials science department to evaluate. After several iterations, these researchers fabricate and test the structure for its light-manipulating capabilities.
The improved waveguide is an example of one of the new crystals emerging from this and similar collaborations. The MIT researchers introduced a flaw in the crystal by carving a path in its high-index/lower-index periodicity. In a crystal made up of cylindrical high-index materials surrounded by air, for example, they simply removed a row of cylinders. "Forbidden" frequencies now pass smoothly through the corridor, directing the light but still containing it within the photonic crystal.

This construct is a higher-tech version of a fiber optic waveguide, and it was recently patented. Joannopoulos's city of light is becoming more real every day. ■ This research was supported, in part, by the MRSEC Program of the National Science Foundation.

Microcavities

In a recent letter to Nature ("Photonic-bandgap microcavities in optical waveguides," 13 November 1997, 390:143-5), Joannopoulos's team reported the properties of another structure, called a microcavity, which can act on light frequencies that are used in optical communication networks. Here, the high-index/lowe-index periodicity is broken up at just one spot — the gap between depressions on the cylinder is wider in the middle. As a result, a frequency of light that is forbidden to exist along the cylinder can be trapped and held at the center by the microcavity. Like the improved waveguides, microcavities may enable engineers to produce much smaller and more precise optical filters and diodes than is currently possible. That, in turn, may enable them to pack many more into a miniaturized device, dramatically improving its efficiency and power. ■
Giving Electrons a Nudge — One at a Time

Plants convert sunlight to energy by moving single electrons from one point to another. Scientists would like to replicate this simple, yet amazing feat.

by Kenneth Chang
A particle of light flies out of the sun, travels 8 minutes, 20 seconds, 93 million miles, down through Earth's atmosphere and hits a leaf on a plant. Within the leaf, a molecule of chlorophyll absorbs the energy, knocking one of its electrons into an excited state. A ring-shaped molecule called a quinone transfers the excited electron away from the chlorophyll and then shepherds it to a second quinone. The plant has now stored the sunlight energy in the electric field between the negatively charged quinone and the now positively charged chlorophyll — a tiny battery, in other words.

These are the first steps in photosynthesis, the energy factory of plants. These are also steps that Ralph A. Wheeler, a chemistry professor at the University of Oklahoma, would like to understand better using the computers at NCSA.

Electron transfer — transferring an electron from Point A to Point B to Point C — sounds simple. "You can think of it like water running down a hill. The water wants to run down to the more stable point, and so do the electrons," says Wheeler. "The plant stores energy by using sunlight to pump electrons uphill and does this exquisitely well." Although scientists know the basics of photosynthesis, they have not yet entirely replicated this efficient feat of chemical engineering experimentally.

One reason is that the study of the movement of single electrons in electron transfer is still relatively new. Chemists traditionally have focused on pairs of electrons, which is how electrons generally arrange themselves as they seek the lowest, most stable energy states. For example, two atoms of hydrogen, each with one proton and one orbiting electron, will join into a dumbbell-shaped hydrogen molecule. With the two electrons now paired, the molecule has a lower energy state than the two separate hydrogen atoms.

That same electron craving makes molecules with an odd number of electrons highly unstable, likely to react with the first molecule that passes by, and thus difficult to study. In some processes, like photosynthesis, the movement of single electrons is paramount. Beginning in the mid-1950s, chemist Rudolph A. Marcus pioneered this research in the field — much of it performed while at the University of Illinois at Urbana-Champaign in the 1960s and 1970s. Marcus, now at the California Institute of Technology, explained how interactions between molecules momentarily increase the energy of a system, driving the electron to jump from one molecule to another. That work won him a Nobel Prize in 1992.
Wheeler's simulations examine the same essential puzzle — the movement of single electrons — but they look at specific reactions in detail. For photosynthesis, the subtle dynamics in thousands of atoms nudge electrons from one quinone to another. This degree of detail is needed for future breakthroughs, and they are why Wheeler requires computing resources like those at NCSA.

**Proteins tip the balance**

During photosynthesis, chlorophyll and the electron-shuttling quinones are joined by a protein and other molecules into what is collectively called the *photosynthetic reaction center*. Most synthetic models of photosynthesis ignore the protein because it is just too big and complicated to fit into the calculations. Although this protein does not play a starring role in this reaction, it plays an important bit part. "The protein contact with the quinones drives photosynthetic electron transfer to the second quinone," Wheeler says.

The machinery of photosynthesis is remarkable in that it stabilizes otherwise unstable molecules and provides a "laboratory" to track electron transfer. After the first electron is transferred from the chlorophyll to the first quinone, the protein nuzzles the quinone, forming a weak bond between the protein's hydrogen atoms and the quinone's oxygen atoms. "The protein pulls on the oxygen and deforms the quinone," Wheeler says.

These bonds tip the energy levels of the second quinone, dumping the electron into it. What makes this such a difficult problem is that as the electron shuffles from chlorophyll to quinone to quinone, it's also changing the molecules' shapes and electronic properties. Understanding the specifics of what's going on requires tracking thousands of individual atoms in the proteins and quinones, and that's what Wheeler and his research team are doing.
Shifting bonds

On the SGI Power Challenge Array, Wheeler and former graduate students Kurt Grafton and Scott Boesch first perform quantum chemical calculations to determine the positions of the electrons orbiting the quinone, a computationally intensive task in itself. Each run ties up eight processors and 256 megabytes of memory for several days.

“Our quantum mechanical calculations of only the quinones currently use 27 functions to approximate the electron distribution around each atom,” Wheeler says. “Adding more functions — by adding more atoms or adding more functions per atom — would increase computer time by the fourth power of the number of functions. Computer time quickly becomes prohibitive for anything more than about a dozen atoms.”

The next step is to figure out how the forces from the electric fields of the quinone electrons push around the atoms in the protein. For this they use the same F=ma formula learned in high school physics to figure out how the forces from the electrons’ electric fields move the atoms around. But even these Newtonian dynamics are not simple. The protein consists of hundreds of amino acids. Even with some approximations, Wheeler’s simulation tracks the movement of 4,500 atoms. That takes weeks on the Power Challenge Array.

“We couldn’t do these calculations without the resources of NCSA,” Wheeler says. “We could use gigabytes of memory and weeks of time, even on multiple processors.”

And it is the high degree of detail that is producing results. They’ve already seen that the extra electron changes the quinone’s chemistry. The second quinone normally binds to the protein in one particular place. Add the electron and the quinone binds in a slightly different spot, 12-billionths of a centimeter away.

How much of a difference that shift causes is unknown. Wheeler says that’s the next step in the calculation: “Seeing how it affects the energy.”

Turning that idea around a bit, Wheeler’s simulations are now “shaking” the molecules in certain ways to induce electron transfer reactions. By making a molecule vibrate a certain way, perhaps by shining an infrared laser on it, “you might be able to put the electrons exactly where you want them,” he says.

If that idea — manipulating reactions electron by electron — ever makes it out of the computer and into the laboratory, it could prove chemistry’s equivalent of genetic engineering: not only understanding how nature works, but creating things that would never have existed.

This research is supported by the U.S. Department of Energy, the Oklahoma Center for the Advancement of Science & Technology, and the National Science Foundation. Previous support came from the National Institutes of Health.
Industrial Strength Antigols

Antibodies can bind to just about anything. Surprisingly, that makes them a potential source of new catalysts.

by Michael Schneider
viruses, bacteria, or other harmful invaders enter the bloodstream, the immune system fends them off by generating antibodies — proteins shaped to latch onto the intruders (analogous to a lock fitting a key) so that its ability to harm is immobilized. What if it were possible to harness this remarkable ability to manufacture proteins matched to the 3D features of minuscule intruders? What about commercial applications? Hmmm, said a few chemists about 30 years ago — what about using the immune system to generate catalysts, the substances that speed chemical reactions?

Laboratory work during the past 12 years has shown that this once far-fetched idea is more than an idea — catalytic antibodies work. "The immune system can produce antibodies that bind to almost anything," says UCLA chemist Kendall Houk. "It's really quite a marvel this way. The idea is to use these proteins to catalyze reactions; after all, nature uses proteins as enzymes to catalyze reactions in the body. Maybe we can use antibodies to catalyze reactions that don't have catalysts in nature."

For the pharmaceutical industry, catalytic antibodies offer promise for rational drug design — creating molecules with 3D features sculpted to interact with other molecules. "Lots of synthetic chemists are trying to develop methods to produce drugs with the proper stereochemistry, the right 3D arrangement in space," says Houk. "One possible use of catalytic antibodies would be to let nature produce these drugs."

Houk leads a group of UCLA researchers who tackle problems in organic and bio-organic chemistry, applying theory and computation in conjunction with laboratory work. Catalytic antibodies are one among a number of areas his group is interested in, with a focus on understanding the atomic-level details of why chemical processes work the way they do. "We've used the techniques of computational chemistry," he says, "to elucidate what's going on in this hybrid of chemistry and biology."

Although experiments with catalytic antibodies — by several research groups — have validated the basic idea, better understanding is required for it to have practical value, says Houk. "We're in a situation where we have exciting initial observations, but frustrating lack of success in getting high acceleration of reactions. They occur, but not well enough yet if we're going to use this for something."

With a series of computations on SGI's CRAY Origin2000 at NCSA, Houk's group applied quantum mechanics and classical molecular dynamics to illuminate how a particular catalytic antibody speeds up a well-known organic reaction, called the Diels-Alder reaction (see page 42). Their results — reported in Science (March 20, 1998) — provide detailed new understanding of the interactions. The calculations identify which 3D arrangement of transition-state molecules — out of four possible — involved in the reaction binds with the antibody. They also reveal for the first time that two separate hydrogen bonds — formed between the antibody and the transition state — work together synergistically to catalyze the reaction. "By doing calculations on how that antibody binding-site interacts with the possible transition states," says Houk, "that tells us specifically what's going on."
A shot in the dark

To catalyze a reaction means to make it go faster, and the speed of a reaction — chemists tell us — depends on its transition state. When molecules react, they go through in-between, transient stages of molecular transformation before arriving at the stable end-products, and a reaction's in-between molecular state of highest energy is its transition state. The higher the activation energy of this state — which means the harder this state is to form — the slower the reaction proceeds.

If you can make the transition state more accessible, lowering its activation energy, the reaction will go faster — and that's what a catalyst does. Much like an antibody latches onto a microscopic intruder, immobilizing its ability to do harm, a catalyst binds with a transition-state molecule, stabilizing it, making it more available to other chemical species that interact with it — thereby speeding up the reaction.

If it were possible to inject a reaction's transition state into the bloodstream (rats are used for these experiments), presumably the immune system could generate the appropriate protein catalyst. But it's not that simple. Since transition states are transitory, chemists must find a stable molecule that's a close cousin to the transition state they're interested in and use this "transition-state mimic" to stimulate the immune response, which, if all goes well, generates a catalytic antibody that works.

While there are many difficulties, several research groups have produced catalytic antibodies over the last decade. For the Diels-Alder reaction that was the focus of Houk's recent work, a group led by Kim Janda of Scripps Research Institute in La Jolla, CA, succeeded in 1995 at producing an antibody that catalyzed the reaction, an important development. Still, the laboratory success offered relatively little understanding.

"It's kind of a shot in the dark," says Houk. "We try to have something that's like the transition state, and we inject it into the rat, which does everything else, and something comes out. If we're lucky it works. To make progress, we need to better understand what's going on."

Quantum mechanics and activation energy

Another collaborator on the project, along with Houk's and Janda's teams, was Scripps molecular biologist Ian Wilson, who took a major step forward by determining the structure of a crystallized form of Janda's Diels-Alder antibody, known as 13G5. Working with this exact structural data, Houk's group undertook a series of computations to provide quantitative understanding of why 13G5 catalyzes one particular Diels-Alder reaction, as opposed to other similar reactions.

The first step was quantum-mechanical calculations (using GAUSSIAN 94) to determine the transition states involved in the reaction, information unavailable from the experiment. These calculations gave precise geometries for four different transition states, each of them a molecular variation on a theme — the same atoms in four slightly differing 3D arrangements.
To quantify how much acceleration of the reaction resulted from catalytic interaction, the quantum-mechanical calculations determined activation energies for the transition states. First, the researchers calculated activation energy for the reaction occurring by itself, without catalysis. Then further calculations quantified the effect of hydrogen bonds at particular atoms of the transition states, replicating interactions in the binding site of 13G5. These calculations showed a large reduction in activation energy (from 22.2 to 12.2 kilocalories/mole) resulting from synergistic interaction of two hydrogen bonds.

This was an interesting finding in itself, says Houk, which accounts for the catalytic effect of the antibody: "We found a new effect that people hadn't known about before, which is that if we put a hydrogen-bond donor on one part of the transition state, and a hydrogen-bond acceptor on another part, we get a very large transition-state stabilization. People are kind of surprised you could get this much speed-up from a catalytic antibody, and this is a significant discovery."

**Docking with the antibody**

Experiments showed that the catalyzed reaction formed only one product, so it was clear that 13G5 stabilized only one of the four possible transition states. To determine which one, Houk carried out a series of molecular dynamics computations (using software called AUTODOCK, developed by Arthur Olson at Scripps) to simulate each transition state docking in the antibody binding-site.

These calculations, necessarily less exact than quantum mechanics, track all the atomic interactions of both molecules. "They're gigantic calculations," says Houk. "We have the whole protein and the transition state binding with it, and we're calculating the energies of interaction between every pair of atoms."

For each transition state, Houk carried out a form of molecular dynamics known as simulated annealing, which computes the interaction energies at high temperature, where the atoms are more free to move around. As the computation then lowers the simulated temperature, the molecules settle into the favored low-energy arrangement.

"To make sure you've done it right," says Houk, "you do it over and over again, up to 100 times, and then look at these 100 different docking calculations, which will tend to show you the same arrangement repeating itself a high percentage of the time. This gives us confidence of what the preferred binding arrangement is."

These computations showed that the antibody stabilizes a transition state involved in a "disfavored" form of this Diels-Alder reaction, a result that suggests the potential of catalytic antibodies to shift the natural tendencies of a reaction. "When you run the reaction in solution," explains Houk, "you get predominantly one product. But the antibody specifically favors a different product. Normally, you'd be out of luck if you want this product, but with catalysis of the disfavored product, it becomes possible to get that selectively."

With this work, Houk's group opened a new door in the use of molecular dynamics. "The unique thing we did, which people haven't done before," says Houk, "is to use this kind of program to directly study the energy of interactions of a transition state with a protein. This is a new application for this kind of calculation."
Blueprints of the Protein Factory

As in all organisms, our proteins are assembled in intracellular factories called ribosomes. A new imaging method is producing some of the first interior views of this factory as well as details of how the nascent proteins are exported.

by Michael Schneider
All cells come from cells. It's a basic proposition of life. Each of us started as a single fertilized cell, a minuscule volume containing all the equipment needed to sustain itself, thrive and multiply, along with the instructions for creating the diverse, intricate multicellular organization of full-grown human life.

A cell is so small that not until 300 years ago, with the invention of light microscopes, did we learn that there was such a thing. Now we know this tiny volume, 1,000 times smaller than the tip of a ballpoint pen, itself contains an energetic, teeming world of small and large structures, moving parts, interacting subunits, a world easily more complicated and active than rush-hour Manhattan.

In the 1940s, a new tool, the electron microscope, gave scientists their first look at the details of structure within the cell. Since then biologists have pieced together a basic map of the intracellular world, one of the most essential components of which is ribosomes, the miniature factories that build new proteins.

Three years ago, using the electron microscope in a powerful new way — relying on computational methods to construct 3D images from 2D samples — Joachim Frank produced the first detailed, 3D picture of this protein factory. A senior scientist at the Wadsworth Center in Albany, New York, Frank in 1996 followed his 1995 breakthrough with detailed views of the ribosome's inner space, images that show clearly for the first time how transfer-RNA molecules interact with the ribosome as they bring amino acids to form the nascent protein. As a result of this work, in 1997 Frank received a seven-year award as a Howard Hughes Medical Institute investigator, one of the most prestigious grants available for biomedical research.

Last year, Frank ported his software system to NCSA, gaining the advantage of more powerful computing and, at the same time, making his imaging method more useful to others — through the infrastructure of the Alliance. Since then, using the SGI CRAY Origin2000, Frank and his colleagues put into place another piece of the ribosome jigsaw puzzle. Their recent results (Science, Dec. 19, 1997) trace the pathway followed by a freshly minted protein chain as it slips out the ribosome's back door — actually the exit from a tunnel — into a donut-shaped opening in the intracellular membrane known as the endoplasmic reticulum.

The key to these revealing — and to biologists invaluable — pictures of ribosome structure is single-particle reconstruction, a method of electron-microscopy imaging that Frank's team at the Wadsworth Center developed and refined over 20 years.

Crystallography without crystals

Most of the detail work of molecular biology comes from x-ray crystallography. Determining molecular structure in this way is a painstaking, iterative process that relies, to begin with, on being able to extract molecules from their watery, cellular environment and prepare them as ordered arrays of regular, repeating structure — crystals. James Watson and Francis Crick's breakthrough, determining the double-helix structure of DNA, for instance, started with x-ray diffraction data from a crystalized DNA molecule obtained in the early 1950s by physical chemist and molecular biologist Rosalind Franklin.
Large macromolecular structures like the ribosome, however, present especially tough cases. The ribosome is a combination of about 80 different proteins and ribonucleic molecules, 100 to 1,000 times larger than most proteins — and extremely difficult to crystallize. Conventional electron microscopy (EM) — has delineated the overall shape and outline of the ribosome, its surface topography, but this is a far cry from the atomic-level detail of crystallography. Frank's success has been to coax higher resolution from EM than previously possible.

The main obstacle to high-resolution EM is inherent in the technology; EM irradiates the specimen under observation with an electron beam, which damages molecules. "Macromolecules are very radiation sensitive," says Frank, "so we want to reduce exposure to the absolute minimum." This eliminates the usual approach to building a 3D image from 2D samples — doing multiple tilts of the specimen to capture it from multiple angles.

Frank's central insight was to devise a reconstruction technique that works with a single tilt. First, the researchers prepare a sample of ribosome particles in a liquid buffer, which is then frozen into a thin layer of ice. This cryo-EM preparation captures a natural distribution of orientations. "We use images of single macromolecules," says Frank. "They're randomly oriented, which implies that we get all the different aspects of that single particle at once."

The process focuses on a subset of particles that can be computationally classified as the same 2D view, distinguished only by rotation. When the specimen tilts, these subset particles no longer look the same, but since the tilt angle and their positions are known, there's enough information to put together a 3D image. "That experiment completely defines the geometry of this entire subset of molecules," explains Frank. "At once we get all these angles, and we can perform a reconstruction in a deterministic way."

**SPIDER and the donut-shaped pore**

The particle images are sorted out, digitized, and analyzed with SPIDER (System for Processing Image Data from Electron microscopy and Related fields), a software system developed by Frank's research team over the last 22 years — and now used by many other research groups. Each field produces a few hundred separate particle images, not enough statistically to assure views from all possible directions. Each reconstruction therefore requires multiple micrographs, producing thousands or tens of thousands of separate images for SPIDER to classify.

Through iterative computations, SPIDER refines the initial reconstruction to higher resolution. "The refinement is done in cycles," says Frank, "and for each cycle we have to compare predicted projections by cross-correlation with the experimental projections. We get better and better angles, and then we perform another 3D reconstruction, and so on. It's an enormous computational expense."

By transferring most of the computational load to NCSA's Origin2000, Frank's team gained the processing power needed for larger-scale reconstructions, and hence, better resolution. SPIDER is fully parallel in all its numerically intensive operations, taking advantage of the Origin2000's shared memory parallel architecture. Because their onsite system was an SGI Onyx Reality Engine, porting the code, says Frank, was "almost trivial."

The latest published results, obtained in collaboration with Günter Blobel's group at Rockefeller University in New York, show how ribosome binds to a donut-shaped pore, called the protein-conducting channel (PCC), in the endoplasmic reticulum (ER). Working with 9,571 separate particle images, SPIDER used up to 12 Origin2000 processors over one-and-a-half days of processing time, to reconstruct the ribosome-PCC complex. The resulting 3D image distinguishes details 20 angstroms in size, one of the highest resolution EM images of large-scale macromolecular structure yet obtained.
Frank's prior work showed details of the ribosome's "inter-subunit space" — the interface between the two subunits, called the large and small subunit, that comprise the ribosome. The Origin2000 reconstruction of the ribosome-PCC connection shows the donut-hole of the PCC aligning perfectly with the tunnel exit from the large subunit, forming a continuous passageway for a nascent protein to slide out the ribosome into the ER membrane.

"This is the first real evidence that the tunnel, which we'd already looked at, must be involved in the export of the polypeptide chain," says Frank. "We had only indirect evidence for this. We knew the tunnel went from inter-subunit space to someplace on the outside, and it had roughly the right length. But this is the first finger that points to it. Seeing the tunnel aligned with the protein-conducting channel is beautiful confirmation of that hypothesis."

**Picture stories of protein translation**

Producing proteins, made to the order of our DNA, is the fundamental, bottom-line business of the cell. Biologists call this process protein translation, and ribosomes are the assembly lines where this happens. Raw materials enter (amino acids carried by transfer-RNA), assembly plans (messenger-RNA) are deciphered and followed, and new protein chains get put together, one amino acid at a time, until the assembly instructions say stop and the new protein rolls out the door.

Improved knowledge of protein translation gained from Frank's detailed pictures are certain to pay off in many ways impossible to predict. One promising avenue for applying this research is bacterial resistance to antibiotics, a current, pressing public-health concern. Some of the most potent, useful antibiotics interfere with the ribosome of bacterial cells, killing the germs by shutting down their protein factories. Over time, the bacterial DNA responds, mutating to create a resistant strain.

"Before there is a detailed understanding of how the whole translation process works," explains Frank, "the fight against drug resistance is somewhat of a random approach. If you have detailed understanding, you know exactly what you can do — how to develop a drug, for instance, that provides multiple interference with ribosomal function, so that a single step of circumvention by the ribosome could not by itself produce drug resistance."

Other research groups are using crystallographic and nuclear magnetic resonance imaging methods to determine the molecular structure of ribosomal proteins, individual components of the ribosome composite. The larger-scale work of Frank's group will eventually intersect with these atomic-level efforts. The Wadsworth team has completed a 15-angstrom ribosome reconstruction and is now working toward a 12-angstrom resolution. When they reach about 10 angstroms, says Frank, the EM image will begin to show individual protein helices. "At that level, it becomes possible to fit existing x-ray structures into our EM map, and we'll start to build an atomic-level map with EM as a framework."

Until atomic-level resolution is achieved, Frank's team will continue to push toward finer detail and better knowledge. His first published ribosome reconstruction came from 4,300 separate particles. In current work, relying on NCSA's computational resources, he's using tens of thousands.

A related goal involves using the visualization resources of the Alliance to depict the translation process as a time sequence: computer-generated movies can show the ribosome up close in 3D from multiple angles, with start-stop motion in millisecond time-steps — a guided tour of the protein factory at work. ■

This research is being supported by the Howard Hughes Medical Institute and by grants from the National Institutes of Health.

**ACCESS** Fall/Winter 1998
Striking Gold in Mountains of Data

The rush is on. With data mining techniques developed at NCSA, companies are discovering valuable nuggets of information buried in mountains of data.
Data are piling up—in company files, in government databases, in scientific labs. Doubling every 20 months, by some estimates. When NASA's Earth Observing System swings into operation by the end of this decade, it will transmit more than two terabytes (trillion bytes) of data—the equivalent of two million books—to Earth daily. At that rate, a single day's worth of data will occupy a researcher for several years. Wal-Mart, the merchandising giant, records more than 20 million sales transactions each day. Every swipe of your credit card or beep of the checkout scanner adds more bits to business's cache of sales data.

With this exponential growth in data, the "information at our fingertips" has too often meant a flood of data rather than a stream of knowledge.

Data mining tools and techniques being developed at NCSA are helping change that. These algorithms and techniques are not only expanding the realms of science but also strengthening the economy through projects with the center's industrial partners. These data mining tools are finding innovative applications in marketing, sales analysis, fraud detection, quality control, and investments, to name a few.

These tools are uncovering patterns and detecting trends hidden in data. The tools are, quite simply, turning information into knowledge and data into dollars.

What is data mining? or, Teaching machines to learn

Data mining is a generic term for all the new—and not so new—computational techniques and methods for extracting useful information from billions of bits of data. It is actually one step in the larger process of knowledge discovery, whereby data are collected, cleansed of errors, and warehoused in any of the new unified computing environments that make it easier to keep more data online and accessible. Data mining is the step at which specific algorithms are applied to extract useful information from the data.

Most data mining tools rely primarily upon a short program that says, "If X happens, do Y." You've used one such program if you've queried a database or searched the Internet. Businesses are turning to them to predict customer purchasing patterns or to detect fraud. These data mining tools comprise the bulk of the burgeoning business in data mining products. An article in the December 1996 issue of Application Development Trends quoted Meta Group, a Stamford, CT-based consulting firm, as projecting that sales of data mining products and services would total $80 million in 1996. That market is predicted to reach $800 million by 2000.

Among the Rolls Royces of the data mining tools are algorithms that grew out of 1980s work in artificial intelligence. These are dubbed "machine learning" algorithms because they are primarily adaptive systems that can learn on the job. Some people refer to them as ways of teaching machines to think. All of these algorithms rely on an iterative process through which the programs learn from examples.

If these sound sophisticated, you're getting the picture. These aren't your standard query-and-report tools for tracking sale trends or for informing managers when to replenish stock. Nor are they simply the just-in-time kind of data processing required for managing an organization's day-to-day operations.

Machine-learning algorithms can be applied for tactical and strategic decision making. NCSA researcher Michael Welge is doing so to tease out relationships and discover patterns that will be the foundation of decisions affecting both the near-term and long-range plans of a company.
"We are actually applying data mining technology to problems that have impacted companies' bottom lines — either in helping them make better decisions as decision-support tools or actually reducing costs," says Welge.

Welge and other NCSA data mining experts are building the data mining algorithms that often find patterns so subtle as to be unnoticed or that perform with a degree of accuracy unattainable with off-the-shelf programs. They are tackling the difficult problems — the ones others have just thought about.

"We're at the leading edge of this stuff," says Welge. "What we are after is an additional 4 to 8 percent accuracy. It doesn't sound like much, but it is. Think about it this way: if you were a portfolio manager and I could give you an extra 4 percent accuracy in your predictions and you're managing $20 billion, you would be very pleased. If scientists have 4 percent more accuracy in a DNA base, they are very pleased."

Cluster, classify — technology for data mining

Squeezing performance from data mining tools doesn't require writing new algorithms so much (though Welge's group does that, too) but customizing existing ones and implementing them in novel, hybrid ways. Many of the machine-learning algorithms the group adapts have been tested on small datasets — maybe 250 records with only ten attributes, or fields that characterize a dataset (typical attributes in a financial record, for instance, are name, Social Security number, and credit card totals). The added value that Welge's group brings to these tools is to extend their capabilities for very large datasets and to build scalable approaches; that is, for these tools to operate as efficiently on desktop computers as on large computing systems. "We're talking about company databases that may be half a terabyte in size, with maybe 500 million records and 300 attributes."

Some databases are not large yet still complex. To understand what sets apart Welge's work from others, you have to understand the basics of data mining.

Data mining problems fall into one of two categories: clustering or classifying. Clustering discriminates among data so that they can be divided into groups. Often companies are simply looking for interesting patterns in their data or more efficient ways of organizing them. Postdoctoral student David Tcheng says the goal is often to please the customer, and elicit the Eureka response: "Ah ha, I've never seen that before."

Whereas clustering identifies groups into which data fall, classifying predicts which group particular data will fall into. "In classification you want to take a set of data and separate the fruit from vegetables, based on predetermined characteristics," says Rick Kufrin, a research programmer at NCSA who collaborates with Welge. (Kufrin is optimizing several algorithms for parallel computing architectures so that codes written for workstations can be scaled to run on high-performance computers.) "Clustering helps you separate fruit from vegetables, then determine what distinguishes the two groups. How is an apple different from a green bean? How is a green bean like an ear of corn?"

Naturally algorithms cluster and classify data in different ways. All, however, involve some kind of iterative process so that the models they generate "learn through example." Take, for instance, a neural net algorithm.
Classifying — supervised learning

Neural nets, which mimic the processes of the brain, can be trained to separate fruits from vegetables in ways surprisingly similar to teaching a child to distinguish a cat from a dog; that is, with examples. With the child, you say “dog” as you point to malamutes, beagles, and collies. You say “cat” as you point at Siamese and tabbies. All the while the child is subconsciously making distinctions — a cat has whiskers and meows; a dog has a snout and barks. The more examples you give the child, the more quickly he or she learns the difference between the two as well as which characteristics are more discriminating than others; for instance, a meow is a more definitive indicator than tabby coloring in identifying that the furry animal lying near the TV is a cat.

In this process of discrimination, the child’s brain assigns greater significance to some characteristics than to others so that it can evaluate and categorize new information. A similar process of discrimination also occurs in machine-learning models.

The neural net receives inputs and outputs. Using the dog and cat example from above, the inputs might be attributes such as whiskers, size, voice pitch, and hair length. Each attribute is connected to the output — cat and dog. To train the model, mathematical weights are assigned to each connection between an attribute and each animal. Each time the model correctly associates an attribute and animal, the value of the weight is increased slightly in the direction calculated to reduce error. The weight is decreased if the association is incorrect. After repeated examples, these weights come to reflect the probability that a particular input will correspond correctly to an output. The process is often more complicated than that but the principles are the same.

“When the model sees another one, it adjusts the weights to reinforce that right answer, and so on,” says Welge. “The model sets weights and reinforces characteristics in such a manner that when it is given something new, with a certain level of accuracy, it will be able to make predictions.”

To predict whether, say, an auto accident is staged, a computer scientist using neural nets to classify data would begin with examples of fraudulent and nonfraudulent accidents. Each accident may consist of thousands of attributes — age and sex of the driver, time it happened, whether or not the driver called a lawyer, whether he called three lawyers, whether the lawyer was called before the doctor, and so on. Algorithms would then assign weights to generate accuracy as a percentage; that is, whether an example is 70 percent likely to be fraudulent and 30 percent likely to be nonfraudulent.

Because examples are used to train the model to obtain a desired outcome, this type of machine learning is called supervised or directed. When the goal is more abstract, as when clustering data, it is called unsupervised or undirected learning.

Clustering — unsupervised learning

Neural nets also offer good examples of clustering data. Look, for instance, at the warranty modeling done for Caterpillar Inc., a manufacturer of earthmoving equipment and an NCSA industrial partner.

Caterpillar researcher Syamala Srinivasan approached Welge about designing a data mining system that would help managers in the company’s 15 engine divisions more quickly assimilate the information in the voluminous weekly warranty reports. It is “more information than anyone can possibly absorb,” says Srinivasan. Every Friday each manager
receives a 2.5-inch thick warranty report containing copies of the more than 4,000 claim reports filed that week at the company’s network of 192 dealers. Managers scan this tome to detect trends in engine problems. They catch the typical engine ailments but Caterpillar would like to zero in on minor problems indicative of new or more significant problems to come.

Welge calls this warranty issue a classic data mining problem. “They wanted a good organization of everything. They were getting reports of failed engines. They wanted to view their data from a perspective they hadn’t viewed it before. And they wanted to use tools to help them do that. There really was no preset goal, as in classification.”

Working with Hillel Kargupta at Los Alamos National Laboratory and Hsinchen Chen at the University of Arizona, Welge’s group attacked the problem in three ways. Kargupta developed the encoding scheme that translated the warranty reports into binary numbers the computer could read. This encoding scheme had to handle misspellings and grammatical errors as well as work in any language.

From Chen, Welge’s group obtained a modified version of the Kohonen Self-Organizing Maps (SOM), which was geared toward clustering of the warranty data. (SOMs originally sprang from the work of Teuvo Kohonen, a professor of computer science at Helsinki University of Technology.) SOMs map, or translate, documents that are represented as vectors in high-dimensional space to lower-dimensional space — such as a 2D grid — so that similar documents in high-dimensional space are clustered together and similar clusters migrate together. The SOM notes these locations and labels them as correct spots for a related attribute. As the model accumulates feedback about correct responses, it adjusts its weight vectors so that similar documents are placed in similar places. It is parallel to tossing a baseball into the air and deciding that wherever it lands is home plate. From then on, the accuracy of all other tosses are weighted by how close to or far from this mark they land.

Welge’s group assembled the pieces and built the last two pieces of the model. Computer science graduate students Jason Hibbler and Nina Mishra applied algorithms to first cluster the output, then visualize the clusters of data. The result was an interactive 3D graphic.

The 4,000 warranty reports were replaced with 200 color-coded columns organized into regions summarizing engine problems. Three colors are associated with particular regions of the engine: yellow for the heads, blue for the pistons, and green for the cylinders, making it easy to identify engine problems at a single glance. The taller the column, the larger the cluster. The closer to neighboring columns, the greater the overlap in problems. By “drilling down” into each column, a manager can review every report or simply look at a single representative sample.

Clustering techniques also work for managing risk — against either unprofitable research ventures or even losses in the stock market. Marketing managers like the technique for identifying customer populations they may want to target, such as with a special promotion. These and other neural networks are by far the most popular learning algorithms because of their adaptive, predictive capabilities. They are not, however, the only ones, nor the only ones Welge’s group works with. In fact, central to scaling up algorithms is choosing the best algorithm or combination of algorithms for a problem.
One algorithm won’t do it all

Sometimes the simplest solution is the best.

Recently Welge’s group designed a rules-based system for Sears, Roebuck and Co., an industrial partner with NCSA. Though few companies like to admit their employees are stealing from them, it is a major problem that costs the retail industry $27 billion a year. For Sears, implementing a neural network would have required that the company collect thousands of examples of fraud — a task that would have consumed a year. A rules-based system, on the other hand, adapted the rules the company’s fraud experts used to detect theft into computer code that implemented those rules. The system could detect employee theft by analyzing cash register receipts for abnormalities. Though not as glitzy as a machine-learning algorithm, the system was effective because Welge’s team took the time necessary to thoroughly understand the problem. Also, the code that NCSA’s Tcheng wrote was able to handle the 50 to 60 gigabytes of sales data that had to be analyzed weekly. It worked; people were caught. Says Sears’s technical consultant Chris Herr: “NCSA accomplished in six months what another company failed to accomplish in two years.”

Sometimes a problem requires several algorithms.

Most people practice the winner-takes-all approach to data mining, says Welge. That is, after running several algorithms on a dataset, they select the algorithm that performs best overall. Welge’s group often takes a hybrid approach. Rather than applying one algorithm to an entire problem, it splits the problem into segments that can be attacked by whichever algorithm is most suitable for that segment. This approach is similar to hiring different chefs to prepare each course in a seven-course meal rather than relying on the pastry chef to also master the bouillabaisse.

Some problems contain steps that are both linear and nonlinear — that is, they contain data that change in simple to complex ways. “A nonlinear model would be the price of a given stock over time — every day the price changes but in a way that is not a straight line. If it were linear, and I knew that on day two the price was $2, then on day three I would know it was $6,” says Welge.

Sometimes solving a problem requires a mix of commercial software and customized code.

One-of-a-kind systems are expensive to build and maintain. Companies hesitate to abandon the computing systems they’ve pieced together over decades. Nor do they always have staff skilled in the programming languages of newer codes. Companies want a system that not only finds the information they seek but is also easy to maintain. That’s why NCSA is converting Sears’s rules-based system from C code running against flat files into a simpler SQL (standard query language) running against the database.

“The original C program was part of proof of concept: Can we do this?” says Sears’s Herr. “Now the question is: How do we make that into a production system that we can run at Sears as opposed to something that requires the resources of NCSA to run?”

The systems employed by Welge’s group clearly aren’t for everyone. When businesses first apply data-mining techniques to a problem, they are likely to get a big bang for their buck from generic programs. Only when they need better performance will they reap benefits from Welge’s custom fit.
Expanding Alliance resources to keep pace with User demand

University of New Mexico/Maui High Performance Computing Center
- IBM SP
  - 603 nodes

NCSA
- SGI CRAY Origin2000
  - 768 processors
- HP-Convex Exemplar SPP-2000
  - 64 processors
- NT Supercluster
  - 192 processors

Alliance visualization and virtual reality environments
Alliance computing resources

vBNS network*
- OC-12
- OC-3
- DS-3
- Alliance sites
- Backbone nodes
- Connection sites
- New connection awards

*The vBNS is an advanced network funded by the National Science Foundation to support scientific research at U.S. research and academic institutions. The vBNS serves as the high-speed backbone of the Alliance and will extend to nearly every partner site by the end of 1999.
Requests for Computing Allocations
Researchers and educators at U.S. institutions are eligible to apply for time on the Alliance's computing resources. Instructions for submitting a proposal for a peer-reviewed allocation can be found at http://www.ncsa.uiuc.edu/alliance/applying.

University of Wisconsin at Madison
Condor Flock
600 workstations

Boston University
SGI CRAY Origin2000
192 processors
SGI Power Challenge Array
38 processors

University of Kentucky
HP-Convex Exemplar
SPP-2200
64 processors

Disciplines Represented in Large Academic Projects at NCSA, the Alliance's Leading-Edge Site

- Computer Science
- Mathematics
- Physics
- Atmospheric Science
- Engineering
- Materials
- Biology
- Astronomy
- Chemistry

Large Projects are >5,000 NUs Annually

(1 NU = 1 Cray X-MP CPU hour)
Researchers have witnessed extraordinary advances in science and engineering since the advent of supercomputing. What can they expect from the Grid? According to researchers such as Glen Wheless and Cathy Lascara from Old Dominion University in Norfolk, VA, the Grid will mean more discoveries — faster — and with the help of colleagues in scientific disciplines that never crossed paths.

Wheless and Lascara are physical oceanographers modeling the intricate biological and physical processes within the Chesapeake Bay, the nation’s largest estuary. Their model depicts how winds, tides, geography, water temperature, and the different salinity levels of the ocean and fresh water influence how propellerless larvae are transported about the estuary of the Chesapeake Bay and, in the case of some species, back to the coastal shelf. Declines in larval populations are cited as one of the reasons for plummeting fish catches in the bay. Oyster harvests that peaked at 15 million bushels in 1884 are now around 10,000.

Five years ago, Wheless and Lascara teamed with researchers from two future Alliance sites — NCSA and University of Wisconsin at Madison — to create a virtual reality model of the bay so that they could better analyze the causes of the declines. The results were spectacular. The software they borrowed from Wisconsin called CaveSD — named after the roomlike virtual reality environment in which the data are often viewed — re-created the bay in their laboratory so that they could seemingly soar above the bay and dive below, searching for patterns. When Wheless lingered in the first few meters of the bay, he identified the previously underestimated influence of wind, tides, and runoff on the movement of larvae within the estuary and the nearby coastal shelf. Their model showed how these factors churn the waters in the shallow bay (which averages only 20 feet deep), often carrying larvae to the oxygenless lower depths, where they perish.

That was only the start. The visualization program upon which CaveSD was built, Vis5D, written by Wisconsin’s William Hibbard and Brian Paul, is used by hundreds of weather modelers worldwide for simulating the ocean and the atmosphere. Recognizing the benefits to their research and those of others if these data could be integrated into a single model, Wheless and Lascara began modifying CaveSD to be more compatible with all data visualized with Vis5D (as it was, it required extensive modifications to accommodate each dataset). This past year, they released a generic version of CaveSD so that other researchers could import their data directly into virtual reality. Their longer-term goal is to integrate Cave5D into an Internet-based environmental hydrology "workbench" that they are building with other members of the Alliance. Last year, Wheless and Lascara, working with Donna Cox, Robert Patterson, and Stuart Levy from NCSA, incorporated NCSA’s Virtual Director software into the model so that researchers can view and interact simultaneously with the same simulation — even if they are using virtual reality systems located in other states.

These efforts are precursors of how the Grid will change the collaborative way in which scientists work.
What, exactly, is the Grid?

A computational infrastructure. Cyberspace on steroids. A virtual machine. Imagine taking apart various desktop computers and souping up each of the components — memory, CPU, graphics devices — so that they are thousands of times faster and orders of magnitude more sophisticated than they are now. When you connect them via a network that shoots data across the country at speeds approaching that of light, what you end up with is no longer a computer but a portal to a flexible and powerful computing environment. That's the Grid.

Or, at least, that is the Alliance's vision of the Grid.

What exists now are the pieces. The Alliance is building the system of software that will integrate this ensemble so that its many and varied pieces will operate as if they were one. We know this type of distributed computing can work today, on individual heroic projects — the challenge is to make it work seamlessly, efficiently, and routinely so that it becomes as ubiquitous and encompassing as the electrical power grid is today.

When you compose a letter on a computer, hundreds of software components are, in some way, involved in this simple task. There are components that change the size of type, check grammar, format the layout. Others recognize the printer you are using and attach the bits and bytes that instruct your printer what to do. Still others — the operating system and print scheduler — make sure these bits reach their destination. If you diagrammed the relationship among these components, you'd wind up with a pyramid that, in many respects, resembles a corporate organizational structure. At the top, in the CEO's office, is the word processing program as it appears on your screen. Underlying it are the divisions, or group of programs, that encapsulate still more groups that perform related but distinct functions. The further down you drill into any of these groups, the more specific become the tasks that each software module executes.

The software schema for the Alliance's National Technology Grid is a similar cascade of hierarchical layers. The base is the hardware — computing resources, networks, and vast distributed sets of data. But overlying it are three interdependent layers of software in which specific functions are being bundled to perform progressively more complicated tasks.

These divisions are, of course, artificial because unlike conventional software architectures, the Grid's system is flexible. Researchers will assemble functions in whatever way best fits their needs. The system must also accommodate the dynamic nature of the Grid. Bandwidth will fluctuate, computers will be upgraded or unavailable, software will be modified. The software system, however assembled, must deliver reliable, high performance without the user needing to be aware of these continual changes to the underlying system.
Applications answer the question: Why do we need a Grid? Most of these applications now stem from science and engineering, but eventually they will encompass all areas of society. The Alliance is focusing on six data- and compute-intensive areas: chemical engineering, cosmology, environmental hydrology, molecular biology, nanotechnology, and scientific instrumentation. Other applications such as collaboration, knowledge management (or data mining), and visualization are also helping define the performance requirements of the Grid.

Programming Tools make it easier to develop programs for the Grid by masking some of the complexities of the underlying structure. Some of these tools, such as application toolkits, provide flexible frameworks for assembling applications from software modules, some of which may be located on computing systems at other institutions. Newer parallel programming languages, libraries, and compilers map applications onto distributed resources in ways that compensate for what are often dramatic differences in power, architecture, and data representation. Object-oriented techniques and commodity technologies, too, can simplify programming for the Grid.

Services, or middleware, are similar to the operating system on a conventional computer. They are the intermediaries to the hardware and provide such basic services as security, authentication, resource scheduling, and assurances as to the quality of service. The software encapsulating these services will be designed in such a way that the programming tools and applications dependent upon them can assess and dynamically respond to changes in the Grid.

Physical Resources are the computers, visualization environments, mass storage devices, and networks associated with high-performance computing. The specialized software associated with each resource can be made more Grid compatible. For a detailed description of computational grids, read The Grid Blueprint for a New Computing Infrastructure, Ian Foster and Carl Kesselman, eds. (San Francisco: Morgan Kaufmann, 1999).
Technology Deployment

Remy Evard*
Argonne National Laboratory
evard@msc.anl.gov

John W. D. Connolly
EPSCoR and University of Kentucky
connolly@ukcc.uky.edu

Ken Kennedy
Rice University
kenn@cs.rice.edu

George Lake
University of Washington
lake@astro.washington.edu

Miron Livny
University of Wisconsin at Madison
miron@cs.wisc.edu

Rick Stevens
Argonne National Laboratory
stevens@mcs.anl.gov

John Towns
NCSA
jtowns@ncsa.uiuc.edu

Paul R. Woodward
University of Minnesota
paul@lcse.umn.edu

*Team Leader(s)

User Training

Frank L. Gilfeather*
University of New Mexico/Maui High
Performance Computing Center
gilfeath@arc.umn.edu

Leslie Southern
Ohio Supercomputer Center
leslie@osc.edu

John Towns
NCSA
jtowns@ncsa.uiuc.edu

John M. Unsworth
University of Virginia
jmu2m@virginia.edu

NCSA — Leading-Edge Site

Larry Smarr, director
Alliance/NCSA
ls@ncsa.uiuc.edu

Jim Bottum, deputy director
jb@ncsa.uiuc.edu

John Toole, deputy director
jtoole@ncsa.uiuc.edu

Charlie Catlett, senior associate director
for Science and Technology
catlett@ncsa.uiuc.edu

Melanie Loots, senior associate director
for Institutional Relations
mlotts@ncsa.uiuc.edu

Tom Prudhomme, senior associate director
(acting) for Institutional Relations
tip@ncsa.uiuc.edu

John Towns, senior associate director
for Scientific Computing and Visualization
jtowns@ncsa.uiuc.edu

*Team Leader(s)