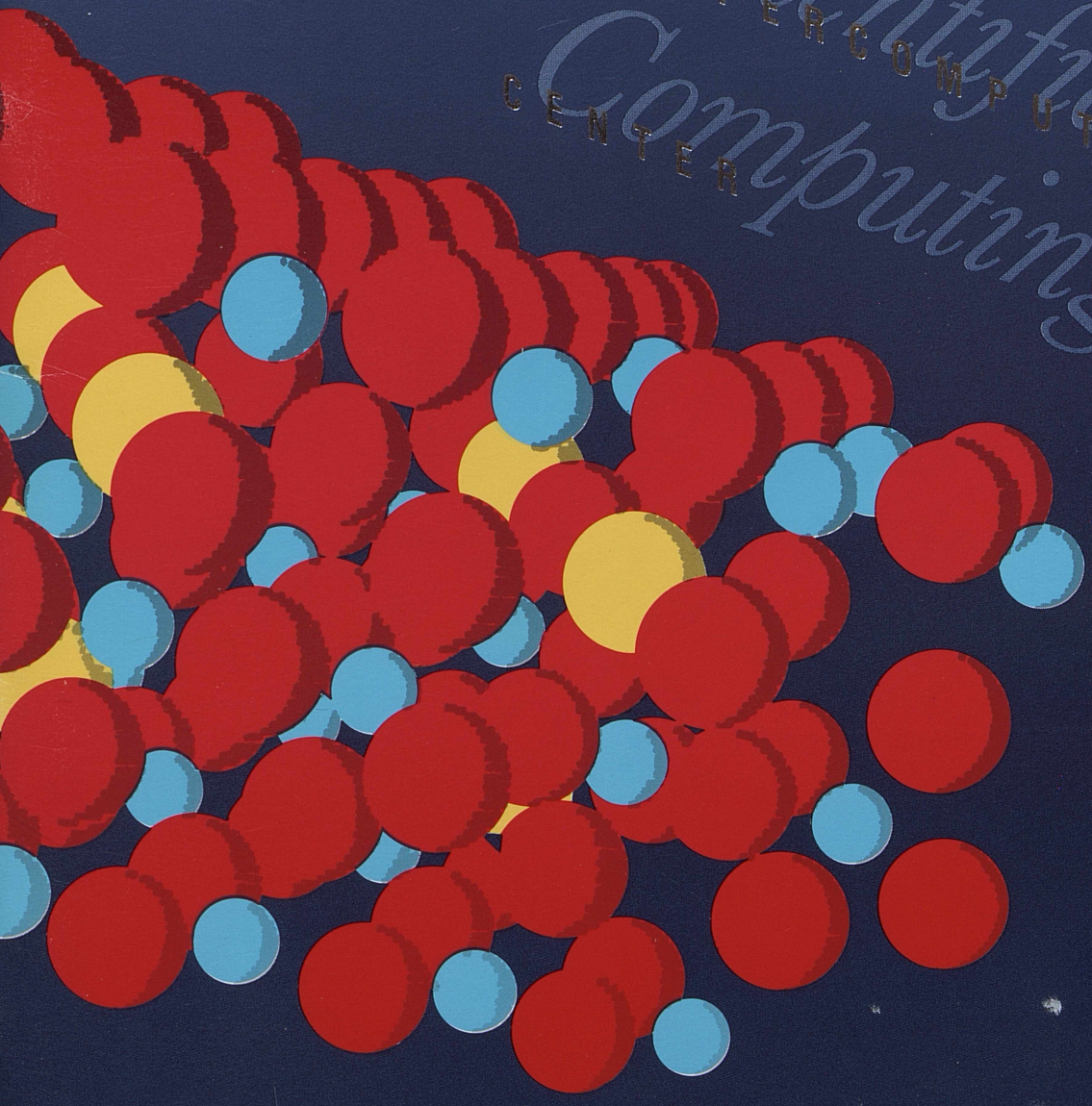
Projects in Scientific Computing

PITTSBURGH SUPERCOMPUTING CENTER



1987-1988

Projects in Scientific Computing 1987-1988

Published June 1988 by the Pittsburgh Supercomputing Center
Mellon Institute Building
4400 Fifth Avenue
Pittsburgh, PA 15213
(412) 268-4960

The cover image is Jeff Lunderstadt's rendering of the perovskite crystal structure which is characteristic of high T_c superconducting materials. Lunderstadt worked from an image which Doug Fox and Joel Welling of the Pittsburgh Supercomputing Center prepared on the Center's Pixar Image Computer.

Foreword

Very large scale computation is coming to full bloom as a research tool in most fields of science and engineering. A qualitative transition is taking place in the use of computation as the capability of the largest machines, the supercomputers, grows to be great enough to allow researchers to turn from using them in simplistic model calculations to properly addressing realistic problems. Put simply, the increasing strength of their computational tools is changing the types of problems on which scientists are willing to work.

Mike Levine and, briefcase in hand, Ralph Roskies.

Recognizing the need to make advanced computational tools available to the members of the national, university based research community, the National Science Foundation helped to create the Pittsburgh Supercomputing Center. Through this and other Centers, the NSF is acting to assure the continuing availability of this qualitatively different form of computational resource for those advanced projects which require it.

Put simply, the increasing strength of their computational tools is changing the types of problems on which scientists are willing to work.

In this booklet, we present brief descriptions of a small sample of the research done over the past year by members of the national scientific community using the facilities of the Pittsburgh Supercomputing Center. This research spans numerous fields: solid state physics studies on the new, high temperature superconductors; theoretical particle physics studies of the structure and interactions of the tiniest known building-blocks of the physical world; work in biomedical engineering which aims to improve the state of the healing arts; work in economic science which seeks to better understand a sector of our national economic structure. We hope that these brief introductions will help those outside of these fields to appreciate the advances being made through the use of supercomputing techniques. We also hope that these examples will encourage other scientists and engineers to consider the potential benefit of using these tools in their own research.

Michael J. Levine, Scientific Director

Michael J. Levine

Ralph Z. Roskies, Scientific Director

Ralph Roskies

CONTENTS

The Pittsburgh Supercomputing Center at Two Years iv

Projects in Scientific Computing, 1987-88

1. MATERIAL SCIENCE

Superconductivity, a Preface with John K. Hulm 1

Oxygen Vacancies and Antiferromagnetism in Lanthanum Copper Oxide
Philip A. Sterne & Ching-ping S. Wang, University of Maryland 2

Commentary: Ab Initio and Computing the Properties of Solids 5

Defects in Molecular Solids and a Pseudopotential for Superconductors
J. D. Joannopoulos, Massachusetts Institute of Technology 6

AC Conductivity in Substitutionally Disordered Alloys
Arthur J. Freeman & Antonios Gonis, Northwestern University 11

2. HIGH-ENERGY PHYSICS

Supercomputing in the Particle World
Quarks, Gluons, Bosons, Leptons and Other Creatures 14

Simulating the Strong Force
The Quark-Gluon Plasma Phase Transition of QCD
Rajan Gupta, Los Alamos National Laboratory
Gregory W. Kilcup, Brown University
Stephen R. Sharpe, Stanford Linear Accelerator Center 16

Observing the Strong Force
HELIOS: Smashing Quarks to Plasma
Julia A. Thompson & Wilfred E. Cleland, University of Pittsburgh 18

Simulating the Weak Force, With Bosons and Quarks
Interactions in Lattice Gauge-Higgs with Fermions
Robert E. Shrock, State University of New York, Stony Brook
I-Hsiu Lee, Brookhaven National Laboratory 20

3. ASTROPHYSICS

Supercomputing in the Cosmos
Sinking Satellites of Spiral Galaxy Systems
Peter J. Quinn, Space Telescope Science Institute, Baltimore
Lars Hernquist, Princeton Institute for Advanced Studies 22

4. THEORETICAL ORGANIC CHEMISTRY
Quantum Chemistry and Molecular Mechanics
Rotational Barriers in Substituted Compounds
Kenneth B. Wiberg, Mark A. Murcko & Keith Laidig, Yale University 26

5. MOLECULAR BIOCHEMISTRY
Molecular Dynamics and the Patient Crystallographer
Protein Structure Refinement by Simulated Annealing
Axel T. Brünger, Yale University 29

Molecular Dynamics and Drug Design
Free Energy Difference in Enyzme-Inhibitor Interactions
Bogdan Lesyng & Edgar F. Meyer, Texas A. & M. University 32

6. BIOMEDICAL ENGINEERING
Supercomputing and the Healing Arts
Finite Element Modeling of Strain-Induced Bone Remodeling
Richard T. Hart, Tulane University 35

7. ECONOMIC SCIENCE
Supercomputing and Wall Street
Dimensionality of the Covariance Matrix for Stock Returns
Keith T. Poole & Scott F. Richard, Carnegie Mellon University 38

8. COMPUTER AND INFORMATION SCIENCE ENGINEERING
Microtasking for Better Supercomputing
"Polite" Multitasking and Improved System Throughput
Michael Bieterman, Boeing Computer Services 42

Wendy Janocha with documents in hand is a familiar sight at PSC. As Allocations Coordinator, Wendy answers questions about research proposals and due dates. With her is Lori Graul, who maintains an abstracts data base for research at PSC.

The Pittsburgh Supercomputing Center at Two Years

"Has it really been two years? Already?" Vivian Benton is one of the "veterans" at the Pittsburgh Supercomputing Center. She's been coordinating documentation since before the Center opened officially two years ago, June 1986, and she edits the **PSC News**, a monthly newsletter that reaches more than 2000 people in the supercomputing community.

"Veteran" isn't quite the right word for anyone at a place this young—and still growing steadily. The staff numbers about 60 now, which is quite a change from that first half-dozen or so: Joel Welling, Doug Fox, Scientific Directors Mike Levine and Ralph Roskies, and Jim Kasdorf at Westinghouse, who have been with the Pittsburgh Supercomputing Center since before there really was one. "We're still getting to know each other," adds Executive Director Beverly Clayton, another "veteran." Beverly arrived in April 1986, when testing of the CRAY X-MP/48 was just beginning, bringing with her 20 years of systems and management experience at Gulf Oil.

Beverly Clayton, Executive Director, and Jim Kasdorf, Manager of Engineering Computing Services at Westinghouse Energy Center, site of PSC's CRAY X-MP/48.

"There's a sense of purpose. People feel they're doing something worthwhile. It begins with Mike and Ralph," says Debbie Nigra, a veteran User Consultant. Roskies of the University of Pittsburgh and Levine from Carnegie Mellon University are physicists—practicing, committed scientists and teachers, who understand the importance of supercomputing in scientific research.

"Supercomputers will allow us to ask questions that couldn't have been asked before; they are helping to create a new kind of science," wrote Roskies in his 1987 article, "Science and Supercomputers."

PSC's CRAY, one of the most powerful computers in the world, is vital to the Center's purpose; yet supercomputing here is much more than equipment. It is dedicated professionals. Education Coordinator Janet Brown organized the Center's well-received 1987 Summer Institute, and she's at it again this year. Brown has held several positions in advanced computing: "This is the most talented group of people I've worked with." Others echo her sentiment.

The managers, Marvin Zalevsky, Bob Stock and Gene Hastings, bring backgrounds spanning industry and academia. Stock, Manager of User Services, like Clayton came from industry—twenty years in management and user services at J & L Steel. "They keep us communicating and in touch with our overall goals," comments Phil Andrews, Scientific Specialist, who joined the PSC during the summer of 1986. Andrews this year developed the integrated graphics package GPLOT.

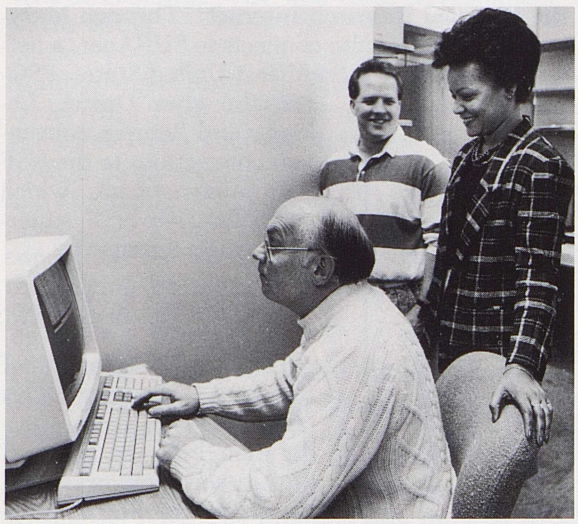
"We've accomplished many things in two years," says Levine, "but we're working to get better. We want to improve our telecommunications links, and our user documentation, which is already extensive; we want to document the research that's done here. We want to provide quality graphic output capability and a smooth transition to UNICOS; and we want to improve efficient use of our hardware."

Two years? Already? Funded in January 1986 by the National Science Foundation with matching funds from the Commonwealth of Pennsylvania, the Pittsburgh Supercomputing Center came up running. Only three months after National Science Board approval, the CRAY X-MP/48 was tested and available for state-of-the-art computational research. Considerable credit for this flying start goes to Jim Kasdorf and his staff at Westinghouse, where the CRAY occupies 10,000 very-well-tended square feet. Kasdorf has been managing supercomputers since the term had meaning.

His participation, through Westinghouse, with Levine and Roskies adds immeasurably to the three-way linkage that created the PSC: Carnegie Mellon—a private university known for its excellence in science, engineering, and computer science; the University of Pittsburgh—a state-related university with a record of accomplishments in scientific and biomedical research; and Westinghouse Electric—a major U. S. corporation with extensive experience in effectively managed research computing. The three offer a synergy of institutional resources that may be unsurpassed in the world.

The Center opened, users came, research began, usage increased rapidly, and within months important scientific results emerged. The Center proudly highlighted some of this research in a booklet, **Projects in Scientific Computing**, published a year ago. "We wanted not only to justify the substantial public investment in what we do," says Roskies, "but we wanted also to show other scientists, skeptical or unaware of the value of supercomputing, its power as a research tool."

Now, another year later, usage has continued to climb. The CRAY is fully utilized. Hundreds of scientific papers have been written based on results obtained at the PSC. Some representative examples of this research appear here—significant work which will again help to show the value of supercomputing to the NSF and to the public, who ultimately pays the bill.



Bob Stock examining CRAY usage on his Micro VAX display. Vivian Benton, editor of the **PSC News**, *looks on. Ken Hackworth of the User Service's staff has something to smile about.*

The Fundamental Goal: Support for Scientific Computing

Referring to the five NSF supercomputing centers, Gordon Bell, former head of the NSF Directorate for Computer and Information Science Engineering, noted recently that they're all different, there's no single best way to operate a center: "I think they're all independently defining what they want to be." After two years, the Pittsburgh Supercomputing Center has come a long way toward defining the kind of supercomputing center it wants to be. It has a philosophy which helps people do their job. It fits with the hard-working, no-nonsense heritage of Pittsburgh; and it fits too with Pittsburgh's reputation for small-town friendliness. Most of all it helps actualize the Center's commitment to science.

"The fundamental goal of the Pittsburgh Supercomputing Center is to provide support for scientific computing. For the User Services staff, this means we make it easy for users to do their work," wrote Bob Stock in the July 1987 issue of **PSC News**. As Manager of User Services, Stock plays an important role in setting a tone and establishing the image users form of the Center.

"There are four ways," says Stock, "of serving the function of making it easy for scientists to use the supercomputer." In order of "idealized importance," they are training, documentation, trouble-shooting and consulting, and projects. "Theoretically, training and documentation should work so well that users could go through training, read their manuals, and be able to sit down and work on the supercomputer. In reality, most User Services' efforts are devoted to trouble-shooting and consulting, because users can't always get to training classes and often don't bother to read manuals. As time permits, we develop projects—the design and implementation of software which makes it easier for users to exploit the supercomputer."

The November/December 1987 issue of **Computers in Physics** noted that the Pittsburgh Supercomputing Center was "rated highly by users for its services and education programs." The Center's user education efforts include a four-day training course every few months and a two-week Summer Institute. One-day seminars and on-line tutorials are also offered. "We've acquired or written manuals and introductory guides for most of the software we have," says Stock. During the year, the Center added more than 40 applications software packages. The **PSC News** publishes updates to the software library and other valuable user information monthly. "And it's free," adds editor Vivian Benton, "just call me."

"We're here to provide a service," Stock continues. "There are four aspects of that service: responsiveness, enthusiasm, support, and software and hardware. Responsiveness means the user knows we take his or her concerns seriously. If a user's question can't be answered right away, we follow a policy of research and regular reporting back. You always have some resolution, a closing . . . even if it means, 'We can't fix this.' You never just leave it hanging."

"Enthusiasm is what I try to impress on all consultants. When we hire consultants, we want people who are enthusiastic about supporting others. We try to impress this need for enthusiasm on all our consultants, and especially the student consultants, who represent the 'front line,' the first people the user will contact."

Support extends from these front-line consultants through the Senior User consultants to the Scientific Specialists, those staff members with research experience at the postdoctoral level. "Because of the academic environment here, and the conviction that we're contributing to leading-edge science, we've been able to attract these people," says Levine. The Scientific Specialists, along with providing the Center's highest level of software development, consult with users in chemistry, fluid dynamics, general relativity, genetics, and crystallography. "The knowledge they bring enables them to go quickly to the heart of many scientific computing problems and suggest efficient solutions."

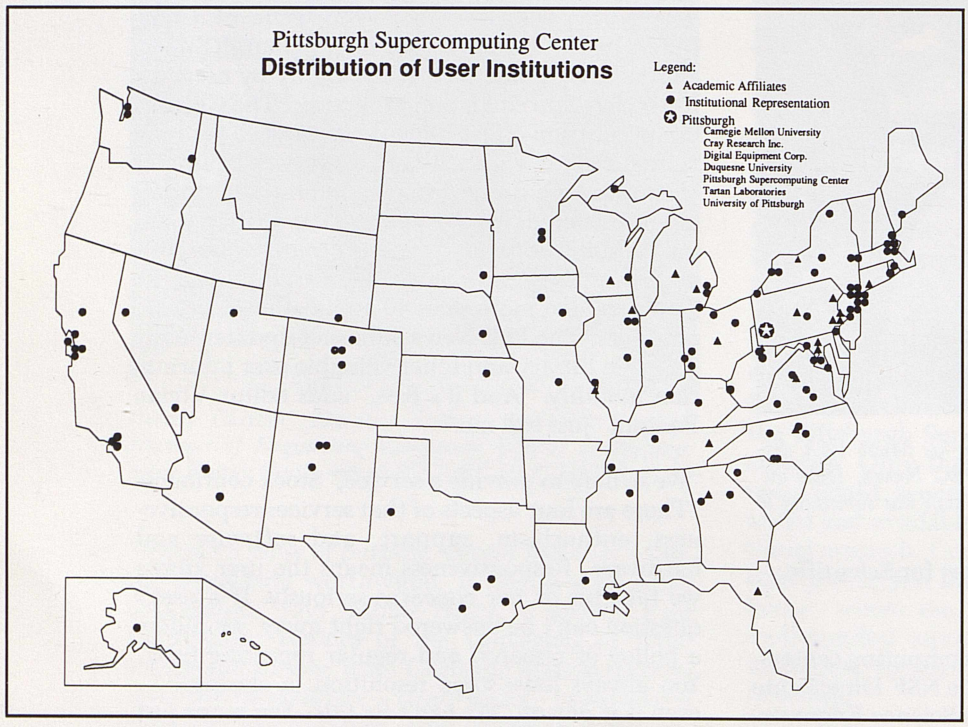
Perfection? Stock knows better. But he has no qualms about saying that the effort to serve users well has had success. "We want all our user-user service contacts to result in a more educated user. We don't want to solve just one problem, one time; we want that user to be able to solve similar problems, all the time."

Support for Research Nationwide: Remote Users

The Pittsburgh Supercomputing Center has served more than 1400 users at 152 institutions in 41 states. During the last quarter of 1987, furthermore, over 90% of the Center's usage came from outside the Pittsburgh region. "With this amount of remote usage, our emphasis has to be on educating the users to answer questions themselves," says Roskies. The Center's extensive documentation and on-line help files are a step in this direction.

The nationwide scope of PSC's outreach to users was signalled early on with formation of the Academic Affiliates. This group of 26 universities established a relationship with the Center with the aim of facilitating access to supercomputing by qualified researchers at their campuses. Representatives from the Affiliates form the Center's advisory committee, which provides useful feedback and input on all facets of operation. A Center-trained representative on each campus serves as liaison to answer user questions and promote awareness of the advantages to scientific research which supercomputing offers.

With the Center's high degree of remote usage, networking occupies a good share of its energy. During this past year, Gene Hastings, Manager of Communications and Operations, oversaw implementation of PSCnet, a regional network that connects many of the Center's Academic Affiliates. "These T1 lines offer the highest speed available today," says Hastings. PSCnet links the Center with Case Western Reserve University, Ohio State University, the University of Michigan, and the University of Maryland. "It connects them not only to us but also to what's being called the National Research Internet." Through Maryland, the Center also connects to SURAnet, a network spanning the Southeast United States. PSCnet also gives Drexel, Lehigh, and Temple convenient connections to the NSF backbone network, which links the five NSF supercomputing centers.



At the national level, Scientific Director Levine co-chairs the Federation of Research Networks. This group works to assure coordination between regional high-speed networks and the NSF backbone network. While these networks undoubtedly need improvement, already they are providing access for much of the remote usage at PSC.

Software Projects at PSC: Making the CRAY Easier to Use

"We choose our software projects carefully and look for the ones we can develop quickly and which our users need," explains Marvin Zalevsky, Systems and Software Manager. The first of these was a software interface which allows remote job submission to the Center's X-MP via BITnet, a convenient store-and-forward mail network. "BITnet was never intended to be used this way, and we're still amazed at how much work it helps to accomplish.

"We also recognized," Zalevsky continues, "that two-dimensional graphics needed a standard output format. So we coordinated every graphics package on the CRAY to produce CGM metafile format. Then Phil Andrews wrote GPLOT, an interface from CGM metafiles to all our graphics hardware devices. It's 20 times faster than what we used before. Phil's code now is used at over a dozen places."

"Most of our users are remote, and the networks can't transmit graphic data routinely. So what can we do when users want graphics, make them come here in person? Building on GPLOT, we developed an automated videotape facility. Users now can routinely make videotapes from images generated on the CRAY. And we simply mail the tapes to the users, who have been delighted with the results."

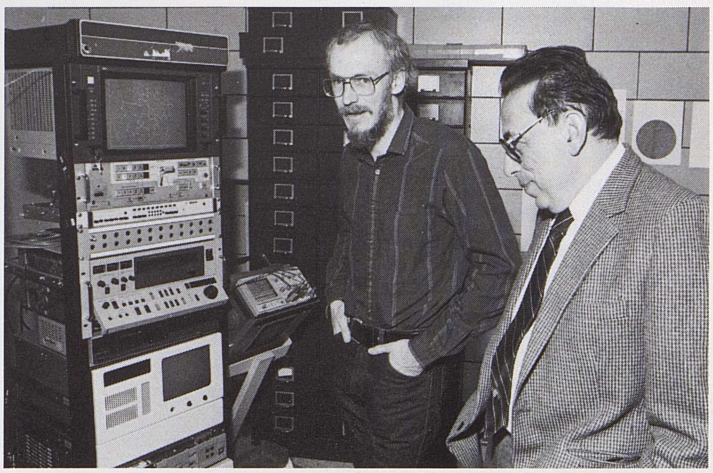
VAX SDE/Science: PSC and Code Development on the CRAY

Another opportunity to serve the remote user arose when the Center entered into a research agreement with Digital Equipment Corporation (DEC). "Our research grant involved adapting and expanding DEC's suite of VMS code development tools to support the interactive writing, compiling and running of CRAY code from within the existing VMS development environment," explains Beverly Clayton.

The Center worked with DEC's Language Sensitive Editor (LSE) and its Module Management System (MMS). LSE is an interactive program development tool that enables a programmer to create syntactically correct code while remaining within the LSE. The program uses internal templates to present choices to the programmer through menus. At any time, the user may ask LSE to test compile the code and then lead him through error correction if necessary. MMS compiles and links separate elements of a complex set of programs.

"An additional thrust of this work is to better support the remote user," says Levine, "who can do a larger fraction of the work of CRAY code development at their home site by using these new tools on their local workstation." Levine enlisted Jon Goldick, a senior in physics at Carnegie Mellon, who had worked with Levine since his freshman year.

"I expected about 30,000 lines of code at the outside," says Goldick. "I was expecting it to be one language, CRAY FORTRAN, and no library." 700,000 lines of code and a year later, Goldick and others had implemented not one, but two versions of CRAY FORTRAN (CFT and CFT77), CRAY C, four scientific libraries, and online documentation and several HELP libraries to go along with them. "The HELP libraries are massive," says Goldick. "LSE helps users search the CRAY scientific libraries for appropriate subroutines and then examine the meaning and types of variables in the appropriate calling sequences."



Erich Bloch, National Science Foundation Director, and Scientific Specialist Joel Welling appreciate the last frame of "minimal spanning trees," a video animation done with GPLOT by William Eddy of Carnegie Mellon University's Statistics Department. Bloch visited PSC on April 20.

The Center finished the extended version of LSE in November '87, and DEC formally announced it as the "VAX SDE/Science Integration Package" on May 17. "The Center's mission is to enable more pure scientific work to be done," says Clayton. "In use VAX SDE/Science allows researchers to spend more time on science and less time on application development. It gets them through the business of developing their applications faster."

Robert H. Ewald, CRAY's Software Division vice-president, added CRAY's endorsement: "We feel that this product will be beneficial to our joint customers in government, aerospace, education, and petroleum. In fact users in all our government and industrial sectors should benefit from using VAX SDE/Science both today and tomorrow."

The VAX Supercomputer Gateway, PSC and Vendor Cooperation

Through its research grant on VAX SDE/Science, the Pittsburgh Supercomputing Center helped consolidate a working relationship between two of the major computer vendors in the country. In a similar cooperative venture, the Center helped with a major hardware development project. "By serving as the site for beta testing of the VAX Supercomputer Gateway, we participated in joint DEC/CRAY development of new technology for the supercomputing community," says Clayton. The VAX Supercomputer Gateway is an enhanced communication path between CRAY supercomputers and VAX BI-based configurations. It has been developed to overcome one of the major problems in supercomputer usage: the need for high-speed data communication.

Joint efforts such as the SDE/Science and Supercomputer Gateway projects will continue at PSC. They've recently entered a research agreement with DEC to provide a living laboratory for distributed processing of applications that run partly on CRAY, partly on DEC equipment. "These efforts improve the competitiveness of both companies," Roskies notes, "and help make the supercomputing effort in this country a more vertically integrated enterprise, which is one of the strengths of the Japanese competition."

Melinda Shore, Senior Systems Programmer, makes a point about migration to UNICOS, the CRAY operating system which PSC is implementing in stages. Lori Garrett of User Services listens attentively. She supports PSC's "friendly users," currently more than 60 who run on UNICOS.

Biomedical Research at PSC, A Grant from NIH

In September 1987, the Pittsburgh Supercomputing Center received word that the National Institutes of Health (NIH) had awarded the facility a $2.2 million grant to provide the biomedical community with supercomputing resources, training, and user support. This three-year grant, awarded by the NIH Division of Research Resources' Biomedical Research Technology Program, represents a major NIH initiative in this area. "The Center is delighted to have been selected from among stiff competition to receive the award," says Roskies.

In December, Hugh Nicholas, a Scientific Specialist with an extensive background in crystallography, genetics, and molecular biology joined the staff. "One immediate need for supercomputers," says Nicholas, "is to help researchers compare newly-sequenced segments of a genome or protein with known sequences. This may be important when the biological function of the newly-sequenced genes or proteins is known—for instance, when a segment of a chromosome is identified as contributing to some disease or gross property of an organism." To make this possible, Nicholas has acquired a number of databases used widely in biomedical research.

Nicholas predicts that supercomputing will become increasingly important in simultaneous comparison of several sequences rather than conventional pair-wise comparison. The Center now has several programs for comparing macromolecular sequences, and Nicholas is planning an NIH-sponsored workshop on sequence comparison on the CRAY X-MP.

Nicholas' background in crystallography combined with the University of Pittsburgh's well-known strengths in this area will be especially valuable to biomedical researchers. Macromolecular crystallography on supercomputers will be the focus of another NIH-sponsored workshop at the Center this October. William Furey of the University of Pittsburgh's Crystallography Department and Axel Brünger of Yale's Department of Molecular Biophysics and Biochemistry will lead the workshop. Brünger's very useful work with simulated annealing as an automated crystallographic refinement technique is reported in the science section.

David Deerfield, a chemist with experience in quantum chemistry and molecular mechanics, joined the staff in March. He'll assist Charles Brooks of Carnegie Mellon's Department of Chemistry with the Center's first NIH-sponsored workshop this July on the subject of molecular mechanics. "Modeling based on molecular mechanics," notes Roskies, "should improve drug design by a factor of 10. Only 1 in 10,000 designs succeed with conventional techniques."

"Macromolecular structure, drug design, reconstructing three-dimensional structures from two-dimensional images, and the massive project to sequence the human genome—this is very important work to which the PSC can contribute," says Roskies, "thanks to the confidence which the NIH has placed in us."

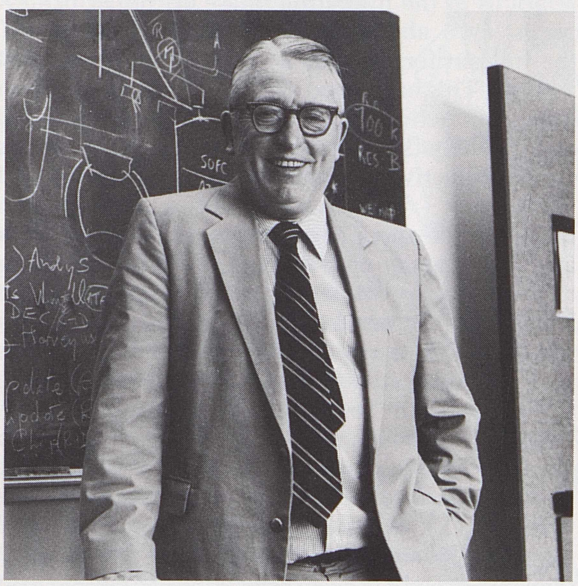
Projects in Scientific Computing 1987-88

The known is finite, the unknown infinite; intellectually we stand on an islet in the midst of an illimitable ocean of inexplicability. Our business in every generation is to reclaim a little more land.

— T. H. Huxley

Superconductivity, a Preface with John K. Hulm

The 1987 American Physical Society Annual Meeting session on superconductivity is jammed, standing room only, and runs almost until dawn. "The Superconductivity Revolution," says the cover of **TIME**. The President gets into the act with an 11-point "superconductivity initiative" intended to assure that the United States wins the international R & D race to cash in the new technology. Various agencies announce research programs almost month-to-month while industry searches quietly for talent.



Dr. John K. Hulm, Chief Scientist, Westinghouse R&D. Hulm chaired the panel which issued NSF's ***1987 Report on High-Temperature Superconductivity.*** *His January 1988 article in* ***Science****, co-authored with Theodore H. Geballe of Stanford, surveys the background, physics, and projected technology of high T_c superconductivity. "Most of the fellas seem to think it's pretty good," says Hulm, who will retire from Westinghouse in July 1988.*

Much of this work aims at the now almost foreseeable "holy grail"—as it's termed so often the metaphor begins to rust—a viable room temperature superconductor. What does it all mean? Possibly the closest thing to a free lunch that Mother Nature has to offer: energy traveling point to point with no charge for the journey, no electrical resistance, whether a mile or a hundred thousand miles. Little imagination is required to appreciate the consequences of a reduction in the real cost of electrical energy to a small fraction of what it is today.

John K. Hulm, Chief Scientist for Westinghouse Research & Development, discovered the A15 class of superconductors, which for 15 years held the field as the highest known T_c—the critical temperature below which a material begins to be superconducting. Hulm believes that the most important applications for the new superconducting materials are electronic. Josephson junction technology holds the potential to provide logic circuitry which can switch between binary states two orders of magnitude faster than the fastest semiconductor, with two orders of magnitude less energy per flip. The latter will become especially significant, says Hulm, as ever more compressed circuitry strains the limits of heat which can be dissipated.

What role will supercomputing play in superconductivity research? "The identification of the primary mechanism responsible for the high T_c is possibly the most challenging problem in condensed matter physics today," writes Hulm. The new superconducting materials are complex structures with molecular unit cells that include typically 20 or more atoms. Computational quantum studies of these materials require enormous computing power and memory. Such studies, more than any empirical method, can hypothesize and test answers to the theoretical question of the day: What causes high-temperature T_c? And such studies can propose detailed band structure data, which can be verified experimentally with diffraction and spectroscopic techniques, to yield accurate images of the physical system on a microscopic level. For such work, supercomputing is the *sine qua non*.

The discovery in 1986 which startled the dormant field of superconductivity into its now thriving activity was T_c above 30K in lanthanum copper oxide (La_2CuO_4). The next two reports of research using Pittsburgh Supercomputing Center's CRAY X-MP/48 involve studies of this compound. In the second, J. D. Joannopoulos of M.I.T. reports on a pseudopotential that achieves good results with transition metal alloys, such as the cuprate perovskite high T_cs. Ching-ping S. Wang and Philip Sterne cast a revealing light on antiferromagnetism, the phenomenon which many see as related to the superconductivity mechanism of the new materials.

Oxygen Vacancies and Antiferromagnetism in Lanthanum Copper Oxide

Philip A. Sterne & Ching-ping S. Wang, University of Maryland

Introduction: High T_c Superconductivity

Bednorz and Müller won the 1987 Nobel prize in physics for their January 1986 discovery of the first in a rapidly expanding line of copper-oxide based high-temperature superconductors. They prepared a version of lanthanum copper oxide, La_2CuO_4, with 5-10% of the lanthanum replaced by barium (barium doping) and raised the critical temperature (T_c), the transition to superconductivity, from 23 K where it had stood for nearly 15 years to above 30K. Not long after, T_c was extended to the 90-100 K range with yttrium barium copper oxide ($YBa_2Cu_3O_{7-x}$), and more recent discoveries of superconductivity in the 120-130 K range in bismuth and thallium compounds have raised hopes for room temperature superconductors, a prospect with enormous scientific and economic implications.

These copper oxides share many features that are different from previously known superconductors, and much of the recent research has focused on determining what causes their superconductivity. Conventional superconductivity is understood within the framework of the BCS (Bardeen, Cooper, Schrieffer) theory as a phase transition to a coherent state of correlated electron pairs. An attractive interaction between electrons—which overcomes the normal repulsion—occurs due to exchange of phonons (lattice vibrations) in the atomic structure of the material. Until the discovery, within the last ten years, of superconductivity in heavy fermion compounds ($CeCu_2Si_2$, UBe_{13} and UPt_3), the BCS theory was widely accepted as the only mechanism for superconductivity. In the case of heavy fermion superconductors, however, there is much experimental and theoretical evidence to suggest that the pairing mechanism is due to electronic excitations (antiferromagnetic spin fluctuations) which are highly detrimental to conventional superconductivity. Unfortunately, the transition temperatures in heavy fermion superconductors are extremely low (1K or less).

Antiferromagnetism, Doping, and Oxygen Vacancies. A unique feature of the new oxide superconductors is that supercomputing properties show extreme sensitivity to small amounts of rare earth doping or oxygen deficiencies. Superconductivity appears to occur near a metal-to-insulator phase transition which can be controlled by doping or oxygen deficiencies. In the metallic phase, the system is superconducting below T_c and normal above it. In the insulating phase, where the electrical resistance increases strongly as the temperature is lowered, antiferromagnetism has been observed. In this state, the magnetic moments (from the electron spins) on the copper ions are aligned, but instead of parallel alignment, as in ordinary everyday magnets, each spin is anti-parallel to its neighbors, leaving a net magnetization of zero. Recent neutron scattering experiments at Brookhaven National Laboratory have found that long range antiferromagnetic correlations in the insulating phase of copper oxides persist even above the Neel temperature, where the full antiferromagnetic ordering has been destroyed. This has fueled speculation that the magnetic interactions may extend into the metallic region where they become the mechanism causing superconductivity.

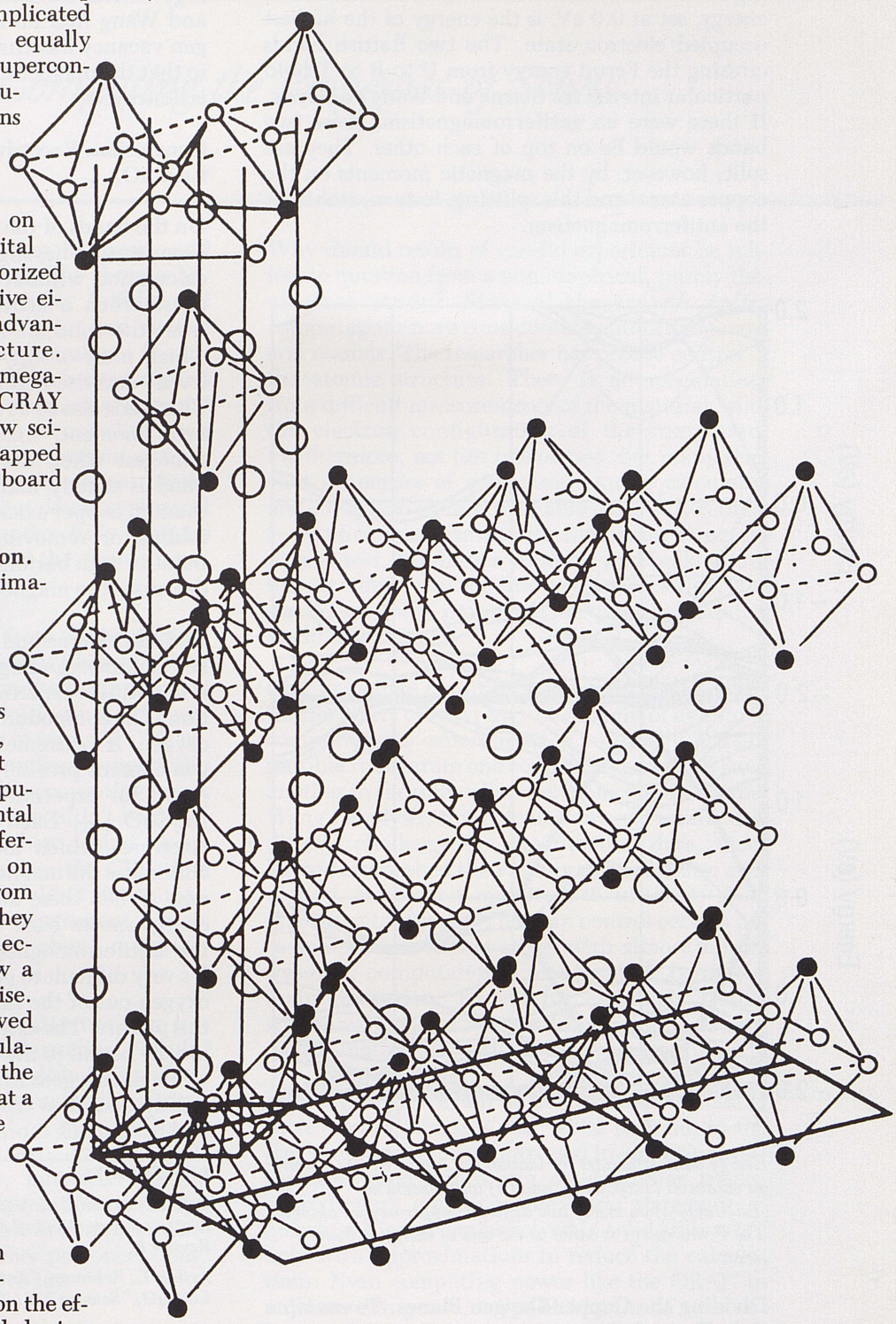
Philip Sterne and Ching-ping S. Wang, physicists at the University of Maryland, used PSC's CRAY to study the electronic and magnetic properties of lanthanum copper oxide. Without barium doping, La_2CuO_4 is a semiconductor, and recent experiments at Exxon Research have shown that an oxygen deficient form ($La_2CuO_{3.97}$) is antiferromagnetic below 220K. Antiferromagnetism has been observed in other high-temperature superconductors. For instance, yttrium barium copper oxide ($YBa_2Cu_3O_7$), which superconducts at 90K, becomes semiconducting and antiferromagnetic with only one-seventh of its oxygen atoms removed. Antiferromagnetism in lanthanum copper oxide is incredibly sensitive to the amount of oxygen in the compound; with 1% less oxygen, the system changes from no magnetic ordering to antiferromagnetism at room temperature. This suggests that oxygen vacancies in the La_2CuO_{4-x} structure may be a factor in shifting the system to antiferromagnetism. Sterne and Wang designed their calculations to explore this possibility. Interestingly, the results suggest that oxygen vacancies actually reduce the tendency to antiferromagnetic order.

A First Principles Approach to Antiferromagnetism

They performed first principles *(ab initio)* calculations of the electronic properties of lanthanum copper oxide with and without oxygen vacancies ($La_2CuO_{3.75}$ and La_2CuO_4). They also examined both crystal forms of the compound: a high temperature body-centered-tetragonal form and a

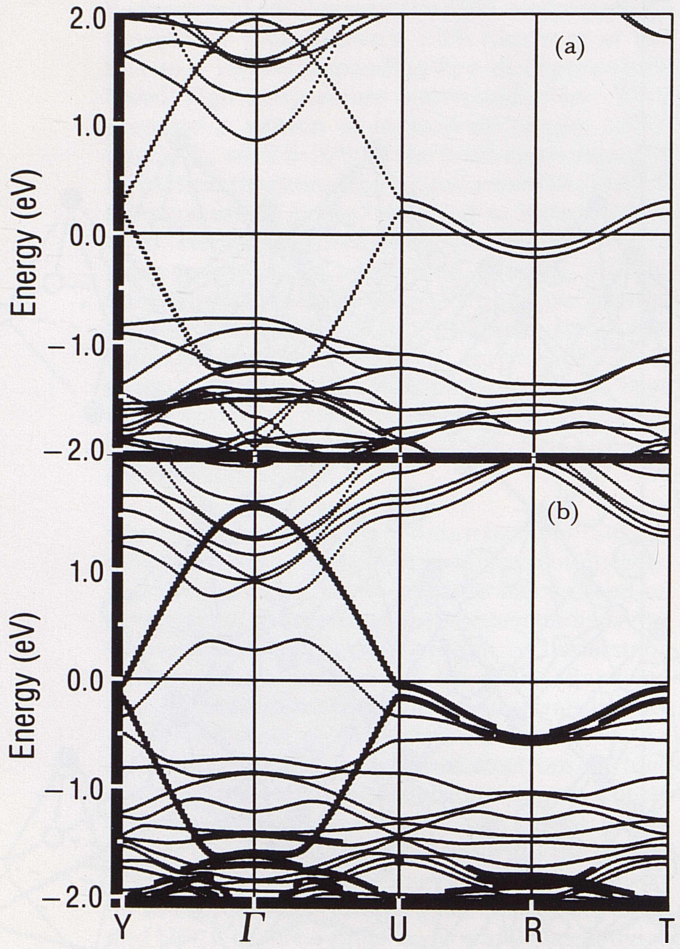
very similar orthorhombic lower temperature form. In these calculations, it was especially important to examine the complicated electron-electron interactions equally crucial to magnetism and superconductivity. To keep their computationally intensive calculations within the realm of tractability, they used the local spin-density approximation. Their program, based on the Linear Muffin-Tin Orbital (LMTO) method, was vectorized carefully to adapt the extensive eigenvalue calculations to the advantages of CRAY architecture. Nevertheless, it occupied six megawords, three-quarters of CRAY memory, an indication of how science on this level would be trapped indefinitely on the drawing board without supercomputing.

A "Constrained" Calculation. The local spin-density approximation works well for a wide range of applications, including magnetic materials, but may fail when localized magnetic moments are involved, as in this case, and Sterne and Wang were not surprised that their first computation, contrary to experimental results, showed no stable antiferromagnetic state. Building on previous work by a group from Ames Laboratory, however, they analyzed the calculated electronic structure and saw how a magnetic moment could arise. This physical insight allowed them to "constrain" the calculation by forcing occupation of the energy bands in such a way that a magnetic moment became observable. Having reached this point of concurrence with the experimental results, they were able to go forward with analysis and further calculations aimed at shedding light on the effects of oxygen vacancies and doping on the antiferromagnetism. Through the combination of a thorough understanding of their system on a physical level and sure-handed control of their extremely complex calculations, the Sterne and Wang calculations exemplify supercomputing at a high level.



Structure of orthorhombic La_2CuO_4. Large open circles represent lanthanum atoms; small open and filled circles represent oxygen atoms. The copper atoms (not shown) are centered in the oxygen octahedra. Used with permission, **Science** 237 (1987) 1136.

Figure (a) shows the energy band structure resulting from the constrained calculation. The Fermi energy, set at 0.0 eV, is the energy of the highest occupied electron state. The two flattish bands crossing the Fermi energy from U to R to T hold particular interest for Sterne and Wang's analysis. If there were no antiferromagnetism, these two bands would lie on top of each other. They are split, however, by the magnetic moments on the copper atoms and this splitting, in turn, stabilizes the antiferromagnetism.



Energy band structures for lanthanum copper oxide (a) with no vacancies (La_2CuO_4), and (b) with vacancies ($La_2CuO_{3.75}$) on one of two different copper-oxygen planes. The Fermi energy is taken as the zero of energy in both cases.

Dividing the Copper-Oxygen Planes. To examine the effect of the oxygen vacancies, Sterne and Wang did another calculation in which they set up a structure with two different copper-oxygen planes; one layer had the full oxygen configuration while the other had oxygen vacancies, yielding an overall stoichiometry of $La_2CuO_{3.75}$. Figure (b) shows the resulting energy bands. The highlighted bands show the copper-oxygen layer without vacancies. It's very similar to the bands in (a). The bands along U-R-T are split, and the copper atoms on this layer retain their antiferromagnetic order. The new, flatter band near the Fermi energy derives from the vacancy layer. Here Sterne and Wang find no antiferromagnetism; the oxygen vacancy has changed the electronic structure so that the magnetic moment on the copper atoms is destroyed.

Conclusion: Excessive Oxygen in Experimental La_2CuO_{4+x}

On the bands of the vacancy free layer in (b), the Fermi energy lies slightly higher than it does in the calculation without vacancies in (a). This is because when a neutral oxygen atom is removed from the solid, it leaves behind two electrons, which must occupy the lowest available energy levels; and the Fermi energy level shifts upward. This leads also to a reduction in the antiferromagnetic moment, since the antiferromagnetism is strongest when the antibonding copper-oxygen band is exactly half-filled, as it is for pure lanthanum copper oxide. Any deviation from this, by adding or removing oxygen or replacing lanthanum with barium, which has one less electron, decreases the magnetic moment.

And this, Sterne and Wang conclude, provides the key to understanding why the antiferromagnetism is so sensitive experimentally to oxygen concentration. The compound actually has a little too much oxygen. It's extremely difficult to control the oxygen content precisely, and the compound as prepared for experiment actually has the formula $La_2CuO_{4+.03}$. Each of the excess atoms binds two electrons, which lowers the Fermi energy and eliminates the magnetic moment on the copper atoms. When these atoms are removed, the Fermi energy moves back toward half-band filling and the antiferromagnetism grows. Experimentally, it's very difficult to get more than about 1% of the oxygen out of the sample, and this fits well with this picture. The extra oxygen atoms will be rather loosely bound to the crystal and easy to get rid of, while the oxygens in the copper-oxygen planes are tightly bound to their neighboring copper atoms and need much more energy to break their bonds.

REFERENCES:

P. A. Sterne and C. S. Wang, "Oxygen Vacancies and Antiferromagnetism in La_2CuO_4," to be published in ***Physical Review B.***

Arthur L. Robinson, "Antiferromagnetism Observed in La_2CuO_4," Science 236 (1987) 780.

Commentary

Ab Initio *and Computing the Properties of Solids*

There are important differences in theory and computational approach among the three solid-state physics studies reported here. Articles and books have been written arguing for and against one or another of the differing techniques for approximating quantum theory which these reports represent. There are, nevertheless, some essential common points which may be more telling than the differences. All three studies represent computationally intensive quantum theoretical work. Each of them derives results that describe physical properties of the compound under study using no empirical data as input, beyond the atomic number of the constituent atoms.

This approach, of beginning from the beginning, using no experimental data even if it's available is called *ab initio*, literally "from the beginning," or first principles. Why are *ab initio* studies important? They are very expensive in terms of CPU usage, much more than parameterized (empirically based) methods. Without supercomputing, most *ab initio* studies of metal alloys couldn't be done. None of the studies included here would be attempted on a VAX class machine; the CPU time would be prohibitive. Why shouldn't the equivalent resources be devoted to experiment, determining accurate information about materials from the materials themselves?

For Joannopoulos, *ab initio* investigations are important "because they involve theoretical calculations or simulations which are able to stand on their own." Since they don't rely on empirical data as input, results don't depend on the validity or limitations of experiment. As such, *ab initio* studies complement experiment, but need not be guided by experimental interpretation. *Ab initio* studies are free to yield a result in complete defiance of experiment or researcher preconceptions. They will say what they have to say.

One need go no further than Wang and Sterne's report on La_2CuO_4 to appreciate Joannopoulos' point. Neutron diffraction experiments, a costly, difficult experimental process, revealed antiferromagnetism when a small quantity of oxygen was removed from what was believed to be La_2CuO_4 rather than, as Sterne-Wang's *ab initio* study shows, $La_2CuO_{4+\in}$. Which is correct? Time will tell; what's important, at least provisionally, is that the question has been raised.

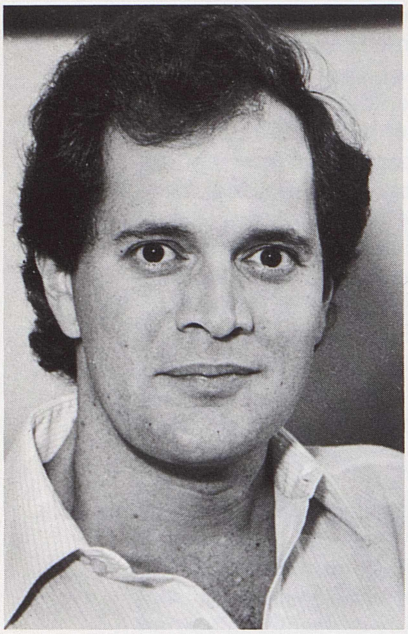
Why should results of careful experiment be subject to question from a non-empirical, purely theoretical study? State-of-the-art *ab initio* computations have considerable authority for several reasons. The researcher has precise control of the atomic structure. There is no speculation, from difficult measurements of the material, as to the electron configuration of the compound. Furthermore, not just one or two, but many relevant properties of a material can be calculated and compared against available data. When there is significant agreement for important structure parameters, like lattice constant and bulk modulus, as there was in Sterne-Wang's study, the prediction for a parameter in question acquires credibility.

First principles quantum studies, moreover, allow the flexibility of what may be thought of as unique computational experiments. Empirically, it's impossible to separate one copper-oxygen band from another in the unit cell of the molecule. In Sterne-Wang's analysis, this calculation gave results that directly challenge the experimental data. Once having established the validity of an *ab initio* simulation, researchers may treat it, in effect, as an experimental device. They can control certain parameters while testing others with almost infinite freedom, computational requirements being the main limitation. Except for extremely simple structures, none of this would be possible in condensed matter work without supercomputing.

Looking briefly again at the differences among these studies, one sees that when it comes to the particular set of techniques used in *ab initio* investigation, good minds run on dissimilar tracks. Schrödinger's equation, the basic statement of quantum theory, can be solved in solid-state work only with approximations to reduce the calculations. Even computing power like the CRAY, in any interesting case, is insufficient for an exact solution. For theoretical work on the cutting edge, the issues often involve approximation methods. Joannopoulos' preferred method, for instance, uses three techniques he has combined and applied to a variety of solids with good results. It's not the same method as Freeman's, and Freeman puts his forward with no less authority. Both would argue that their approach is state-of-the-art for the alloys they study.

Defects in Molecular Solids and a Pseudopotential for Superconductors

J. D. Joannopoulos, Massachusetts Institute of Technology



Introduction: Approximating Quantum Theory for Metals

The general emphasis of J.D. Joannopoulos' research has been to develop realistic microscopic descriptions of material systems using *ab initio* quantum mechanical methods. A first principles quantum study must find a way to represent the positions of the nuclei, the interactions among the electrons, the interactions between electrons and the nuclei, and it must solve the equations. The obstacles that arise in quantum calculations with solids reflect the density of the atomic structure, about 10^{23} atoms/cm^3, and its extent, which in contrast with an isolated molecule, is essentially infinite. Each atom and its associated electrons interact with an enormous number of neighboring particles.

Joannopoulos believes *ab initio* studies in solids at the present can do this most efficiently using (1) density functional theory to model the electron-electron interactions, (2) pseudopotential theory to model the electron-nucleus interactions, and (3) supercell configurations to model systems with aperiodic geometries. The Born-Oppenheimer approximation allows the atoms of the solid to be "frozen" in a pattern; and when the pattern is periodic, or modeled as such through a supercell approximation, Bloch's theorem reduces the difficulty of treating 10^{23} atoms to a representative unit cell containing only several atoms. Bloch's theorem provides basically that electron wave functions moving within a periodic potential, the unit cell in this case, can be represented as a plane wave modulated by the periodicity of the potential.

The supercell approximation applies in the case of aperiodic configurations of atoms, such as randomly disordered alloys. The researcher constructs a large unit cell containing the configuration in question and repeats it periodically. By studying the properties of the system for larger and larger unit cells, it's possible to gauge the importance of the induced periodicity and systematically filter it out. Joannopoulos has tested this approach against exact Green's function methods, an extremely complex calculation which is tractable only for very high symmetry configurations.

In principle, explains Joannopoulos, density functional theory allows exact mapping of a strongly interacting electron gas in the presence of nuclei onto "a single particle moving in an effective non-local potential." Local approximations to this potential seem to work quite well. This "mean-field" approximation treats each electron as if it moves within an electrostatic field determined as an average of other strongly interacting electrons in the unit cell. There are no conclusive *a priori* reasons why these approximations work, but theorists have shown that density functional theory reproduces a variety of ground state properties within a few percent of experiment.

Pseudopotential theory is a relatively common approach which Joannopoulos has refined and extended to broader use in *ab initio* computations. This approach replaces the strong electron-nucleus potential with a much weaker potential—a pseudopotential—that describes the necessary salient features of a valence electron moving through the material under study. The original solid is, in effect, replaced by a system of pseudo-valence electrons and ionic cores. The weakness of the pseudopotential is crucial in enabling a solution to Schrödinger's equation for low symmetry systems. An important limitation of pseudopotentials, however, is that for transition metal atoms even the pseudopotential is too strong and makes the calculation intractable for low symmetry systems. (The transition metals occupy the middle part of the periodic table and include many important elements, such as copper, nickel and iron.) Joannopoulos has succeeded recently in overcoming this obstacle, however, and the last part of this article reports preliminarily on those results.

The Role of Supercomputing. Joannopoulos' work with pseudopotential aims at allowing tractable, reliable solutions where they would otherwise beprohibitive. In two projects undertaken at the Pittsburgh Supercomputing Center, he applies his methods to two different complex structures,

arsenide-triselenite, As_2Se_3, and lanthanum copper-oxide, La_2CuO_4. Both studies were computationally intensive. To develop and test his transition metal pseudopotential, for instance, Joannopoulos performed quantum calculations that ran 50 to 100 hours of X-MP/48 CPU time per bulk crystalline copper geometry, and several different geometries had to be calculated. Both studies yield interesting, significant results which, in both cases, would have been impossible without supercomputing. The two reports that follow, however, don't allude to supercomputing because, as Joannopoulos says, "it is assumed supercomputing is essential."

Defects in Molecular Solids: Crosslinking and Paired Electrons in As_2Se_3

The chalcogenide materials (compounds of S, Se, Te & Po) have held special interest in solid-state physics for many years. These materials are characterized structurally by two-dimensional covalently bonded networks which form in weakly coupled layers, as represented in Figure 1 for arsenide-triselenite. They are excellent glass formers and exhibit a number of unusual electronic properties thought to be caused by structural defects. Experiments have determined, for instance, that it's impossible to change the Fermi level – the energy of the highest electron-occupied state – of the materials by doping. Studies have explained the unusual properties by postulating that electrons near the Fermi energy attract each other and form pairs, a very unusual phenomenon termed "negative effective electron correlation energy."

The chalcogenide materials are extremely complicated, even in their crystalline form. As a result, experimental data have not been well interpreted, and theoretical studies have been limited by insufficient computing capability to perform accurate, realistic quantum calculations. The complex crystal of As_2Se_3, for instance, has 20 atoms in its unit cell (see Figure 1). Within each bonded layer, the As atoms bond to three Se atoms and each Se atom bonds to two As atoms.

Results. In this study of As_2Se_3, Joannopoulos has used the computing power of PSC's CRAY X-MP to perform *ab initio* calculations which he reports as the first reliable investigation of defects in a chalcogenide compound. He found a peculiar structural shift, a defect which begins as an antistructure defect within a layer, then decays to a crosslinking defect between layers. Figure 1 depicts the formation of this defect.

Even more fascinating is that the system can break the covalent bond within the layer and reform another across layers without passing through an intermediate energy barrier. The formation energy of the crosslinking defect is only about one electron volt, which Joannopoulos finds to be remarkably small in relation to the formation of

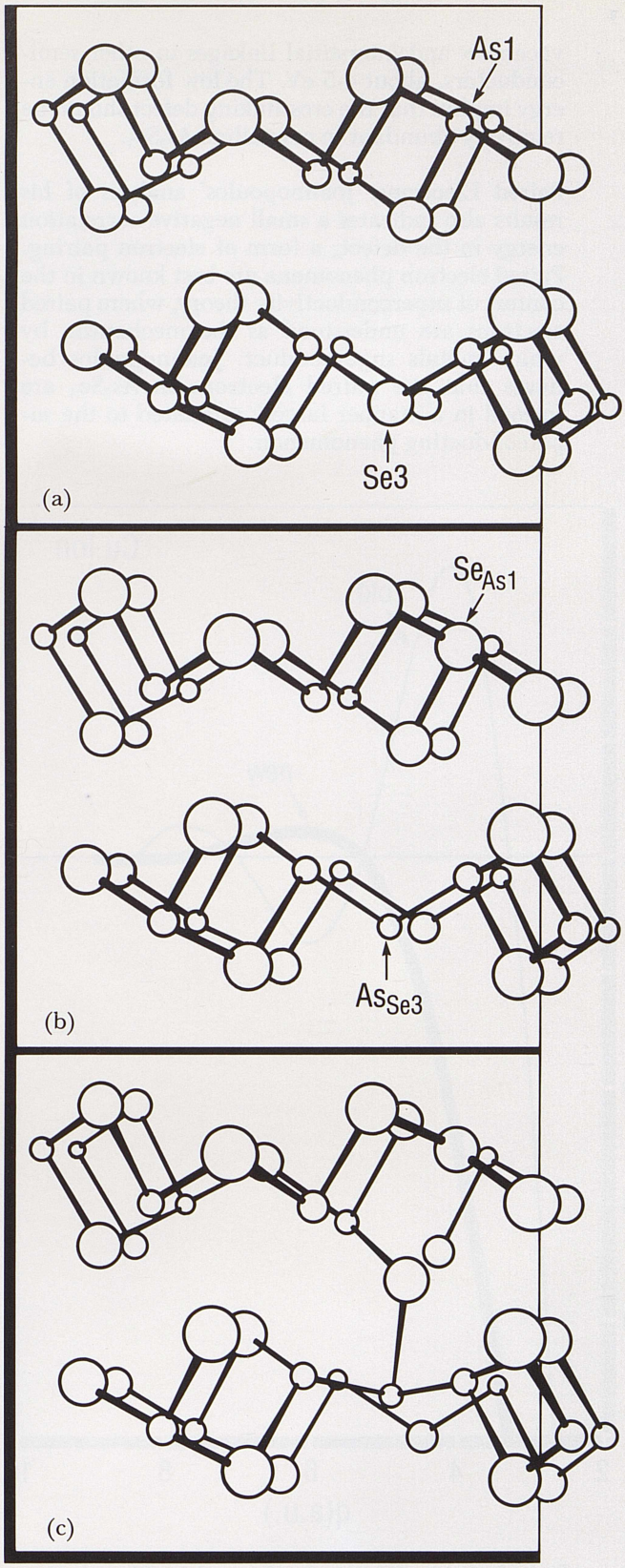


Figure 1: Crystalline As_2Se_3. Panel (a) shows a cross-section through two layers of the crystal. The crosslinking defect proceeds in two steps. First, the interchange of an As and Se atom situated close to one another on adjacent layers creates an anti-structure defect, panel (b). Second, this antistructure defect lowers its energy by a bondswitching reaction that destroys a bond within a layer and creates a bond between the two adjacent layers, a crosslinking defect, panel (c).

vacancies and interstitial linkages in other semiconductors, about 4-5 eV. The low formation energy implies that the crosslinking defect should be relatively abundant in crystalline As_2Se_3.

Paired Electrons. Joannopoulos' analysis of his results also indicates a small negative correlation energy in the defect, a form of electron pairing. Paired electron phenomena are best known in the context of superconductivity theory, where paired electrons are understood as the mechanism by which metals superconduct. Joannopoulos believes that the paired electrons in As_2Se_3 are formed in a manner largely unrelated to the superconducting phenomenon.

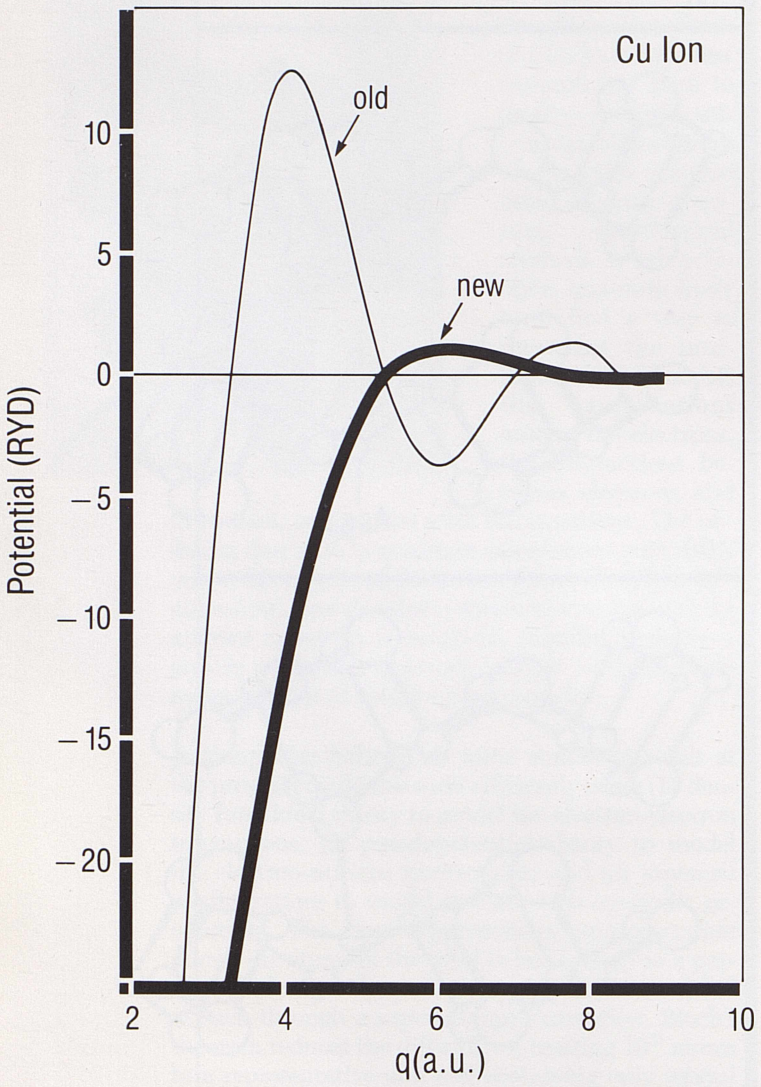


Figure 2: Copper (Cu) ion pseudopotentials. The light line corresponds to the conventional pseudopotential, and the heavy line to the newly determined class of d-electron pseudopotentials. The x axis is the momentum or wavevector of the plane waves. The y axis is the potential energy in Rydbergs (13.6 eV).

"In superconductivity," he explains, "you need many free particles which can exchange phonons, vibrational excitations. In the exchange of these phonons, the electrons bind and make pairs. In As_2Se_3, two electrons like to be in the same site. They're not mobile; they like to be stuck in a particular site, and the energy that drives this is the relaxation energy associated with the defect. In other words, it's coupling between electrons mediated through defect relaxation.

"Suppose, for example, you have two oxygen atoms and you ask: Is it energetically favorable to take an electron from one and put it on the other? . . . What happens is you have O+ and O−, infinitely separated from each other. Is that energetically favorable? . . . The answer is no." But on the lattice, as in this case for As_2Se_3, Joannopoulos' results suggest the answer may be different. "That question is related to this effective correlation energy. If that effective correlation energy is truly negative, then it's possible for an electron to transfer from one atom to the next, or from one defect to another—which is what happens in our case. And that's what's so unusual about it."

A Tractable Transition Metal Pseudopotential and High T_c Superconductors

Background to the Problem. As the first part of this report indicates, pseudopotentials are useful in solving complex, low symmetry configurations prone to vacancies and bond shifting. For this reason, they would seem to offer an ideal approach to *ab initio* investigations of copper-oxide high T_c materials. Like As_2Se_3, the copper-oxide materials have extensive unit cell structures, and the substitutional disorder generated by doping and oxygen vacancies could be systematically addressed with such calculations. The problem, mentioned earlier, is that existing pseudopotentials don't work with transition metal atoms. Their unfilled outer layer of "valence" electrons are in the "d" orbital, and unlike valence electrons in non-transition metal solids, these *d* electrons are highly "localized," close to the nucleus, which makes the potential strong.

"The spirit of pseudopotential," says Joannopoulos, "is to make a weak potential. If all the electrons are deep, near the nucleus, the wave function will have many oscillations. That represents lots of energy and you will need an enormous number of plane waves." The wave function, Ψ, the quantum theory representation of particle motion, can be approximated as a sum of plane waves, which then are used to solve Schrödinger's equation. Plane waves, which basically are ordinary sine waves, in principle can approximate any function; but the number required may make solution impossible. Joannopoulos attempted the

calculations with La_2CuO_4 and found no indication of convergence. The copper ion (Cu) pseudopotential has very large components with large wavevectors (short wavelengths), which requires a correspondingly large plane-wave basis in order to sample the potential effectively (see Figure 2).

The New Pseudopotential. "If you use the standard technique for generating pseudopotential, you get one that is very, very strong. You have to work at trying to do different types of things. You just have to look back at the problem again." Joannopoulos undertook an extensive investigation of possible pseudopotentials and tested them with calculations for bulk crystalline Cu. He succeeded in developing a class of new pseudopotentials which requires far fewer plane waves than before, yet mimics the real *d*-wave function much closer to the nucleus. More importantly, the new pseudopotential is weaker and falls off more rapidly for large wavevectors, as indicated in Figure 2. "We established this class *purely*" to make the pseudopotential as weak as possible. That way "you don't need many plane waves to expand it."

Using this new pseudopotential, Joannopoulos has been able to reproduce the bulk crystalline band structure of face-centered-cubic Cu and predict the experimental lattice constant correctly to within two percent. This work is not published, Joannopoulos explains, "because we're still working on it. The new class of potentials requires far fewer plane waves, but still we ought to be able to do better. We may not. What we've got here may be as good as we can do. But I'm not satisfied."

As a result of his success in finding the new class of pseudopotentials reported here, Joannopoulos believes that transition metal calculations are now feasible with pseudopotentials and plane waves. He's attempting to use this methodology to shed light on the structure of high T_c superconductors.

REFERENCE:

J. D. Joannopoulos, "Predicting the Structure of Solids," ***Physics of Disordered Materials*** *(N.Y.: Plenum, 1986), pp.19-35.*

The Superconductor of Pittsburgh

*Maazel. Of the West Berlin Opera. Of the Cleveland Orchestra, chosen to succeed the incomparable Szell. Of the Vienna State Opera. The National Orchestra of France. Of LaScala, the zenith, where he is acknowledged master and will conduct opera no other place. Of films—**Don Giovanni** and **Carmen**—that set new standards for opera on screen.*

Those who know only the "Smoky City" may wonder what attraction Pittsburgh holds for Lorin Maazel. But the kind of energy he draws from the Pittsburgh Symphony was harnessed long ago to generate an urban renewal that cleansed the air, scrubbed the grime and revitalized the not-so-golden triangle. Today a visitor to Rand-McNally's 1985 "most livable city" will remember a striking skyline, shining bridges and three rivers of a city transformed.

Still rebounding from steel-mill closings, which affected many in the area, Pittsburgh now conducts itself toward a future based in research, technology, commerce and culture. Currently six economic and business development groups join with over 15 venture capital funds to provide a healthy climate for entrepreneurs, which has helped to foster over 600 advanced technology firms in the region. Industrial automation, advanced materials, software engineering, and biomedical research are well represented.

*Led by active coalitions of academia, the private sector, and government, the former Smoky City has taken enormous steps in a short time. "An internationally renowned center for health care, a national computer programming industry and a leader in the nascent robotics field," wrote **The New York Times**, "Pittsburgh is increasingly touted as a yardstick for cities staggered by the decline of the country's manufacturing industries."*

From smokestack to high-tech, Pittsburgh is one of many American urban communities finding the new tools it needs to compete economically. No longer is mass production our golden goose. We will construct an industrial evolution and our economic future from one factor: our capacities, our brainpower, our creative energies. Supercomputing is one of the essential tools.

The arts too are an essential ingredient. Pittsburgh needs to attract and hold that most indispensable resource: human capital, as the economists say, qualified, talented people. In sum, you need supercomputing centers like Pittsburgh's. And superconductors like Maazel.

*"I say to people, if Pittsburgh were in Europe, people would travel one thousand miles to see it. There isn't a city in the world that can match the blazing splendor of Pittsburgh as seen when first you emerge from the tunnel, coming in from the airport." Brendan Gill, **The New Yorker.***

AC Conductivity in Substitutionally Disordered Alloys

Arthur J. Freeman & Antonios Gonis, Northwestern University

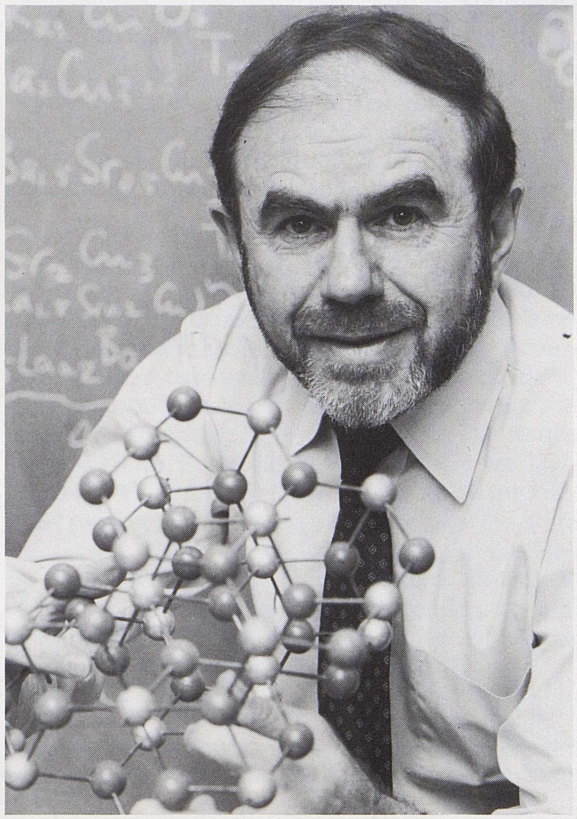
A Three Phase Study: Transport Properties in Binary Alloys

"It would not be an overstatement to say," says Arthur Freeman, "that one of the loftiest goals of theoretical metallurgy and of condensed matter physics is to achieve a microscopic predictive understanding of materials phase formation and stability." In an extensive series of studies of various alloys, Freeman and Antonios Gonis are using the computing power of PSC's CRAY X-MP/48 to advance understanding of transport properties in disordered materials. The alloys investigated include palladium rhodium (PdRh), palladium vanadium (PdV), copper palladium (CuPd), copper zinc (CuZn) and copper nickel (CuNi). The work divides into three phases:

1. Calculating electronic densities of states and AC conductivities for computer simulated disordered alloys.
2. Developing analytic theories for calculating important physical quantities of the systems simulated in the first phase.
3. Developing first principles calculations of AC conductivities for real materials.

Transport properties, such as electrical, thermal or optical conductivity, because they involve the motion of electrons under the influence of an external field, have received relatively slight attention in theoretical studies. This is due to the greatly increased complexity of calculations required for these properties, called two-particle-related properties since they involve the interaction of two particles within the same system configuration. Most studies have focused on single-particle properties, which allow computation of the static configuration and energies associated with it. One of the most important single-particle properties, because it bears a relation to AC conductivity and other two-particle properties, is the electronic density of states (DOS), which gives the number of electronic energy states within a particular energy range.

State-of-the-Art Approximations. Freeman and Gonis's work at PSC is premised on recent developments in the approximations used to calculate the electronic structure of disordered systems. They apply what they regard as state-of-the-art usages of the coherent potential approximation (CPA), the corresponding embedded cluster method (ECM), and the full potential linearized



Arthur J. Freeman

augmented plane wave (FLAPW) method. These provide workable analytic theories and, along with advanced computing capacity such as that provided at PSC, bring the goal of completely *ab initio* quantum mechanical descriptions of electronic structure, including AC conductivity, within reach. Freeman and Gonis seek to establish that the AC conductivity of substitutionally disordered alloys can be calculated as efficiently and accurately as single-particle properties within a well understood and established analytic method.

The Theoretical and Computational Approach

The starting point in any electronic structure calculation is the Hamiltonian operator, a mathematical expression named for Sir William Rowan Hamilton (1805-1865). The Hamiltonian for the system under consideration gives its energy in terms of the momenta and positions of the particles. The Hamiltonians corresponding to real systems like disordered Cu_xNi_{1-x} alloys are extremely

complicated and preclude exact solution of the corresponding Shrödinger's equation. Approximations become necessary, therefore, either in constructing the Hamiltonian or solving Shrödinger's equation.

During the first phase of their study, in order to investigate how much their CPA and ECM approximated solutions differ from the real ones, Freeman and Gonis use a reference system with known exact solutions. For this purpose, they use a more easily solved form of the Hamiltonian, the tight-binding Hamiltonian, and apply it to a disordered alloy, A_xB_{1-x}, where A and B represent two elements and the alloy is configured as a linear chain. An algorithm for generating random results, known as a Markov chain, generates a sequence of A and B "atoms" and distributes them over the possible sites of the disordered alloy linear chain structure.

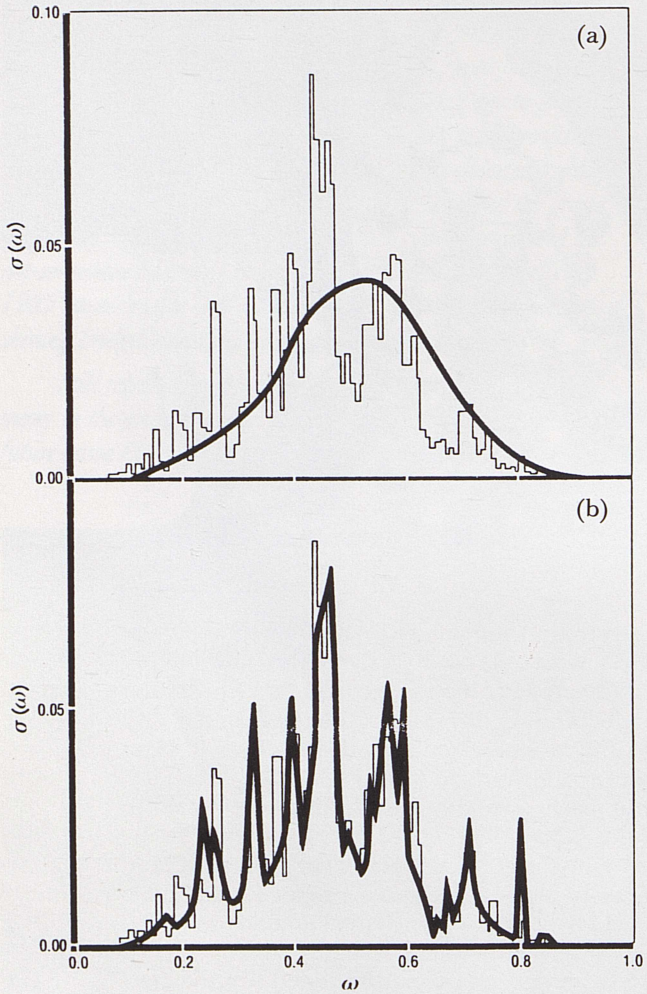


Figure 1. The histograms in (a) & (b) represent the exact numerically simulated AC conductivity, σ(ω), for a disordered linear A_xB_{1-x} alloy system. The x-axis represents frequency, ω. The superimposed smooth curve in (a) shows AC conductivity calculated by CPA, and in (b) shows it calculated by ECM with a cluster of 11 atoms.

Freeman and Gonis obtain exact solutions to the tight-binding Hamiltonians by applying negative-eigenvalue counting theorem, a commonly used technique for obtaining the exact band-width distribution of disordered systems describable by tight-binding Hamiltonians. For every possible combination of A and B, Freeman and Gonis calculate a density of states and then, in effect, average the separate calculations to arrive at a density of states for the overall system of the particular A_xB_{1-x} compound under investigation.

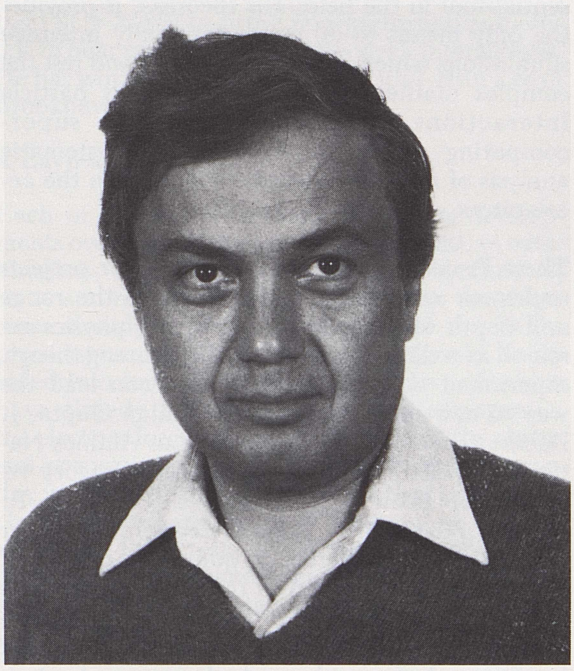
CPA and ECM. After obtaining exact solutions by this method, Freeman and Gonis compare the exact densities of states and AC conductivities of their model system with the results from calculations using the CPA and ECM approximations. They describe the CPA method as the most satisfactory single-site theory for studying the properties of random substitutionally disordered alloys, in particular the one-particle properties such as density of states. It is a single-site mean-field theory. "In the CPA, one considers that the real disordered material is replaced by a self-consistently determined effective medium, which is characterized by an energy-dependent site-diagonal self-energy and which preserves all symmetries of the lattice. The self-consistency condition, which determines self-energy, is that the scattering off of a real alloy atom embedded in the effective medium vanishes on the average. The CPA has many desirable mathematical properties and yields physically meaningful quantities."

In spite of its many desirable properties, the CPA, as a single-site theory, cannot account for the effects of statistical fluctuations in the environment of a site. Consequently, its accuracy decreases with increasing short-range order—the tendency to form ordered phases—or clustering—the tendency to segregation. A further disadvantage of the CPA, particularly in connection to one-dimensional systems, is its failure to yield a vanishing DC conductivity, as is required by formal considerations and verified by computation simulations.

The result of the CPA calculation for AC conductivity applied to the model system is represented in Figure 1(a), where it is superimposed over the exact numerical results. The CPA yields a smooth curve in general agreement with the overall shape of the exact results, but it can't reproduce any of the structure in the conductivity spectra. This structure is caused by fluctuations of the local environment, i.e., in the vicinity of individual atoms the properties of the system may differ considerably from the average properties. The failure of the CPA to reproduce this structure is a manifestation of its single-site nature.

To compensate for the limitations of the CPA method, it's necessary to extend it to a multisite or cluster theory. Freeman and Gonis believe that the embedded-cluster method (ECM) is the most

satisfactory in terms of analytical and computational requirements. The ECM, as a multi-site theory, allows the calculation of vertex corrections to any desired extent, confined to a given cluster, and limited only by computational considerations. As shown in Figure 1(b), the AC conductivity of the one-dimensional model system, calculated within the ECM method, reproduces practically the entire sequence of the exact results. It also properly yields a DC conductivity which tends to zero with increasing cluster size.



Antonios Gonis

The Need for Supercomputing. Analytic calculations of AC conductivities are computationally intensive because all possible configurations need to be considered for a given system. The number of configurations increases very rapidly with the size of the embedded-cluster used. Freeman and Gonis have fully developed and vectorized their code, giving a reduction of more than an order of magnitude over the non-vectorized versions.

These codes require the inversion and integration of large, complex matrices and, Freeman believes, can be run efficiently only on a supercomputer. For example, to obtain effective cluster interactions for all configurations of a nearest-neighbor cluster on an f.c.c. lattice, it's necessary to invert a 117×117 matrix 288 times at each energy of interest less than the Fermi energy, $E < E_F$. The diagonalization of a large matrix is one of the main bottlenecks in scientific computing, and it is only with supercomputers like the PSC's CRAY that it becomes possible. The calculation, furthermore, explains Freeman, "requires knowledge of the off-diagonal elements of the on-the-energy shell-scattering matrices, which in turn must be obtained through delicate integrals of complex matrices in k-space (the space in which the momentum or wave-vector of electron energy may be represented)."

Future Stages of the Work. After the experience gained with the simple one-dimensional model system, Freeman and Gonis are using the ECM method in the second phase of the project to calculate AC conductivities of real systems, such as Cu_xNi_{1-x} alloys and the other alloy systems already mentioned. The results of Freeman and Gonis's AC conductivity calculations will be compared with experiment, an investigation which should yield insight into transport properties in disordered matter. For the CuNi system, considerable empirical data is available, which is one of the main reasons for using this system as part of the study.

Because the configurations of the model system incorporate the set of peaks and valleys that they do, representing a more exotic system structure than encountered in experimental data, Freeman's research assistant, Dr. Miaogy Hwang, who is comparing these calculations against research data, expresses confidence that the analytical computations for AC conductivity will yield reliable results for real alloys. Freeman believes that this work ultimately will help to realize the goal of giving reliable predictions for the properties of solid-state materials: "Our generally successful calculations of alloy properties to this day," says Freeman, "make us confident that our future endeavors will provide both insight into the structure of matter and guidance for future studies."

REFERENCES:

*M. Hwang, A. Gonis & A.J. Freeman, "Transport properties of substitutionally disordered alloys," **Phys. Rev. B.**, 33 (1986) 2872-75.*

*A. Gonis & A.J. Freeman, "A Self-Consistent Treatment of Short Range Order in Substitutionally Disordered Alloys," **Phys. Rev. B.** 31 (1985) 2506.*

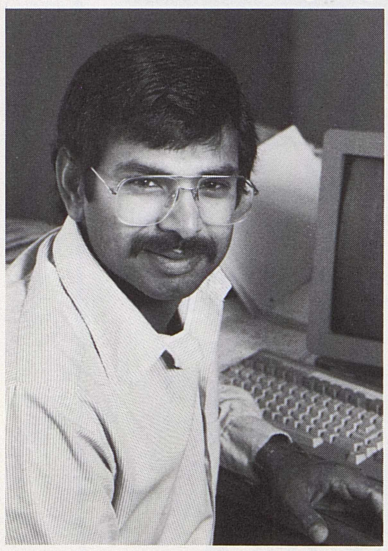
2 HIGH-ENERGY PHYSICS

Supercomputing in the Particle World

Quarks, Gluons, Bosons, Leptons and Other Creatures

Add leprechauns to the above list and it would have about as much reality for the average person. But for high-energy physicists, particles as invisible to mortals as the leprechaun are the basis of their discipline. Many of these elementary particles have no proven existence beyond the equations that describe them; yet for the physicists who study them, they're no less real. This faith in mathematics as a language that expresses real events represents nothing so much as hard-minded reality. How could it be otherwise for particle physicists in the twentieth century, when again and again the predictions of theory have led the way to experimental results?

Background. The goal of high-energy physics, or as it's also called elementary particle physics, is to discover and understand the irreducibly fundamental elements of nature. Recent decades have brought a remarkable surge of progress in this field. It began notably in the mid-1960s with the proposition, arrived at independently by physicists George Zweig and Murray Gell-Mann, that non-physically verifiable entities called, unforgettably, quarks—a name which Gell-Mann took from an obscure line in Joyce's Finnegan's Wake, "Three Quarks for Muster Mark" were constituent parts of sub-atomic particles thought at the time to be elementary. What was then a strange idea has become standard thinking, even though quarks have yet to reveal themselves openly in an experiment.



Rajan Gupta

Inspired theoretical insights, combined with particle collision experiments on the world's largest accelerators, revamped radically the pre-1960s view of the small particle world. By the early 1980s, progress in the field had arrived at a resting place, partly because testing of the new theories required computing capabilities more powerful than those in existence. Probably no other branch of modern science defines itself and its research agenda around problems as inherently intractable. As a result, supercomputing rapidly has become an essential tool in the field. For theorists, it provides the only means to do computationally intensive simulations which, in effect, play out and test the complex mathematics describing small particle interactions. For experimentalists, supercomputing makes possible detailed, systematic analysis of massive amounts of data from the accelerators.

Three Projects at the PSC. Three major projects underway currently at PSC illustrate the range and depth of contemporary particle physics research as well as the vital interplay among theory, experiment and computation that can lead the way to new understanding. Dr. Rajan Gupta, J. Robert Oppenheimer Fellow at Los Alamos National Laboratory, and his colleagues Gregory W. Kilcup of Brown University and Stephen R. Sharpe of the Stanford Linear Accelerator Center (SLAC) are investigating quarks and related gluons—particles which in theory, as their name suggests—exert a force which binds quarks together. This study is designed to examine what happens when quarks and gluons are heated, looking in particular for a phase transition, a structural shift similar to what happens when ice melts at 32F and becomes water. This phase transition, which has important implications for high-energy physics, has been predicted but never observed experimentally.

Julia A. Thompson of the University of Pittsburgh is co-principal investigator with her colleague, W. E. Cleland, in a project using PSC's CRAY X-MP/48 for analysis and interpretation of data from a large-scale high-energy physics experiment. The experiment, titled HELIOS, is a major collaborative effort, involving sixteen institutions from western Europe, North America, the Soviet Union and Israel and more
than one hundred physicists. The project uses the Super Proton Synchotron (SPS) at the European Center for Nuclear Research (CERN) in Geneva, Switzerland, one of the major particle physics research centers in the world. In one part of HELIOS, the participants hope to observe and record experimentally the phase transition which Gupta and his colleagues are examining through theory and computation.

Robert E. Shrock of SUNY at Stony Brook's Institute for Theoretical Physics has undertaken a

noteworthy series of computational studies in lattice gauge-Higgs theory. The basis for this theory developed over a period of years in the late 1960s and 1970s. Three physicists, Sheldon Glashow, Abdus Salam and Steven Weinberg, won a Nobel Prize in 1979 for their contributions to this work, now well accepted, which unifies the so-called weak and electromagnetic forces, two of four basic forces identified by twentieth-century physics. Calculations of electromagnetic and weak-force interactions generally use what's known as perturbation theory, a particular kind of approximation, to arrive at solutions. Shrock's work has employed the PSC's X-MP to bring non-perturbative methods to bear on electromagnetic-weak interactions. He has worked in collaboration with I-Hsiu Lee of Brookhaven National Laboratory.

A Layman's Guide to QCD

Each of these studies relates to major developments over the last twenty-five years; and, in combination, they illustrate how science postulates, tests, and retests, building patiently on the structures already in place. By the end of the 1930s, physicists accepted that interactions could be divided into four categories of force, two of which have already been mentioned: electromagnetic, the electron forces which largely govern the chemical properties of matter and much of our everyday world; strong, the interaction that holds a nucleus together, and the most powerful of all four; weak, the force responsible for some radioactive decays and such phenomena as parity violation (a particle reaction that lacks a symmetrical "mirror image" effect); and, the fourth force, gravity.

A deep understanding of the electromagnetic force emerged in the first few years after World War II. The theory that provided accurate answers, quantum electrodynamics (QED), derived from quantum field theory. For the next two decades, physicists worked with quantum field formulations in efforts aimed at explaining the other three interactions. The search proved unsuccessful, and by the late 1960s most physicists believed quantum field theory was an inappropriate model for the weak, strong, and gravitational forces. Since then, however, several major breakthroughs have led to a quantum field understanding of the weak interactions (now known as the standard model), while another breakthrough theory, quantum chromodynamics or QCD, offers a viable understanding of the strong interactions.

The parallel nature of QED and QCD raises the unified field notion of bringing all four forces within the domain of one grand theoretical explanation—a notion that occupied most of Einstein's energy after the successes he achieved with special and general relativity, and one that vibrates somewhere in the mind of most high-energy physicists. If QED and QCD could be linked with each other, the strong force would be brought into the fold with the electromagnetic and weak force unification. The underlying equations have considerable similarity, and unification schemes have been suggested. The equations of QED find solution only through a physical approximation known as perturbation theory. The physics can be understood as a series of smaller and smaller effects dependent on higher and higher powers of a small parameter (the fine-structure or "coupling constant," about 1/137 for the electromagnetic interaction). QCD has no intrinsic small parameters. The coupling constant is 1.0 or greater, and perturbative calculations don't converge; the sum of quantum terms will give infinity, an impossible solution.

As is often the case, the breakthrough for understanding QCD depended on a rethinking to determine which parts of the theory could be approximated and which had to be retained exactly. Kenneth G. Wilson, the Nobel laureate physicist who currently directs Cornell University's NSF supercomputing center, The Center for Theory and Simulation in Science and Engineering, found that a better approximation for QCD was lattice gauge theory. This theory approximates space and time by representing it as a discrete set of points instead of a continuum. This means abandoning certain aspects of a fully quantum approach, such as continuity and rotation invariance, but in turn, other intrinsic symmetries are retained.

The fundamental constituents of QCD are quarks and gluons. The challenge of QCD is to deduce the properties of protons and neutrons from the equations that describe their constituent quarks and gluons. Even with the lattice approximation, QCD can't be solved exactly, but the equations lend themselves well to computer simulation, and theorists have used tens of thousands of supercomputer hours running variously configured versions of these simulations and examining different aspects of QCD. They've obtained considerable insight, including a clear mathematical statement of what's called the quark confinement problem. The strong force attraction among quarks, a force exerted on them by their ubiquitous gluon companions, actually grows in strength as they separate from each other in distance.

"Trying to separate two quarks," writes Michael Riordan, in *The Hunting of the Quark* (NY: Simon & Schuster, 1987), "is a lot like pulling on two ends of a broken rubber band. You pull and pull, injecting more and more of your energy into the stretched band. To pull two quarks a foot apart, it would require the same energy as needed to lift an automobile a foot off the ground. Long before you can ever reach such a separation, the energy materializes as additional quark-antiquark pairs—other mesons, that is. The rubber band snaps!" Though their existence in theory explains many observed effects, quarks have not been observed directly. The theoretical explanation for this is quark confinement—it has not been possible to shake loose a single quark from the protons and neutrons which contain them.

Simulating the Strong Force

The Quark-Gluon Plasma Phase Transition of QCD

Rajan Gupta, Los Alamos National Laboratory
Gregory W. Kilcup, Brown University
Stephen R. Sharpe, Stanford Linear Accelerator Center

Quark confinement is the theoretical issue at the nub of the phase transition which Gupta, Kilcup, and Sharpe have been examining at PSC. That transition to a high-temperature momentary state known also as quark-gluon plasma holds a great deal of interest for high-energy physics—"the Holy Grail," reported the March 1988 issue of **Physics Today**. It is a problem of enough interest that the Department of Energy has approved construction, though funding must still be raised, of a relativistic heavy-ion collider (RHIC) at Brookhaven National Laboratory in New York. This collider, designed especially to detect this new form of matter, will achieve ten times the energy per nucleon currently available to Thompson's group at CERN.

The simulations of Gupta, Kilcup, and Sharpe extend theoretical studies of simplified quark-gluon interactions which have indicated that a transition to quark-gluon plasma does occur; the simplifications involve ignoring the possibility that gluons could produce quark-anti-quark pairs. Would a full quark-gluon interaction change the prediction? The literature divides on this point. What is certain, however, is that the complexity of the full interaction increases the computational requirements markedly. Studies up to now have been forced to approximate full QCD in one way or another in order to complete the calculations, and these approximations, Gupta points out, introduce "systematic errors" into the results.

A New Algorithm. "The particular twist we add to the study of finite temperature QCD," writes Gupta, "is that we use an exact algorithm." The simulation performed by Gupta and his colleagues treats the full interaction without approximation beyond what is built inherently into the lattice gauge approach. They computed what's called the determinant, a factor necessitated by the presence of quarks in the model, through an accurate algorithm. This is a large calculation which most other studies have approximated crudely or ignored. Gupta and his colleagues used it as the basis of their Monte Carlo simulations, so-called because the mathematical form is based on chance ("stochastic") processes which allow a system to unfold and tell its own story. Monte Carlo simulations are the primary computational mode of addressing QCD theory. In exchange for not approximating the determinant, their simulation compromised on lattice size. Otherwise, the computational requirements would have been unmanageable.

For their first round of calculations, Gupta's group used a lattice with four points in each dimension. As a general rule, the lattice chosen for QCD Monte Carlo simulations would not be smaller than 16 points in the space dimensions; this is considered the sufficient size to eliminate the possibility of inaccuracy due to "edge effects." The results of their simulation, as they themselves recognize, must be interpreted with the understanding that lattice size is a limiting factor, though this hardly diminishes the significance of a fully exact treatment. For different values of the parameters of the theory (quark mass, m_q, and coupling factor, β), they did hundreds of Monte Carlo sweeps, each sweep representing a change of values for each variable.

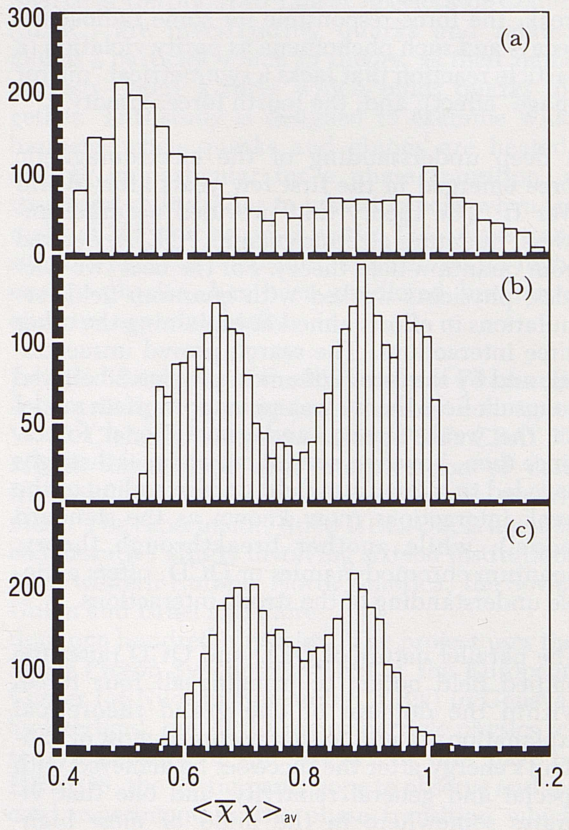


Figure 1: Number of sweeps in which the chiral condensate falls in a certain range for (a) a 4×4^3 lattice, (b) a 4×6^3 lattice and (c) a 4×8^3 lattice.

Signs of the Phase Transition. Their calculation on the 4×4^3 lattice indicates that a phase transition occurs. This can be seen in the histogram of Figure 1(a). The valley between two peaks for the chiral condensate $<\overline{\chi}\ \chi>$ indicates a back and forth movement between different phases. The chiral condensate is one among a number of numerically calculated properties of an elementary particle system which a high-energy physicist reads carefully to gather bits and pieces of the story of the system being examined. When the peaks and valleys of chiral condensate occur as they do in 1(a), at a constant β, it represents an analog to H_2O at 32F, a point at which two states or phases, ice and water for H_2O, exist simultaneously.

When completed, this exact simulation for the 4×4^3 lattice provided Gupta and his colleagues with a point of reference from which to verify and fine tune another algorithm, the Hybrid Monte-Carlo algorithm (HMCA), which Gupta believes to be "the most promising candidate for an exact algorithm which can be used on large lattices." Test runs of HMCA showed a speed-up of two orders of magnitude on the 4×4^3 lattice, with results that correlated with the exact determinant algorithm. Gupta's group then extended their calculations, using HMCA for a 4×6^3 lattice and, for selected values of β and m_q, a 4×8^3 lattice.

They then fall rapidly and fluctuate around a low value ($\sim .62$) for about 1000 sweeps before rising again to a slightly higher value than before ($\sim .9$). This fluctuation indicates what's known as a metastable state; properties of the system bounce back and forth but remain within bounds. This indicates the presence of the plasma phase transition.

The Figure 1 histograms show averaged chiral condensate results, i.e., number of sweeps at a particular $<\overline{\chi}\ \chi>$, for the 3000 sweep runs at $m_q = 0.1$ on all three lattices. Comparing these results leads Gupta to qualify his study: "The figures make clear that the discontinuities are decreasing with volume." Note that the range between peak $<\overline{\chi}\ \chi>$ values decreases from (a) to (b) to (c). "We certainly cannot rule out that the two maxima in the histograms coalesce as the volume increases and the observed metastability is a finite volume artifact. Unfortunately, only runs on much larger lattices ($N_s > 16$) will settle the issue."

The Benefits of Supercomputing. The approach of Gupta and his colleagues, even using HMCA, is feasible only on a small lattice. For larger lattices, the computational requirements exceed those available in today's supercomputers. As it is, these studies have consumed over a thousand hours of processor

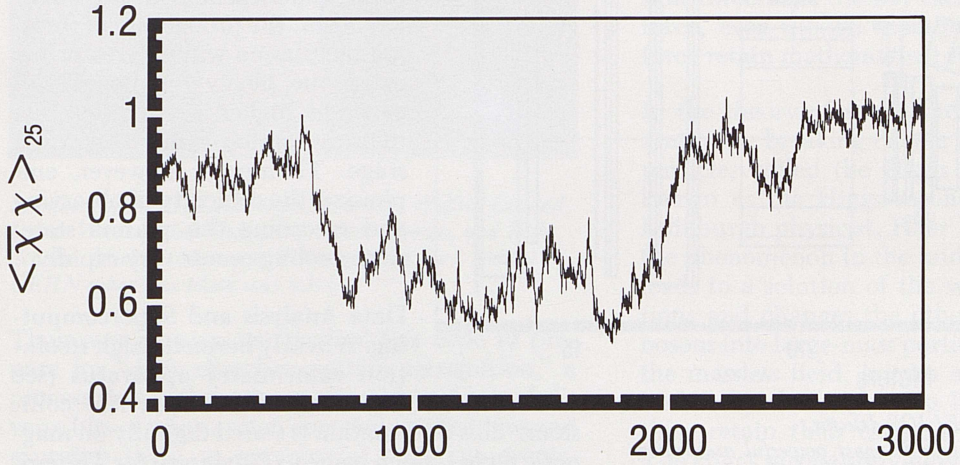


Figure 2: Time history for chiral condensate, $<\overline{\chi}\ \chi>$, on a 4×6^3 lattice, where quark mass = 0.1 and coupling factor = 5.13.

Results for the two larger lattices agree with the 4×4^3 simulation. The phase transition for both is indicated by histograms 1(b) and 1(c), which corroborate the initial simulation in showing a valley and peaks for the chiral condensate. For the 4×6^3 lattice, Figure 2 represents a sweep-time history of $<\overline{\chi}\ \chi>$. For this run, and for corresponding runs at the same m_q and β on the other two lattices, the simulation passed through three thousand Monte Carlo sweeps, a number that was necessary, as Figure 2 shows, for the chiral condensate to complete what appears to be a cycle between two phases. You can see how the $<\overline{\chi}\ \chi>$ values fluctuate, first around a high value ($\sim .84$).

time at the PSC, using very highly optimized code. The benefits to the scientific community extend far beyond the specific results reported here. The speed of the supercomputer has allowed Gupta's group to study what is, in effect, another approximation of the full quark-gluon interaction. The payoff includes a deeper understanding of which approximations are most reliable and what computational techniques will speed up execution in each of the various approaches. Gupta, Kilcup and Sharpe's algorithmic research will benefit all subsequent researchers who want to simulate other conditions and answer other questions about QCD.

Observing the Strong Force

HELIOS, Smashing Quarks to Plasma

Julia A. Thompson, Wilfred E. Cleland, Mark Clemen,
Michael Murray, Yeong Mouk Park, Beth Scholle and Lynne Weber, University of Pittsburgh

HELIOS is a two-stage experiment. The first stage examines leptons, the elementary particles—electrons, neutrinos, muons, and their antiparticles—which don't "feel" the strong force that binds protons in the nucleus. A proton beam from the SPS at two different energies, 200 GeV and 450 GeV, the highest available at CERN, strikes various metal targets. In the second stage, heavy ions are accelerated onto a lead target at an energy of 200 GeV/nucleon. GeV represents giga electron volts, with one giga equal to a thousand-million, 10^9 eV. Using this amount of energy, the experiment is designed to create conditions favoring the observation of quark-gluon plasma. The Pittsburgh group headed by Thompson has responsibility for coordination of data collection and for several aspects of data analysis. For the first phase, Thompson's group is examining a low transverse momentum region where previous experiments observed electrons in excess of what was expected. This data will bear on several questions of interest in the field, among them quark confinement.

Thompson and her group will attempt to determine if the excess electron production may be related to quark-gluon plasma. There is reason to believe that electrons of low transverse momentum, such as those recorded by HELIOS, may be related to a post-collision intermediate stage of the plasma state, which as it cools will, according to theory, in effect "freeze out" some of the transformed nucleon particles released in the collision. In the second stage of the experiment with the heavy ion beam, the probability of forming the plasma will be greater because the heavy ion beam will generate higher energy densities than the proton beam of the first stage. Thompson, however, emphasizes the difficulty of observing and recording this plasma state, since cooling occurs very rapidly.

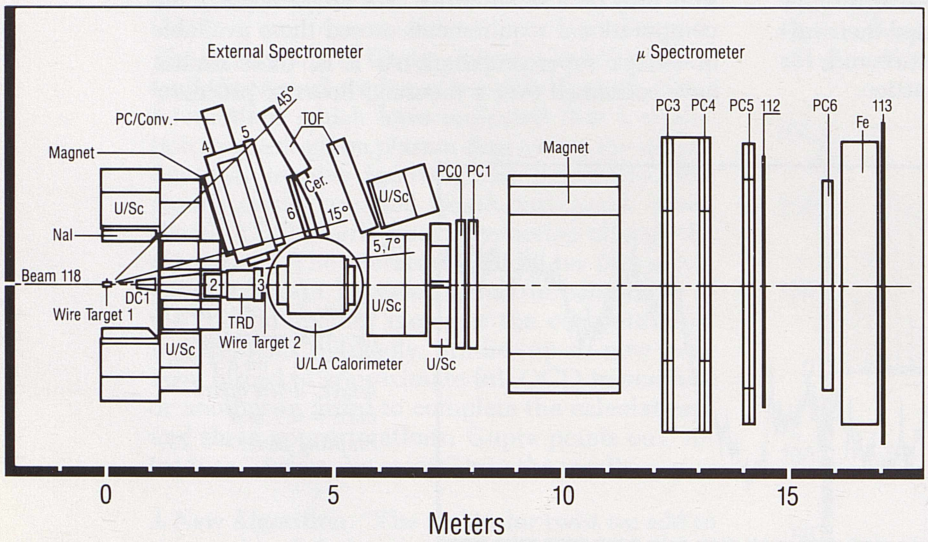


Figure 3: Schematic Drawing of the HELIOS detector.
This system of detection at SPS, the world's most powerful accelerator, measures each particle about 230 times, with one-half of those measurements representing built-in redundancy. The proton beam has a spot diameter less than 50 microns. Target 1 is a 50 micron beryllium wire, barely visible, about the length of a pencil. The proton beam strikes the wire end-on, and one proton smashing another causes a spray of 20 to 30 smaller particles. The drift chambers measure particle position to an accuracy of within about 100 microns. The Calorimeters detect charged reaction products and measure their energies. The transition radiation detector (TRD) allows separation of electrons from the more slowly moving and more common pions. The muon (μ) spectrometer measures muons produced in the experiment.

Data Analysis and Supercomputing. A nearly hermetic high resolution calorimetry apparatus (see Figure 3) records particle collisions. This information is stored digitally on magnetic tapes, which come to Pittsburgh for Thompson's group to analyze. They have tested a package of data analysis routines which they have used for a series of preliminary passes on the data. The routines are designed to assist in reliably eliminating data that shows collection errors or holds no analyzable results. From 1986 alone, Thompson's group has three-million triggered events, recorded on 200 tapes. To date, they have compressed this to about 60 tapes, a more manageable quantity. HELIOS generated additional data during the summer of 1987 and will be collecting data again this summer.

Much of the preliminary data analysis routines don't use the CRAY's vector capabilities effectively, a notorious problem. The solution would

be a boon to particle physics. These routines and the "drift chamber code," a complex set of algorithms used for final analysis, were originally written for the VAX 8650. One of Thompson's objectives is to adapt this code to the CRAY in preparation for CERN's own newly installed X-MP/48, which as of April 1988 was undergoing testing and was not yet available to users. The group has gained a speed-up through optimization of about 30% on the analysis routines, a result which Thompson finds disappointing since vectorized programs usually yield greater improvement than this. Beth Scholle of Thompson's group obtained better results from her work with the drift chamber code, an early version of which she optimized by a factor of 20. On scalar speed alone, PSC's CRAY ran this code twice as fast as the VAX; and with optimization, the total speed-up was a factor of 42. The group is now working with a newer version of this code which is less CPU intensive.

Members of the HELIOS team (L to R): Wilfred Cleland, Lynne Weber, Julia Thompson, Mark Clemen, and Beth Scholle. Michael Murray and Yeong Mouk Park were at CERN when this photo was taken.

Thompson's group has studied this code to prepare for "drift chamber pattern recognition," a process that gathers the many ambiguous and diverse bits of information encoded on the tapes and expresses them as meaningful coherent events. Drift chambers like those used by HELIOS at CERN (see Figure 3) are the most precise measuring devices available for particle collision data over any substantial volume. They measure spatial position of particles with accuracy to the order of 100 microns. Nevertheless, error in interpretation can occur easily due to the quantity of particle tracks recorded and the many possibilities for ambiguity and particle misidentification, which necessitate many different measuring devices.

Thompson explains that the HELIOS apparatus measures each track about 234 times. Different particle types trip different measuring devices, and the devices are arranged to cover a variety of possibilities. These measurements become the recorded data on the tapes, but there's no way to know which measurement goes with which track, and the points can be grouped many ways. The drift chamber code tries many configurations and has a number of algorithms to analyze the data. The speed-up of this code which Thompson's group achieved, plus the intrinsic speed of the CRAY, lets them try several analysis algorithms; VAX speeds would restrict them to one analysis. The difference makes it possible to do thorough, systematic comparisons of data ambiguities in much less time.

A Layman's Guide to the Higgs Boson

As they have with the strong force, physicists also have developed remarkable insights into weak interactions using quantum field theory. Many physical quantities can be computed using perturbation methods similar to those which work with QED for calculating electromagnetic fields. Several effects of the weak interaction, however, don't fit with perturbation theory. One of these, usually called spontaneously broken symmetry, involves the idea that equations of a physical system may retain a certain symmetry, even when the solution doesn't. For example, the laws describing magnetic force have rotational invariance; i.e., the equations have no concern about the direction of the force, one direction is as good as any other. But a magnet magnetizes only in one of many possible directions, thereby exerting an asymmetrical force, even though the equations describing the force retain mathematical symmetry.

In the theory of weak interactions, spontaneous symmetry breaking occurs by introducing a new particle, called the Higgs boson. This effect is known as the Higgs mechanism, so named for Edinburgh physicist, Peter Higgs, who identified the phenomenon in the middle 1960s. This effect leads to a solution of the weak force field equations and changes the otherwise massless vector bosons into large-mass particles. The symmetry of the massless field, known as gauge symmetry, is broken by the large mass bosons, but the equations retain their essential algebraic symmetry. The Higgs mechanism allows calculation of many physical effects and extends the viability of weak-force field theory, even without a detailed understanding of the interactions of the Higgs boson. But to understand the phases in weak interactions, it's necessary to understand the interaction between the Higgs boson and other particles of the theory. Since this is a non-perturbative problem, it's not surprising that lattice approaches have been used.

Simulating the Weak Force, With Bosons and Quarks

Interactions in Lattice Gauge-Higgs with Fermions

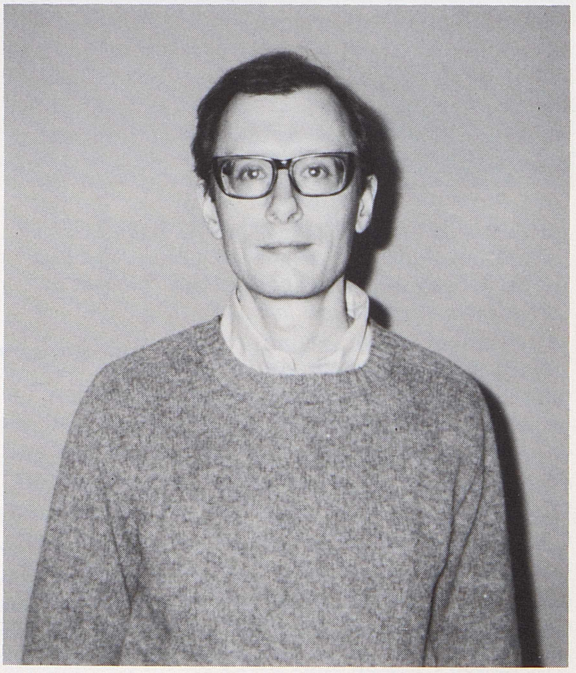
Robert E. Shrock, State University of New York, Stony Brook
I-Hsiu Lee, Brookhaven National Laboratory

Robert E. Shrock and I-Hsiu Lee have carried out a series of computational studies at PSC which have contributed significantly to an understanding of these interactions. This work builds on Lee's work with J. Shigemitsu, the first to analyze weak interactions including the Higgs bosons and quarks with the vector bosons of weak force field theory. These vector bosons are analogous to gluons in QCD in that they exert force on other particles and are not, like fermions—quarks and leptons—subject to the Pauli exclusion principle.

Previous Work. The Lee-Shigemitsu studies formed the opening round in a beautifully developed series exemplifying the synergy between computation and analytic reasoning. Lee and Shigemitsu showed numerically that the phase transition to broken symmetry of the weak interaction depends strongly on whether quarks are included. Lee and Shrock followed with analytic studies aimed at understanding the consequences of the strength of the interaction between the Higgs and vector bosons relative to that between the vector bosons and quarks. These studies, which simplified analysis by ignoring the vector boson interactions with each other, indicated that the phase structure of the system, as determined by the transition to broken symmetry that occurs with the introduction of the Higgs boson, depended on the strength of the Higgs to vector boson coupling factor, β_h.

The PSC Project. Lee and Shrock confirmed and extended this insight with a set of simulations that included quarks in various couplings with vector bosons and with self-coupling among the vector bosons. In these simulations, performed on PSC's CRAY X-MP/48, Shrock set up a lattice with 8 points in each dimension and swept the lattice thousands of times for each set of parameters studied. For different values of the Higgs coupling, the numerical results allowed comparison with the analytic results for the relation between β_h and the broken symmetry phase transition. The quark-vector boson interaction of the simulation ignored the possibility that quark-anti-quark pairs could be produced. This approximation of the full quark interaction, called the quenched approximation, is necessary because the unquenched interaction requires an unreasonable amount of computing time, even on the CRAY X-MP, an indication of the magnitude of elementary particle calculations, even for a relatively small (8^4) lattice. The quenched approximation seems to be a reasonable approximation to make in this case, since the simulation results agreed with the analytical study, which included quark pair production.

Shrock and Lee's initial round of studies assumed a simplified set of symmetries for the weak interactions, U(1), defined in a form of advanced algebra known as group theory. The next logical step was to examine the same system of interactions using the more realistic SU(2) symmetry. With the more complex SU(2) formalisms governing the process, the calculations also become considerably more complex. These calculations, however, are within the range of X-MP/48 supercomputing, and the analytic studies were again extended and tested through Monte Carlo simulations on an 8^4 lattice. Figure 4 shows the results.



Robert E. Shrock

Results. Different values of β_h yield a corresponding result for chiral condensate. The relation between these two parameters, as the chiral condensate approaches zero, determines β_h for the onset of Higgs spontaneously broken symmetry. From the data shown in Figure 4, Shrock and Lee determined the critical point of β_h to be 2.75 ± 0.2, a result in excellent agreement with

the corresponding analytically determined value of 2.76 ± .02. As in the simplified U(1) studies, the quenched approximation required to perform the simulation seems to have negligible effect on the result, judging from agreement with the analytic calculation, which is unquenched, and it appears that quark pair production is unimportant in the phase structure of fermion-boson weak-force interactions.

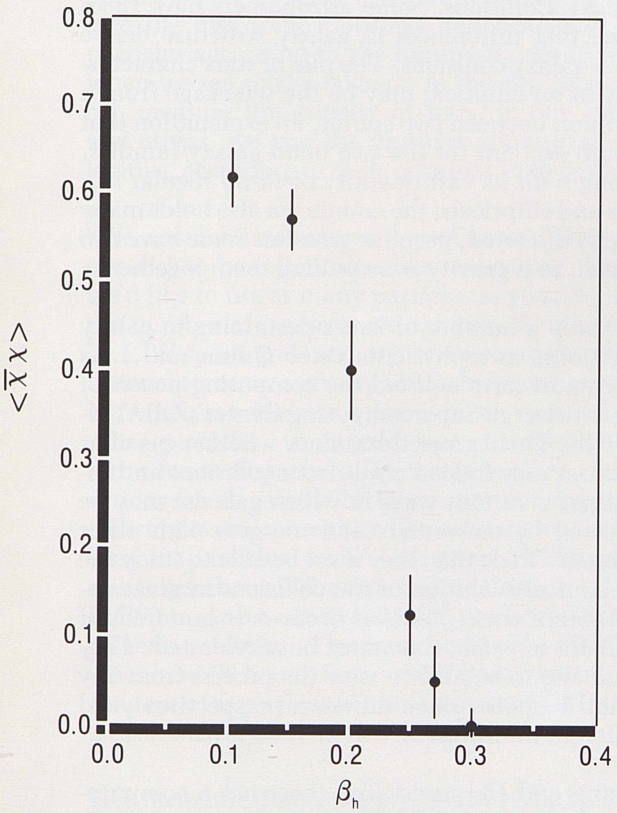


Figure 4: Chiral condensate $<\bar{\chi}\chi>$ (vertical axis) as a function of Higgs coupling, β_h, (horizontal axis) determined by Monte Carlo simulation for symmetry SU(2).

A New Insight. Lee and Shrock also performed these calculations assuming that quarks belong to higher representations of SU(2), an assumption that results in higher strengths for quark coupling with vector bosons. As it turns out, the theory is consistent in the latter case only for values of quark mass much larger than real quark masses, as determined experimentally (Accelerator studies show the indirect effects of quarks and provide data about them even though these reclusive particles are themselves unobserved). Until this study, high-energy physicists had no reason to suspect such a relation between quark coupling to vector bosons and quark mass. This sort of information, because it derives from calculations using parameters different from those observed in nature — in this case, quark-vector boson coupling — gives insight into theory which cannot be acquired experimentally.

In this sense, Shrock's work at PSC illustrates how computational methods are an invaluable complement to experimental methods in ongoing high-energy physics research. Commenting on his studies with Lee, Shrock says they've yielded information "about the nonperturbative behavior of gauge-Higgs theories with fermions, and, by including fermions, have advanced lattice theories a step closer to the real world."

REFERENCES:

R. Gupta, G. Guralnik, G. Kilcup, A. Patel and S.R. Sharpe, "On the finite temperature transition of QCD," LA-UR-87-3068, to appear in ***Phys. Lett. B.***

B. Schwarzschild, "Nuclear Matter in Extremis Sought with Ultrarelativistic Heavy-ion Beams," ***Physics Today*** *(March 1988) 17-21.*

R.E. Shrock, "Lattice Higgs Models" (Invited Review Talk at the International Symposium on Field Theory on the Lattice, Seillac, France, 1987), ***Nucl. Phys. B,*** *in press.*

R.E. Shrock and I-H. Lee, "Analytic and Numerical Study of the Chiral Transition in SU(2) Gauge-Higgs-Fermion Theories," ***Phys. Lett.*** *201B (1988) 497-502.*

Supercomputing in the Cosmos

Sinking Satellites of Spiral Galaxy Systems

Peter J. Quinn, Space Telescope Science Institute, Baltimore
Lars Hernquist, Princeton Institute for Advanced Studies

Introduction: A Universe of Galaxies

During the 1920s, astronomers began to realize that many of the diffuse masses known as nebulae, which they sketched and photographed through their telescopes, were a great distance beyond our own Milky Way. Close inspection of the nearest of these revealed that they were galaxies, composed of large numbers of stars, dust and gas. As the power of telescopes increased and photographic emulsions became more sensitive, astronomers began to see the universe as composed of many galaxies, great distances apart and seemingly self-contained—"island universes" within a vast cosmic sea.

Astronomers and astrophysicists now recognize galaxies as the macroscopic building blocks of the universe. Questions about how they form and evolve can lead to a deeper understanding of nature. When in the history of the universe did galaxies first appear? Are those we see today like the first ones? What are the differences among different galaxies? Although they look like islands in a vast sea of space, are they really isolated, self-sufficient entities? To what extent have external processes like gravity or other forces influenced them?

Two Kinds of Galaxies. With respect to galaxy structure, we're most familiar with the disk-like spirals. Astronomers call these systems "spiral galaxies." Consisting of gas, dust, and many young, blue stars, these spiral galaxies represent one of the two main families of galaxies, the other being "elliptical galaxies." We reside in a spiral, the Milky Way, with a mass about 10^{11} times the sun, a diameter of 100,000 light years, and a disk thickness of 3000 light years. In its center, the Milky Way has a spheroidal "bulge" of old stars with a radius of about 6000 light years. Similar bulges occur commonly in spirals.

Astronomers believe ellipticals are ellipsoids, three-dimensional football shaped distributions of stars which appear spherical or elliptical in flat projection on the sky. Ellipticals may be as large in size and mass as spirals, but consist mostly of old stars. They lack the gas and dust common to spirals. Ellipticals commonly favor areas of the universe where galaxies congregate at high-densities. Spirals, on the other hand, head for the suburbs; they prefer low density environments outside large galaxy clusters.

Galaxy Collisions. Some astronomers have theorized that differences in galaxy structure derive from galaxy collisions. The pile of stars characteristic of an elliptical may be the wreckage from a collision between two spirals, an explanation that would account for the two main galaxy families. Along with its vast quantity of fairly regular spirals and ellipticals, the cosmic sea also holds many highly distorted, peculiar galaxies. Some have two nuclei, as if gravity were pulling them together.

To study a number of issues pertaining to galaxy collisions, astrophysicists Peter Quinn and Lars Hernquist have enlisted the computing power of the Pittsburgh Supercomputing Center's CRAY X-MP/48. Their project examines whether peculiar galaxies may indeed result from collisions and investigates various ways in which galaxies may be affected by encounters and mergers with other galaxies. To do this, they must be able to study the dynamical evolution of the collisions in great detail. Since these processes occur over hundreds of millions of years, time must be accelerated. They need also to be able to view the process from any point in space, since different perspectives yield different information.

Quinn and Hernquist first construct a computational model of the galaxies they want to study. They choose an initial distribution of particles, which represent stars, and apply the laws of gravity to evolve the galaxy systems forward in time. With the large number of particles involved, this is an extremely complex undertaking, feasible only with the supercomputing power of a CRAY or equivalent state-of-the-art system.

Building Galaxies Using Tree Structures

Galaxies like the Milky Way have about 10^{11} stars orbiting their centers on roughly circular paths. Each of these stars exerts a gravitational attraction on all the other stars in the galaxy, and this collective gravitational field holds each star in orbit. For a particular star, this gravitational tug may be thought of as having two components: the smooth field of many distant stars and the occasional hard tug when a particular star comes close. Calculations show it would take longer than the age of the universe for these occasional hard tugs to have a significant effect. Quinn and Hernquist's model, accordingly, should treat the gravitational effect of one star as trivial compared to the overall force exerted by the entire field of

many, many stars. The model should represent galaxies as smooth, "collisionless" stellar systems.

Quinn and Hernquist's model galaxy takes a number of other factors into account. First, explains Quinn, you take the number of particles, N, needed to represent the galaxy's stars and distribute them into a disk-shaped space. "Since spirals are observed to decline in surface brightness—stars per unit area—exponentially as a function of radius, we do the same with our particles. We also give the disk some thickness and distribute the particles according to the luminosity we observe when we see spirals edge-on in the sky. Since we can measure the velocity of stars, in our galaxy and others, we give the particles appropriate velocities, according to their positions in the disk.

"We now face a dilemma," says Quinn. "We want our disk to be collisionless like real spirals. Hence we'd like to use as many particles as possible. But the number of particles we can use, N, is limited by CPU speed and memory. If we compute gravitational force on a given particle by classical physics, applying Newton's formula to each of the other particles, we would have to compute N terms for each particle. To compute all the interactions would require N^2 terms. For our disk to be collisionless for more than 10 rotations, we need at least 10^4 particles. An N^2 code using 32,768 particles takes about 750 days of CPU time on a VAX 11/780 to evolve one disk 6×10^9 years or about 30 galactic rotations. The same code on a CRAY 1 uses about 90 hours. Since we'll need, typically, to run tens of experiments, this amount of CPU time is prohibitive, even on the CRAY."

Tree Structures. Fortunately, a new technique for computing gravitational interactions gives Quinn and Hernquist a way around this dilemma. "The technique we use," says Hernquist, "employs a data structure known as a tree to reduce the number of interactions that need to be computed. The tree structure organizes particles into groups. The further away a group is from a given particle, the more the group acts like a single large particle and the less we need to worry about the distribution of particles in the group. The net result is that for each of our N particles we need to compute only in the order of log N terms."

For a 10^4 N system, in other words, Quinn and Hernquist can apply the tree structure method and compute about 4 terms instead of 10^4 terms for each particle. For the system as a whole, they would compute about 40,000 ($10^4 \times 4$, N log N) terms instead of N^2 or a hundred million terms. They now can run the 32,768 particle simulation in under 3 CPU hours on PSC's CRAY X-MP. Because of this CPU time advantage, which is gained with an accuracy essentially equivalent to the N^2 method, tree structure may become very important in astrophysical studies.

Tree Structures and the CRAY X-MP. Hernquist has studied tree structures in detail and found that the improvement in CPU time comes "at the expense of a small, controllable error in the force computation." He sees this computation method as particularly applicable to CRAY architecture. "Given the rapid rise in interest in parallel computation, it is indeed possible that tree codes may represent the future of N-body simulations over the next decade." "Machines like the X-MP," says Quinn, "are at present the only way many of the most important questions in galaxy formation and evolution can be tackled."

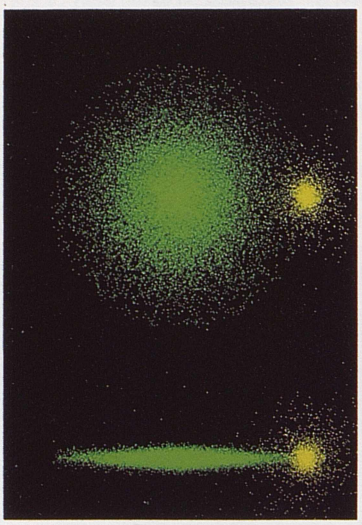
Peter Quinn seated at Wombat, his Sun workstation, with "my colleague, Dr. Wallace Wombat" on top. Quinn is Australian and says wombats are somewhat like beavers, "strong and determined yet mild and friendly."

A Satellite Sinking and its Parent Spiral

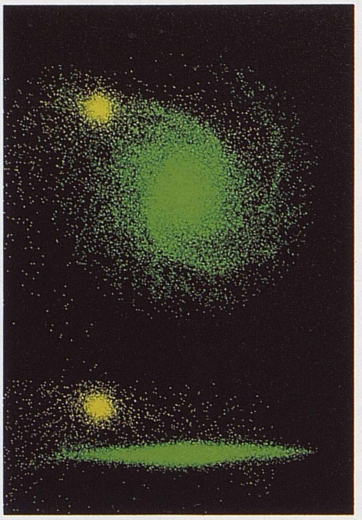
Most large galaxies have smaller satellite galaxies orbiting around them, and the merger of a satellite with its parent represents one likely scenario by which a galaxy collision could occur. The Milky Way, for instance, has two main satellites; the Large and Small Magellanic Clouds. The Large Magellanic Cloud (LMC) is a spiral with a mass about 4% of the Milky Way.

One of Quinn and Hernquist's galaxy modeling studies examines evolution of a satellite as it enters its parent spiral's gravitational pull. For this project, they had to model the satellite as well as the parent spiral. For simplicity, they chose a spherical satellite with a total mass about 10% of the disk mass. Based on observations of satellites, they set the mean density of particles equal to the peak central density of the disk. This means the central regions of the satellite have higher density than the disk, explains Quinn, "and we would expect some piece of the satellite to survive the destructive effects of the encounter."

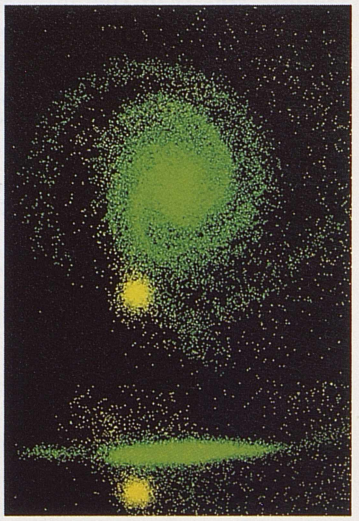
Based on observations of the LMC, Quinn and Hernquist gave their 10% mass satellite a circular



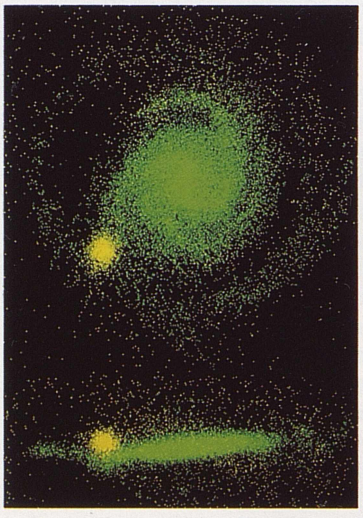
t = 0



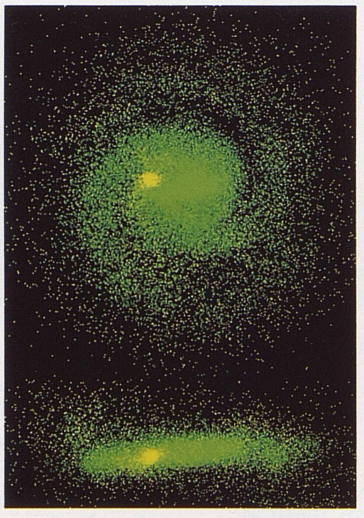
t = 360



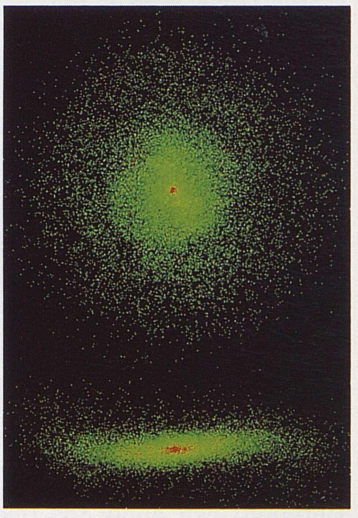
t = 720



t = 1260



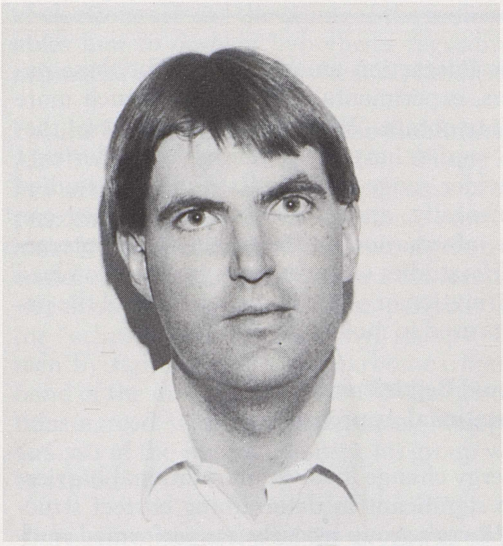
t = 3000



t = 3420

Time is in units of a million years (10^6), and each frame represents about 180,000 light years across. The x-y plane view in the upper section of each frame represents 180,000 light years along the vertical axis. The lower edge-on view is from the same spatial scale as the upper view and represents about 70,000 light years in its vertical axis. The disk and satellite rotate counter-clockwise. In the final frame, to heighten clarity, the satellite particles are red.

orbit at the edge of the disk inclined at a 30° angle with respect to the disk. If the satellite had zero mass, it would orbit forever at the initial fixed radius. The axis of the orbit would move slowly around the disk, like a precessing top, but the satellite radius wouldn't drift in or out. When the satellite has mass, however, the stars in the satellite and disk tug on each other, and their systems begin to distort in response to these tidal forces.



Lars Hernquist

Gravity and Collision. The six-frame sequence shows the evolution of the disk and satellite. Each frame shows a view from above looking onto the plane of the disk with a corresponding edge-on view below it. Time ticks in units of a million years, and each frame is about 180,000 light years on an edge. As you can see, the disk responds violently to the presence of the satellite. "Imagine you are riding along with a particle at the outer edge of the spiral," says Quinn. "As you see the satellite ahead of you, in the direction of your motion, you feel a pull in the same direction you are traveling around the disk. As a result, you accelerate and move out to a larger radius behind the satellite. If the star you are riding were initially ahead of the satellite then you would feel a tug in the direction opposite to your orbital motion, and you would lose velocity and begin to sink into the disk ahead of the satellite.

How does this affect the satellite? The disk builds up particles behind the satellite and reduces particles ahead of it(t = 360-720). This means the satellite will feel more particles pulling it in the direction opposite to its orbital motion than in the forward direction. This backward tug causes the satellite to lose velocity and sink closer to the disk. The edge view shows the same simulation from a point in the plane of the disk. As the satellite plunges backwards and forwards through the disk plane, it pulls on the disk particles and causes them to oscillate with larger amplitudes through the disk (t = 360-1260).

The work that the satellite does in thickening the disk reduces its own orbital motion and causes it to sink into the plane of the disk. As the satellite sinks, the tides raised on it by the disk remove particles from its outer regions. These freed particles become part of an atmosphere of rapidly moving particles both in and above the disk(t = 3420). After about 3×10^9 years (t = 3000), the surviving core of the satellite has lost its kinetic energy and angular momentum and has settled into the center of the disk. In the process, the disk has become much thicker than it was originally (cf. t = 0 and 3420).

Results. Quinn and Hernquist relate the results of this study to observations of the Milky Way. The Milky Way's bulge and thick disk of old stars, like those of the satellite model, may be the tell-tale sign of past encounters with satellites. The satellite core that finds its way to the disk center will either increase the bulge or become the bulge of the expanded spiral galaxy. Since bulges are common to spirals, they may represent the undigested remains of previous encounters with satellites. If the satellite has gas and perhaps a massive object like a black hole, then it could spark activity in the nucleus of the spiral such as the optical, ultraviolet and X-ray radiation seen in many disk galaxies, including our own.

By studying spiral galaxies and their satellites and modeling these systems on the CRAY, Quinn and Hernquist aim to bring increased understanding to one aspect of the life history of the universe. Their work bears on many unanswered questions. Does capturing small satellite galaxies restructure a disk and provide gas to fuel some central engine like a massive black hole? What happens when colliding galaxies are nearly equal in mass and destroy each other? Is an elliptical galaxy formed? How were spiral galaxies formed in the first place? Did they grow by slow accretion of smaller systems of gas and stars, or were they entirely isolated clouds of hot gas that cooled and formed stars in a disk?

Astrophysics and Supercomputing. Such questions are at the forefront of modern astrophysics. "We're lucky enough," says Quinn, "to live when both the numerical methods, like tree schemes, and the hardware, the CRAY X-MP and beyond, can address the essential physics of these problems for the first time. By using the CRAY as a laboratory, we can conduct experiments, uncover new dynamical processes and come closer than we have to understanding the birth and life cycles of the building blocks of the universe, the galaxies."

REFERENCES:

L. Hernquist, "Performance Characteristics of Tree Codes," ***The Astrophysical Journal Supp. Ser.*** *64 (1987) 715-34.*

P. J. Quinn & J. Goodman, "Sinking satellites of spiral systems," ***The Astrophysical Journal*** *309 (1986) 472-495.*

Quantum Chemistry and Molecular Mechanics

Rotational Barriers in Substituted Compounds

Kenneth B. Wiberg, Mark A. Murcko and Keith Laidig, Yale University

The Relation Between First Principles Quantum Chemistry and Experiment

Kenneth Wiberg and his colleagues Mark Murcko and Keith Laidig have aimed a major part of their research at the interface between quantum and experimental organic chemistry. They've done first principles quantum computations on organic systems as a basis from which to evaluate experimental results and deepen understanding on a microscopic level beyond what experiment alone can give. On the empirical side, they've devised experiments to test their theoretical predictions and, in other cases, have formed working relationships with groups doing empirical research on similar problems.

Molecular Mechanics and Experimental Data. A computational method called molecular mechanics has become increasingly important in the study of organic compounds, including proteins. Molecular mechanics reliably characterizes the structure (conformation) and energy of compounds containing just carbon and hydrogen, but has been less reliable with "substituted compounds," so named because oxygen (O) and nitrogen (N) substitute for some of the hydrogens (H) or carbons (C) in common hydrocarbons. For substituted compounds, Wiberg has found that molecular mechanics is hampered by lack of detailed experimental data from which the needed parameters may be derived.

Molecular mechanics describes a molecule in terms of classical Newtonian equations and depends upon experimental data to provide physical parameters for the interactions among the molecules' constituent atoms. These intramolecular parameters include equilibrium bond lengths and angles between atoms of the molecule, and force constants, which describe the energy change caused by perturbation away from equilibrium values. Experimental data from infrared or Raman spectroscopy will frequently give these numerical values.

Molecular mechanics depends also on other quantities describing other molecular interactions: rotational barriers (the energy required to rotate one group of atoms about another with bond lengths and angles constant), interactions with nonbonded nearby atoms, and the Coulomb or electrostatic interaction among atoms. For these parameters, experimental data may be much more difficult to obtain. Wiberg believes that here "theoretical studies have the potential for exploring a much wider range of cases than may be studied experimentally, and at the same time they can provide information on the origin of the interactions." His studies of particular systems often lead to confirmation or questioning (or both) of the parameters used in molecular mechanics.

Rotational Barriers and the Computational Approach

The energy change involved in rotational barriers is often significant in determining correct structure. Wiberg's group recently has performed studies at the Pittsburgh Supercomputing Center concentrating on this aspect of intramolecular potential energy. "It's the availability of the CRAY," says Wiberg, "that's made it possible to do many of our studies." Their approach has been "to first determine the level of theory required in order to reproduce the known experimental data and then use this level to explore cases for which experimental data are not available." In practice, this has meant precise specification among the levels of approximation available in GAUSSIAN, the quantum chemistry software package developed by John A. Pople and colleagues at Carnegie-Mellon University.

Wiberg's group begins with the Hartree-Fock approximation of full quantum theory, standard in much quantum chemistry work, and adjusts the initial guess for the geometry systematically through a number of iterations to find the structure with the lowest energy. This conformation should represent nature's choice, the ground state of the system, which derives from quantum principles analogous to the law that water runs downhill. It can be difficult to know if a particular low-energy conformation is actually the ground state or merely a local minimum, a low energy resting place along the quantum path to the ground state. The researcher's comprehensive understanding of his system, as well as intuition, come into play in determining if the ground state, in effect the true geometry of the molecule, has been reached.

Wiberg's approach often involves several rounds of GAUSSIAN calculations in order to arrive at

the best agreement between the theoretically computed rotational barrier and the experimentally derived value. Once the computations have been validated in this respect, Wiberg's group uses them as a model for other parameters of the system. The model provides a multitude of information which may be evaluated against experimental data. In many cases, Wiberg's analysis of the computed electron densities, the distribution of electrons around the atoms of the molecule, enables him to develop hypotheses regarding data anomalies and to posit physical explanations.

Results for Two Studies: a. The Conformation of Aldehydes. In their work at PSC, Wiberg's group has arrived at explanations for some previously puzzling experimental observations. From experiment it's known, for instance, that a group of molecules called "aldehydes" prefer a conformation in which an alkyl group (-CH_3) is perfectly aligned (or "eclipsed") with a carbonyl group (-CO), as seen by sighting down the carbon-carbon (C-C) bond of the molecule, Figure 1. Wiberg's calculations agreed with this preference. Moreover, by analysis of the electron density, his group was able to explain it.

The electro-negativity of the oxygen in aldehydes creates an "electric dipole moment." A dipole moment is a separation of electric charge that behaves in many ways like a bar magnet, with the direction of the "electric magnet" pointing from negative to positive electric charge. The dipole created by the oxygen substitution also causes further charge separation in the rest of the molecule. The laws of electrostatics indicate that ground state configuration will have dipole moments aligned but oppositely-directed, like bar magnets aligned with opposite poles together. In aldehydes, this implies the eclipsed conformation. The staggered conformation is electrostatically unstable and will rotate counterclockwise about the C-C bond until the molecule stabilizes in the eclipsed position.

b. Amides and Resonance Theory. Wiberg's group also studied a group of molecules known as "amides," examining the effects of rotation about the carbon-nitrogen bond. The rotation barrier for the C-N bond of the simplest amide, formamide (Figure 2) has been experimentally determined, and Wiberg's theoretical computations yielded a rotation barrier in satisfactory agreement with the experimentally observed value. With their model validated in this respect, they were able to show that conventional notions about the distribution of electrons in the stable conformation of formamide were incorrect.

One established theory, known as "resonance," predicts that electron density at the nitrogen atom in formamide should be less in the ground state than in the higher energy (saddle point) conformation. Wiberg's group found just the opposite:

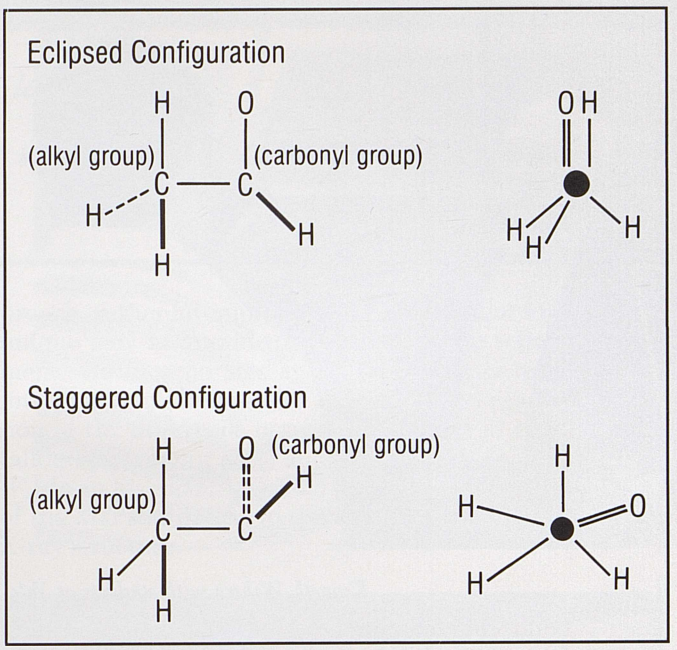


Figure 1: Acetaldehyde in two possible conformations: eclipsed and staggered. The eclipsed conformation is preferred by aldehydes. The views on the right indicate how the molecule would look if the viewer sighted along the C-C bond.

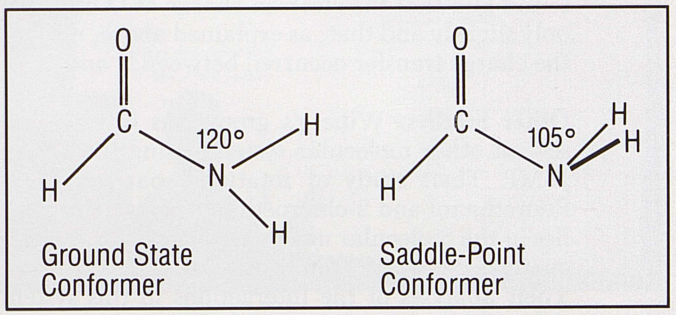
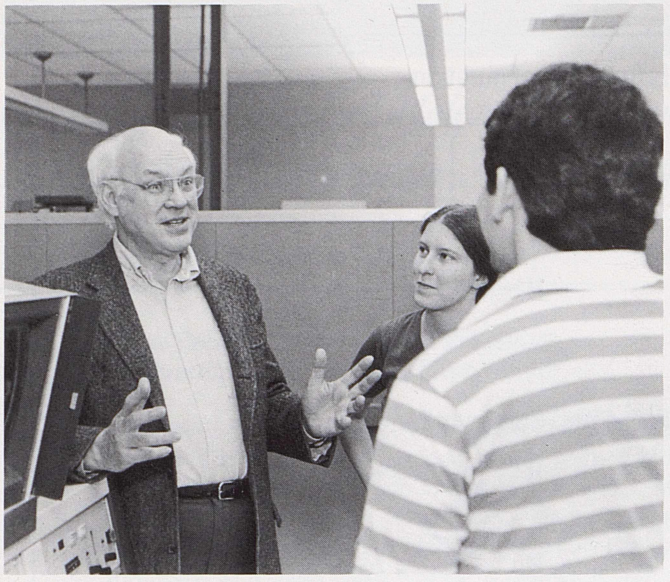


Figure 2: Formamide ($HCONH_2$) in its lowest energy, or ground-state, conformation (left) and in the saddle-point conformer (right), the higher energy conformation for rotation about the C-N bond. The ground state is the preferred geometry.

the nitrogen atom was more negative in the ground state. They believe this is due to the "hybridization" at the N atom, meaning that its electron orbitals are a combination of spherically shaped *s* and figure-eight shaped *p* orbitals. The N atom in the ground state has a larger percentage of *s* orbitals, in which electrons bind more tightly than *p*, and this gives a viable explanation for higher electron density in the ground state conformation.

Conventional resonance theory predicts also that formamide in the ground state will transfer charge along the N-C C = O axes. The theory says that this charge transfer reduces the double bond character of the C = O bond, which lengthens as a result, and the C-N bond in turn takes on double



Kenneth Wiberg with students at Yale.

bond character and decreases in length. Going from ground state to the saddle point conformer, with the 90° rotation about the C-N bond, should eliminate the resonance interaction and lead to an increased C-N bond length and decreased C=O bond length. Wiberg's group found an increased bond length for C-N, 0.08 Å, but only a slight decrease in bond length, 0.01 Å, for C=O. They found also that the electron charge at O changed only slightly and that, as explained above, most of the charge transfer occurred between N and C.

Other Studies. Wiberg's group has investigated several other molecular systems using the CRAY X-MP. Their study of rotational barriers in 2-fluorethanol and 2-chloroethanol showed anomalies in the molecular mechanics values for relative energies between conformers of these systems. Their analyses of the interactions in this system showed that the rotational barrier about the C-O bond was due to Coulombic interactions between the hydroxyl hydrogen and the C-F bond.

Future Directions and the Importance of Supercomputing

These studies require supercomputing power. Wiberg and his colleagues used a MicroVAX in previous studies, and sometimes used many CPU days on a single calculation. With the Pittsburgh Supercomputing Center CRAY X-MP, similar calculations require one to two hours. For convenience, they still do some of the initial searching for ground state on a MicroVAX at Yale. For later stages, they expand the quantum-mechanical basis set using GAUSSIAN and shift their computations to Pittsburgh. Wiberg is especially pleased that the transition from MicroVAX to CRAY X-MP supercomputing flows as smoothly as it does: "It's marvelous that I can take my input file that runs on my MicroVAX, send it here and it runs—unchanged."

These studies suggest that further increases in understanding may follow from similar studies. Known experimental results, even puzzling ones, have been verified and explained. "The phenomena are those commonly encountered in experimental studies," says Wiberg, "but which are not well understood at a fundamental level." His work gives not only a deeper understanding of the physical mechanisms that underly molecular interactions, but also Wiberg believes that the information obtained in his group's rotational barrier investigations, "will be used to reparameterize the molecular mechanics force field. We hope these investigations provide chemists with a much better understanding of these systems, and will help to predict the nature of interactions in other systems."

Wiberg sees supercomputing as playing a central role in this progress: "I think it's a crucial time for determining the intramolecular potentials that go into molecular mechanics, and I think of computational methods as one more experimental tool. Before we used to do simple calculations just to see trends, even though the numbers were often wrong. Now we can reproduce data well enough that we have some hope of getting right numbers as well."

REFERENCES:

*K. B. Wiberg and K. E. Laidig, "Barriers to Rotation Adjacent to Double Bonds. 3. The C-O Barrier in Formic Acid, Methyl Formate, Acetic Acid and Methyl Acetate. The origin of Ester and Amide 'Resonance'," **J. Am. Chem. Soc.** 109 (1986) 5935-43.*

*K. B. Wiberg and M. A. Murcko, "Rotational Barriers. 1. 1,2-Dihaloethanes," **J. Phy. Chem.** 91 (1987) 3616.*

*K. B. Wiberg and M. A. Murcko, "Rotational Barriers. 2. 2-Haloethanols," **Theoretical Chemistry** 163 (1988) 1.*

Molecular Dynamics and the Patient Crystallographer
Protein Structure Refinement by Simulated Annealing

Axel T. Brünger, Yale University

Introduction: The Process of Crystallographic Refinement

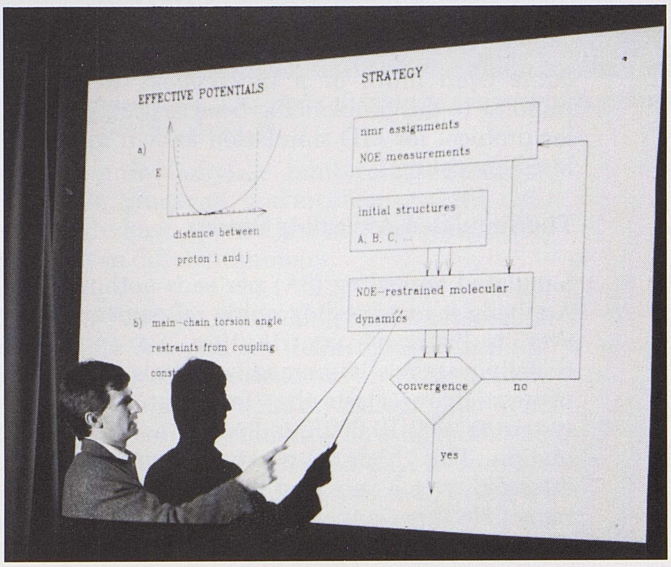
Hours turning to days and weeks, the crystallographer works patiently at his Evans & Sutherland interactive graphic display. He shifts a bond angle slightly, rotates a protein residue. Fine-tuning. Intuition, good guessing, or bad—it's part of the process. He checks the mylar electron density plots he started with six weeks ago.

Maybe this time the model on the screen will fit the X-ray data. No need for another least-squares calculation to move the structure out of one more of what has begun to seem, to the patient crystallographer, like an infinity of energy minima. It's all part of the process. To take the raw data of X-ray crystallography and build an accurate, physical model of a protein is a drawn-out, iterative procedure. For one protein, it can take several months to a year.

Skilled, patient crystallographic refinement has made a substantial contribution to our understanding of proteins. The difficulties of the task have been part of the price of knowledge. For Axel T. Brünger, however, of Yale's Department of Molecular Biophysics and Biochemistry, there has to be a better way. And judging from his work at PSC on a new "automated" refinement process, he's found one. Using the CRAY X-MP/48, Brünger has been developing and testing a molecular dynamics (MD) computational technique which holds promise for eliminating much of the time-costly manual work of conventional crystallographic refinement.

Molecular Dynamics. Simulation has proved itself in recent years as a viable approach for predicting movements over a time range for many macromolecular systems. Refined crystallographic structure data usually provide the starting point for these computations. Theoretically, MD simulations from first principles, beginning with nothing more than the atomic numbers of the atoms involved, would be capable of generating information from which accurate models could be built. At the present state of computing power and potential energy functions, however, such simulations aren't feasible for three-dimensional many-body macromolecular systems like proteins.

Brünger doesn't attempt a first principles approach, rather he employs MD in a particular and unique way as an aid to crystallographic refinement. He incorporates an effective energy term into his potential function, a least-squares expression of the difference between the observed crystallographic factor (F_{obs}) and the factor calculated (F_c) from his model. This term confines the range of the MD simulation to the neighborhood of the X-ray structure.



Axel Brünger at a Pittsburgh Supercomputing Center seminar, April 8, 1988. He's explaining a molecular dynamics strategy for refining protein structures based on nuclear magnetic resonance interproton distances.

Time-Consuming Minimization. Typically, the most time-consuming aspect of refinement has been minimization, calculations to locate minima close to the global minima of the system. Once this energy minimum is located, the atomic coordinates and other structure parameters can be determined with considerable accuracy. The problem has been that restrained least-squares minimization is a gradient descent method; it uses the first derivatives of the squares of the F_{obs}-F_c term. To put it simply, it can take you down to the nearest minimum, which may be global but more often is only local, but it can't take you up and out the other side.

Often the crystallographer can't know if a particular minimum is local or global except by deter-

mining the corresponding electron density map and checking it against the crystallographic data. Once you know you're "trapped" in a local minimum, you can manipulate the calculation artificially, adding energy to "kick yourself out" of the valley you're in, then up and over the next peak, where the gradient descent function will take you down again, to the next, hopefully last, valley of your computational search for the final resting place. (In this regard, a crystallographer once said that the objective is to get there metaphorically before you get there literally.)

Brünger himself has found that least-squares refinement won't correct residues (amino acid structures that attach to the backbone of a protein) misplaced more than about one angstrom ($1 Å = 10^{-10}$m) without intervention to kick it out of local minima. And in recent years, with increased availability of fast diffraction devices and successes at crystallizing larger macromolecular complexes (which makes them accessible to diffraction techniques), the refinement stage has become even more of a bottleneck. Brünger's response to this problem has been to use a particular protocol for MD simulation known as "simulated annealing."

The Simulated-Annealing Process

Simulated annealing (SA) per se is nothing new. Applying it to crystallographic refinement, however, is Brünger's contribution. SA computationally parallels its namesake metals hardening process. Brünger sets the "temperature" for the system to 1000-10,000K and performs an MD simulation. This "high temperature" computation is extended over a specified picosecond (10^{-12} sec.) range. He then "anneals" the system by reducing the temperature slowly.

Before and after these SA simulations, Brünger does a minimization, in the first instance to relieve the system from strain or bad contacts, and in the second to fully optimize the structure. His experience with this technique indicates substantial time-saving in arriving at the global minimum. When it works optimally, the "high-temperature" simulation overcomes local minima. Results such as those included here for aspartate aminotransferase (Figures 1 and 2) suggest that SA-refinement holds real potential for becoming the "automated" refinement technique Brünger would like to see.

Representative Results. Both figures represent examples of the kinds of errors in initial structure that SA-refinement can correct. Figure 1 shows the proline-138 amino acid attached to the backbone or chain of the protein near its center. This structure exhibits what is known as "cis-trans isomerism," which means (1) it has two different conformations with the same chemical formula, an isomer, and (2) the two conformations involve a shift in positions, like musical chairs, between two of the structure's functional groups, molecules or macromolecules that form a bond in the structure. For any given protein to which it is attached, one or the other of the conformations will be in the preferred low-energy state.

Figure 1 indicates that this proline-138 residue was built initially (green lines) in the "trans" conformation, based on reasoned, educated guesses from the crystallography. Conjugate gradient minimization (blue lines), a minimization technique more advanced than restrained least squares, didn't change the conformation. SA-refinement (red lines), however, changed the conformation to "cis," which checked with the manual refitting (yellow lines) which Brünger did

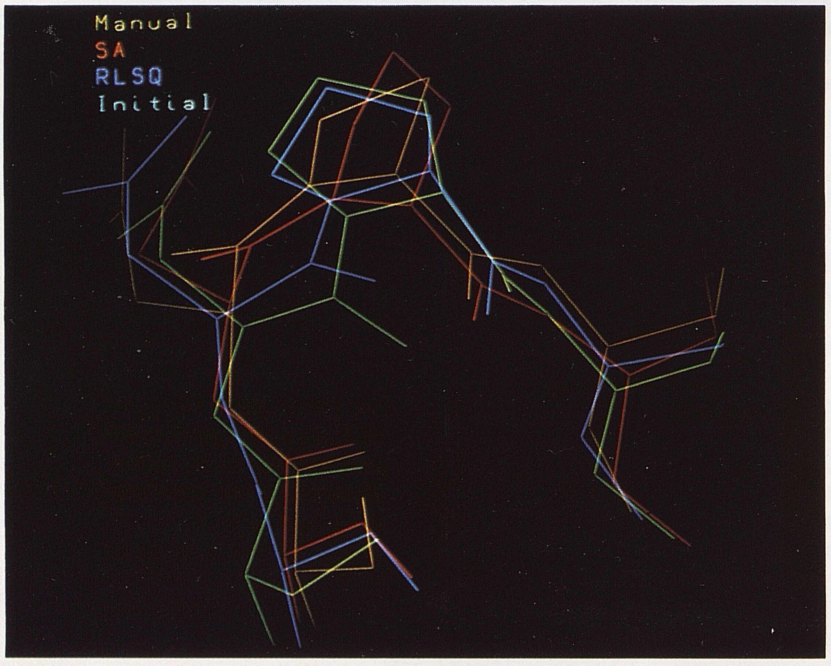


Figure 1: Structural representation of a segment of the protein Aspartate Aminotransferase consisting of residues Ser-136, Asn-137, Pro-138, Ser-139. Superimposed are the manually refined structure (yellow lines), the SA-refined structure (red lines), a conjugate gradient minimized structure (blue lines), and the initial structure (green lines).

to verify the SA computation. This is an error in structure which SA-refinement was able to correct in an essentially automatic manner, where other methods failed.

In the case of Figure 2, the histidine-193 ring, which extends downward to the left from about the center of the aspartate aminotransferase backbone, has undergone a 90 degree rotation during SA-refinement (red lines). This rotation also rearranged some of the backbone atoms. Once again, this rotation is structurally a profound shift from the initial (green) and gradient minimized (blue) structure, and once again the SA-refined structure shows essential agreement with the manually refined version. These structural changes were not possible to attain using conjugate gradient minimization (blue).

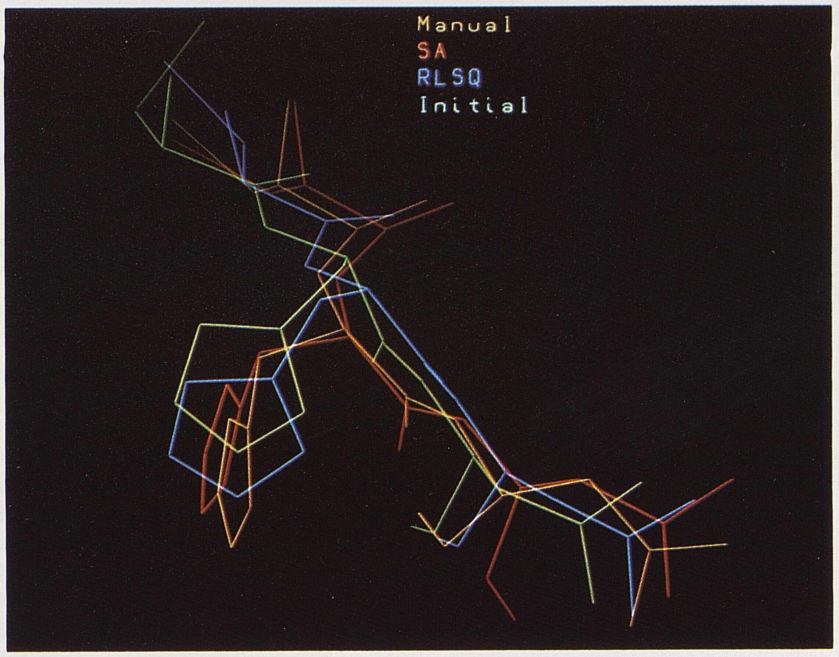


Figure 2: Structural representation of a segment of Aspartate Aminotransferase consisting of residues Cys-192, His-193, Asn-194.

Time-Savings with Supercomputing. For a molecular dynamics simulation involving a protocol not always found in MD work, SA-refinement uses relatively modest amounts of supercomputing time, typically 20 minutes to 10 hours on PSC's CRAY. Compared to the VAX 8800, moreover, turnaround time, which is highly important to the crystallographer, is reduced by days and weeks. The most time-consuming step on the CPU in Brünger's method is the computation for checking a structural model against the crystallographic structure factor amplitudes, which includes calculation of first derivatives with respect to the atomic coordinates. He's gained a speed-up in this process by evaluating the electron density of the model on a finite grid followed by Fast Fourier transformation (FFT), the mathematical transformation which relates the model's electron density to the observed data. Brünger's program X-PLOR, which he evolved from the CHARMM program of Martin Karplus at Harvard, uses a vectorized FFT algorithm. As shown below (in seconds), the availability of FFT routines on the CRAY- X-MP made this algorithm very efficient.

	VAX 8800	Convex-C1	CRAY XMP
Electron Density Calculation	56.0	41.3	2.97
FFT	373.2	21.8	1.67
Accumulation of Data	34.0	8.1	0.64

Future Directions. Brünger's continuing work with SA aims at developing a generally applicable protocol for the process and at fixing optimum levels for temperature, time span, and other parameters according to the type of system under examination. In addition to the aspartate aminotransferase represented here, Brünger has used SA-refinement successfully with crambin, human a-lactalbumin, amylase inhibitor tendamistat, myoglobin, glucagon, and myohemerythrin. Others already have applied the method successfully to about ten different proteins.

Currently Brünger is directing his SA research at cases where conventional refinement is "locked," i.e., when parts of the structure are difficult to interpret. If you imagine the crystallographer at his Evans & Sutherland, thinking how incredibly tedious it used to be, in the old days before SA-refinement, you not only appreciate Brünger's work, but also expect that, with the help of the PSC's powerful research tools, his new research will uncover more and better ways to help the patient crystallographer.

REFERENCES:

*A. T. Brünger, J. Kuriyan & M. Karplus, "Crystallographic R factor Refinement by Molecular Dynamics," **Science** 235 (1987) 458-460.*

*A. T. Brünger, "Crystallographic Refinement by Simulated Annealing: Application to a 2.8 Å Resolution Structure of Aspartate Aminotransferase," **J. Mol. Biol.**, submitted (1988).*

*Martin Karplus, "Molecular dynamics simulations of proteins," **Physics Today** (October 1987) 68-72.*

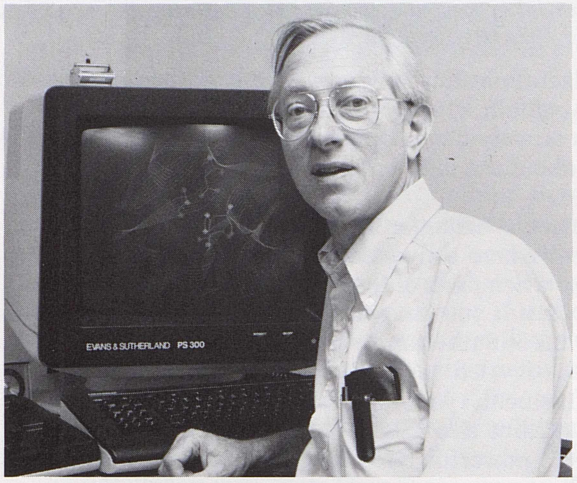
Molecular Dynamics and Drug Design

Free Energy Difference in Enzyme-Inhibitor Interactions

Bogdan Lesyng and Edgar F. Meyer, Texas A. & M. University

Introduction: Enzymes, Inhibitors, and Drugs

One of the characteristic properties of most biological molecules is their ability to interact in a highly specific way with other molecules. Enzymes, for instance, catalyze (speed up) many reactions involving the breakdown of nutrients (substrates). These substrate molecules tend to be much smaller than the enzyme, and a particular enzyme usually acts as a catalyst only for a particular substrate reaction and only within a definite range of temperature and pH.



Ed Meyer in the Texas A & M Biographics Lab. The Evans & Sutherland PS330 displays a representation of the active site of porcine pancreatic elastase.

Enzymes interact also with non-substrate molecules that affect their catalyzing function. A class of these molecules, called inhibitors, locks the enzymes by binding to them, making them unavailable as catalysts, and unlocks them by dissociating. Nearly all drugs exert their effect because they inhibit some enzymatic process. Antibiotics, for instance, inhibit cell processes of the bacteria they attack, while tranquilizers inhibit a process in the brain. Research aimed at understanding enzyme-inhibitor interaction has led to development of other inhibitor drugs. One class of anti-depressant drugs, for instance, inhibits a neural reaction which can decrease alertness and a sense of well-being. Other inhibitor-based drugs suppress specific aspects of the body's immunological response, which may be important during and after transplant surgery or cancer therapy. Well–designed inhibitors provide fairly precise control of the mechanism which the inhibited enzyme catalyzes.

Free Energy and Binding. The binding efficiency of inhibitors depends on the free energy difference between the bound state and the dissociated or unbound state. Free energy difference, a thermodynamic concept, determines in which direction a reaction will occur spontaneously. If the free energy difference between two states is positive the reaction will tend to occur much less frequently than reactions in which the free energy difference between two states is negative. Free energy of zero represents an equilibrium state. A negative free energy difference, therefore, represents the direction that nature goes, like water running downhill, to reach equilibrium.

Binding efficiency is a quantitative measure of how much of the enzyme in a system binds to its inhibitor molecule. Within a given system, the degree to which an inhibitor binds its enzyme is a random, statistical function. They may bump into each other, like strangers in a crowd, and form a bond or they may not. If they bind, it's not permanent; they'll break away after awhile, freeing the enyzme for another bond. The enzyme may be unbound even in the presence of many inhibitors. An enzyme might be compared to a great offensive basketball player who'll break free and score against even the most tenacious defense. But as more defenders guard this player, and he becomes surrounded by "inhibitors," his scoring most likely will decrease.

Molecular Dynamics and Enzyme-Inhibitor Interactions

Professor Edgar F. Meyer and his colleague, Bogdan Lesyng, performed molecular dynamics (MD) simulations on PSC's CRAY X-MP/48 to calculate the relative free energies of binding processes for two similar inhibitors of a given enzyme. Their simulations used AMBER, an MD package of programs developed by Peter Kollman from the University of California at San Francisco. AMBER uses classical mechanics potentials and allows the user to apply free energy perturbational techniques, as Meyer and Lesyng did in this case. The perturbational technique obtains an approximate solution to Schrödinger's equation that applies especially where motions occur over a short distance, such as the fluctuations often characteristic of protein movement. Surrounding water molecules can significantly affect the molecular recognition processes and must be included in the simulations even though they increase the computational task substantially.

Relative free energy, the quantity that Meyer and Lesyng's computations calculate, is a macroscopic thermodynamic property. Many such properties, temperature for one, can be measured easily and accurately. The fundamental properties of individual atoms, however, which allow them to form molecules and compounds having particular macroscopic properties, are not easily determined. Meyer and Lesyng's MD models can be viewed as theoretical statements about how large-scale thermodynamic properties result from the microscopic properties of individual atoms. The accuracy of such MD computations, involving complex protein structures, reflects and measures the accuracy of our knowledge of atomic interactions in complex protein molecules.

Binding Constants and Drug Design. Once the free energy difference between bound and unbound states of the enzyme-inhibitor interaction is determined, binding constant ratios for the two states can be calculated from thermodynamic laws and the binding constant for the inhibitor can be estimated. Binding constants are important in rational drug design because the stronger the binding constant, the smaller the amount of inhibitor required to achieve the same effect. Lower levels of drug reduce the chance of unwanted side effects, many of which result from inhibition of enzymes beyond the therapeutically desired one. If we know which processes will be inhibited, we can predict side effects rather than discover them through long, costly clinical studies on animals and humans.

MD applied to Elastase Enzyme-Inhibitor Interaction. In their project using the Pittsburgh Supercomputing Center's CRAY, Meyer and Lesyng sought to use MD to improve the binding properties of inhibitors for elastase enzymes. Elastases belong to the serine protease family of enzymes and take part in a number of important physiological processes including fertilization, protein turnover, and blood coagulation. The elastases are relatively rigid enzymes that interact with their substrates/inhibitors in a manner analogous to a "lock and key." This provides an attractive model for theoretical investigations.

Meyer and Lesyng chose to work in particular with pancreatic elastase (PPE) and human leucocyte elastase (HLE). Texas A & M University laboratory has determined the structures of a number of PPE complexes to a high resolution, less than two angstrom ($1\,\text{Å} = 10^{-10}$ m). Thus the structural result of MD simulation may be compared with X-ray crystallographic analysis of the parent complex.

Meyer and Lesyng began by analyzing the system consisting of PPE with a trifluoroacethyl-dipeptide-dimethylanilide inhibitor, I (Figure 1) and a cluster of 450 water molecules. This system contains about 3700 atoms. Experimental binding constants are available for I and its modification I′. I′ has an oxygen atom in the place of the N-H group (Figure 1). This I′ "mutation" has a binding constant about 100 times less than I, a marked difference that makes the system a good model for testing MD techniques, for parameterizing the potential energy functions, and for interpreting the inhibition mechanisms.

The "mutation," the substitution of O for N-H, is a relatively slight structural change, and since the binding constant shifts so drastically, it indicates that the N-H group plays a critical role in the binding of I with PPE. Particularly due to the rigid "lock" and "key" structure of PPE, the number of variables that describe the PPE-I binding process would appear to be substantially reduced.

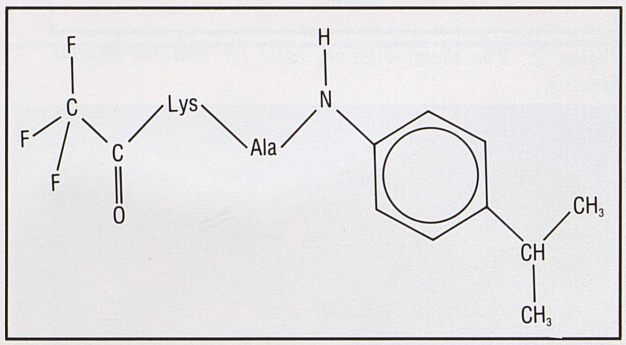


Figure 1: The inhibitor molecule. During the perturbation procedure, the N-H group changes to the oxygen atom.

MD Tractability and the Thermodynamic Cycle. MD simulation of the free energy difference for an inhibitor moving from a free state until it binds to the enzyme is practically impossible to compute for systems of this size. Such binding is a diffusion driven process, which means that the enzymes and inhibitors move on random zig-zag paths in solution, bumping constantly into many molecules. They can't bind until they happen to bump into each other. Diffusion processes occur over much longer distances and times than the vibrational motion usually studied by MD, and it takes many more cycles to simulate diffusion than to simulate vibration, too long for even the fastest of today's supercomputers. Even if tractable, this calculation would represent a statistical "average" of many different movements over a span of time. To compute the free energy difference between two bound states, as in this case, would mean subtracting two such results from each other, each subject to some uncertainty, yielding a doubly unreliable figure.

The thermodynamic cycle perturbational technique offers a tractable, more exact procedure (Figure 2). Instead of simulating "horizontal" processes, ΔA_1 and ΔA_2 as represented in the thermodynamic cycle, Meyer and Lesyng simulate the "vertical" ones, ΔA_3 and ΔA_4, which correspond to changes of I into I′ and the complex E:I into E:I′ rather than diffusion driven binding between the elastase and its inhibitor. For a closed thermodynamic system, such as that represented here, the sum of the free energy differences equals zero, so that $\Delta A_1 + \Delta A_4 = \Delta A_3 + \Delta A_2$. From which it follows that $\Delta A_1 - \Delta A_2 = \Delta A_3 - \Delta A_4$, which gives the desired relative free energy difference.

An MD simulation determines atomic coordinates and velocities as a function of time and is calculated by integrating the equations of motion over the time period necessary to complete the reaction under study. A quasi-stationary "vertical" MD simulation such as this one usually requires 40-50 pico-

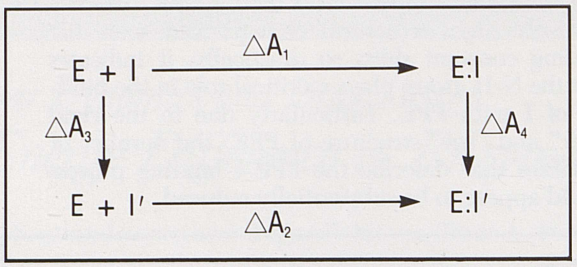
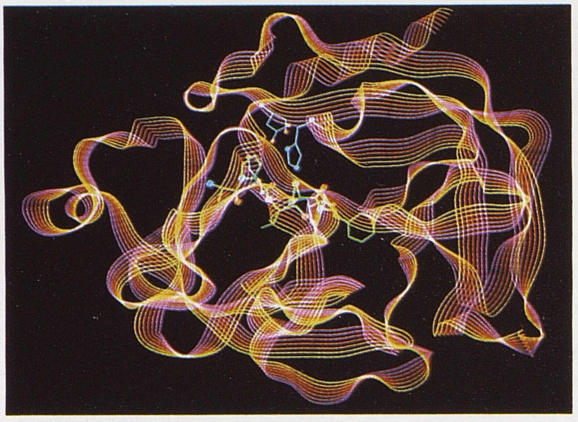
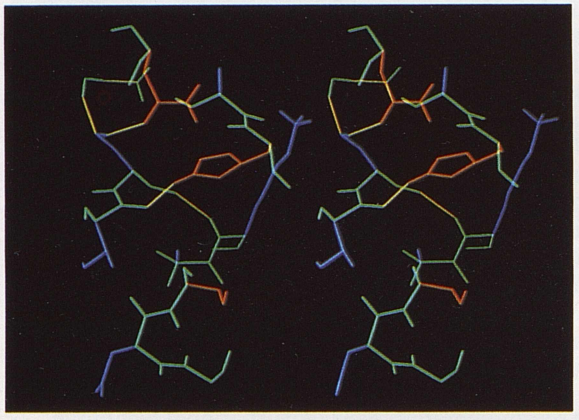


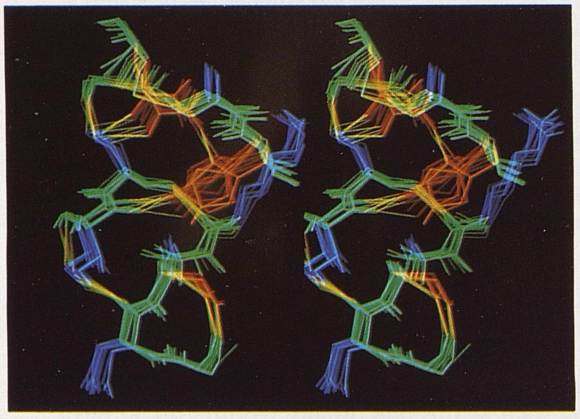
Figure 2: The thermodynamic cycle for inhibitor-enzyme binding.



Polychromic "Ribbon" model of the backbone conformation of PPE with an inhibitor, petidyl-difluoroketone, superimposed.



PPE (stereo view). Key amino acids (Ser 195, His 57, Asp 102, Ser 214) are orange-gold. Adjacent amino acids are lime-green. Peripheral amino acids are blue. Hydrogen bonds are yellow.



Molecular dynamics simulation of PPE (stereo view) for about 1 picosecond duration (10^{-12} sec.)

seconds (10^{-12} sec.) of reaction time in order to reach equilibrium. The corresponding integrals, taken over time steps in femtoseconds (10^{-15} sec.), are highly computationally intensive. In this case, for instance, Meyer and Lesyng calculated 40 ps of dynamics to simulate the E:I to E:I′ transition. Using AMBER, which Meyer and Lesyng adjusted to their purposes, this simulation ran approximately 12 hours on PSC's CRAY X-MP, about an hour of CPU time for each 3 ps of reaction.

The Results: Theory versus Experiment

From their perturbational MD simulation, Meyer and Lesyng derived a theoretical free energy difference of 4.3 kcal/mol. The ratio of the experimental binding constants for I and I′, approximately 100, yields a relative free energy difference of about 2.7 kcal/mol. Meyer and Lesyng's method predicted the right direction for the free energy change, but overestimated de-stabilization binding effects progressing from I to I′. They believe the discrepancy may be explained in part by pH, since experiments have measured binding constants at pH 8, and the simulations correspond to a neutral solution, pH 7. None of the existing MD program libraries treats pH influence on thermodynamic properties of macromolecular systems. Furthermore, Meyer and Lesyng used a "united atom model" approximation of the atom-atom interactions and believe the more exact "all atom model" would further reduce the difference between the experimental and theoretically determined binding constant.

Meyer and Lesyng see two aspects as important in their future MD perturbational investigations: (a) a practical one, aiding development of new inhibitor-based drugs, and (b) a more theoretical one, refining and reworking the method with the aim of greater computational tractability combined with reliable, realistic results. Because of the important clinical aspects of the problem with respect to drug development, Meyer and Lesyng want to concentrate on HLE inhibitors . As for the method itself, they found it to be quite sensitive to the choice of interaction potential, in particular the electrostatic term of the interaction energy. Though the literature occasionally reports close correspondence between experimental and theoretically determined free energies, within a few tenths kcal/mol, Meyer and Lesyng believe such results may be accidental. A precision range of 0.5 - 0.8 kcal/mol for a free energy difference of 3.0 kcal/mol represents probably the best that can be achieved at the moment for macromolecular structures of this size.

REFERENCES:

J. A. McCammon and S. C. Harvey, ***Dynamics of Proteins and Nucleic Acids*** *(London: Cambridge University Press, 1987).*

B. Lesyng and E. F. Meyer, Jr., "Energy minimization and molecular dynamics studies of Asn-102 Elastase," ***Journal of Computer-Aided Molecular Design*** *1 (1987) 211-17.*

Supercomputing and the Healing Arts

Finite Element Modeling of Strain-Induced Bone Remodeling

Richard T. Hart, Tulane University

Orthopaedics and Bone Remodeling

The "sawbones" image of a bone surgeon is almost entirely a thing of the past for orthopaedics, especially with the public attention given to athletes and the glamor of arthroscopic surgery. With less public clamor, orthopaedics also has gained from the tremendous advances made possible by supercomputing. At PSC, Richard T. Hart of Tulane's Department of Biomedical Engineering is attempting to solve a major problem in contemporary orthopaedics, the problem of bone "remodeling."

Bone Remodeling. Bone remodeling, also called "functional adaptation," is a micro version of the self-repairing process that heals broken bones. It refers to bone tissue's ability to respond to strain from external forces. When subjected over time to increased loads, bone cells actually reshape the bone. The process includes a bone surface self-maintenance phase called "resorption" in which microscopically damaged tissue is removed. Then new tissue is overlaid—called "formation"—thickening and strengthening the bone as it hardens. Remodeling also adapts and strengthens the inner structure of the bone. Generally speaking, bone under strain will grow stronger and thicker at points of increased strain.

Though remodeling is a healing process, part of the body's system of adapting mechanisms, it nevertheless poses a major problem for contemporary orthopaedics. An unacceptably high percentage of prosthetic joint replacements fail—36% of knee replacements, for instance, after five to ten years. Many of these failures occur because the anchoring device of the prosthesis, often a steel pin, loosens from the bone. Orthopaedists believe this is due largely to remodeling, which occurs from the changed strains which follow total joint replacement. Various empirical studies and traditional mathematical modeling have made advances, but as yet no mathematical model has been formulated to predict the remodeling response precisely and reliably across a range of situations, a state of affairs which seriously limits the effectiveness of prosthesis design and treatment.

A Three-Dimensional, Self-Adaptive Finite Element Approach

Using PSC's CRAY X-MP/48, Richard Hart has been able to do finite element modeling more sophisticated than other work in the field. His preliminary results show promise of bringing new clarity to some of the muddier issues of bone remodeling. Much of the available research data comes from cross-sectional (two-dimensional) calculations that assume an average load, which remains constant over time. The computing power of the CRAY enables Hart to use a self-developed FORTRAN program (RFEM3D, Remodeling Finite Element Method 3 Dimensional) that operates on three-dimensional bone models. The program uses what Hart calls a remodeling finite element which updates geometric shape and ma-

Richard Hart stands over a monitor which displays a two-dimensional finite element model of the human tibia with the lower portion of a prosthetic knee replacement. The closely packed stress contours indicate where the prosthetic material could fail.

In planned future work at PSC, Hart will extend knee remodeling to include the effects of dynamic loading, loads that vary with time. Recent research has shown that dynamic—as opposed to static—loading significantly affects the bone remodeling response. Ultimately, Hart hopes to develop a three-dimensional model with dynamic loading, a computationally intense project that realistically can be carried out only in a supercomputing environment.

terial stiffness of the model for each iteration. This update is based on strain history and distribution across nodal elements of the model.

In a recent study, Hart used supercomputing runs of the RFEM3D code to compare several remodeling theories with measured results from the front leg bones (radius and ulna) of a sheep. In the empirical study, a researcher used strain gauging to measure and establish equilibrium values in both forelegs. Then a section of ulna was removed from the sheep's right foreleg and strains measured again. With the left foreleg as control, strain was recorded immediately after the removal and at various intervals up to a year, by which time the remodeling process had stabilized.

influenced remodeling results. Somewhat to Hart's surprise, however, the present study, based on empirical data from living bone tissue, suggests otherwise. In each maturation case, including the "no-maturation" model—i.e., newly deposited bone has mature properties—the remodeled geometry and strain distribution were essentially as shown in Figure 2c. Hart speculates that this result may be valid only for remodeling over extended time; in such cases, the effects of different maturation rules may cancel out.

Why Supercomputing? Since reporting these results, Hart also has developed a 4000 node finite element model of a sectioned human jaw bone and performed some preliminary calculations

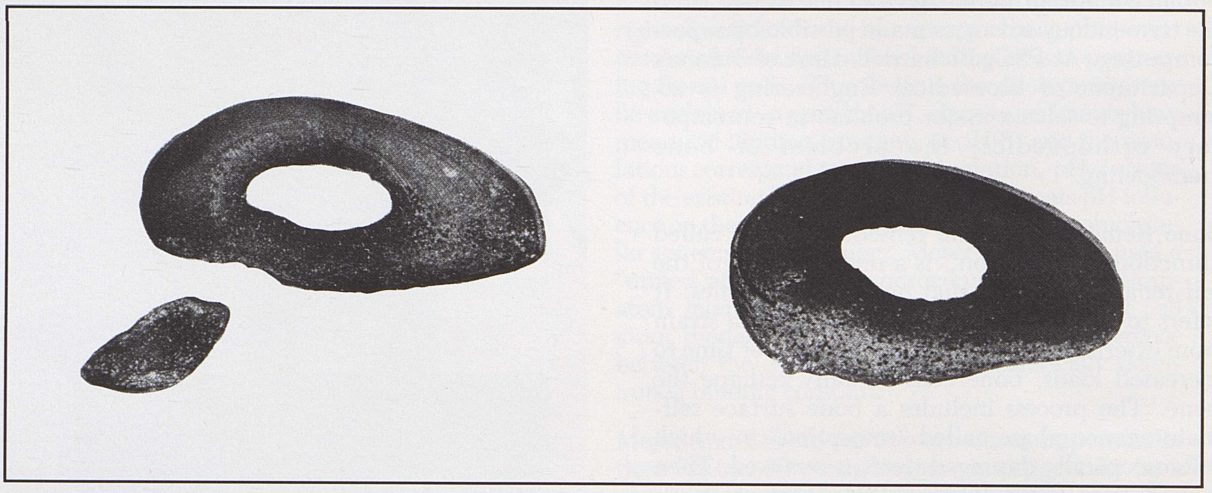


Figure 1. Mature sheep forelegs one year after ulnar ostectomy (removal of a section of the ulna): The left leg, with ulna intact, was used as the control. The radius of the right foreleg shows the effect of remodeling; surface deposits of new bone material resulted in substantial asymmetrical enlargement to the cross-sectional shape.

Maturation Rules. Hart used the PATRAN finite element pre- and post-processor program from PDA Engineering to construct a three-dimensional finite element model of the sheep foreleg bones (see Figure 2). He then made several RFEM3D runs, each with a different "maturation rule," a theoretical description of the stiffening process for newly deposited bone tissue. Until quite recently, reported work assumed that new surface bone material added by remodeling has the same density and stiffness as the old bone material which it overlays. Hart recognizes, however, that the new material is likely to be relatively compliant, stiffening over time as it calcifies. He has added this increased complexity to his modeling. Another maturation hypothesis—among several that Hart tested—assumes that in the initial response to strain, remodeling deposits excess material and that this excess is resorbed over time to arrive at a stable bone geometry and stiffness.

Results. In a prior study, based on constant static strain in an idealized tubular model, Hart and colleagues found that different maturation rules

aimed toward modeling bone deposition in the socket formed by tooth loss. Hart has also mapped plans toward a finite element study of internal bone remodeling and a dynamic load analysis of the human knee. Hart's impressive work with complex finite element analysis demands computing accuracy and speed available only in a supercomputing environment. Each run of the sheep bone study reported here, for instance, used 26 minutes of CPU time on PSC's CRAY X-MP/48, but would have required more than eleven days on MicroVAX. "Access to the PSC CRAY," says Hart, "makes it possible to solve biomechanics problems that simply cannot be solved otherwise."

REFERENCES:

Hart, R. T., "Remodeling finite element calculations of strain-induced in vivo bone remodeling," 34th ORS, February, 1988b, Atlanta, Ga.

Hart, R. T., Chapters 4, 11 & 12, ***Bone Mechanics****, ed. S. C. Cowin (CRC Press, 1988).*

Lanyon, L. E. and Rubin, C. T., "Static v. Dynamic Loads as an Influence on Bone Remodeling," ***J. of Biomechanics****, 14 (1984) 897-905.*

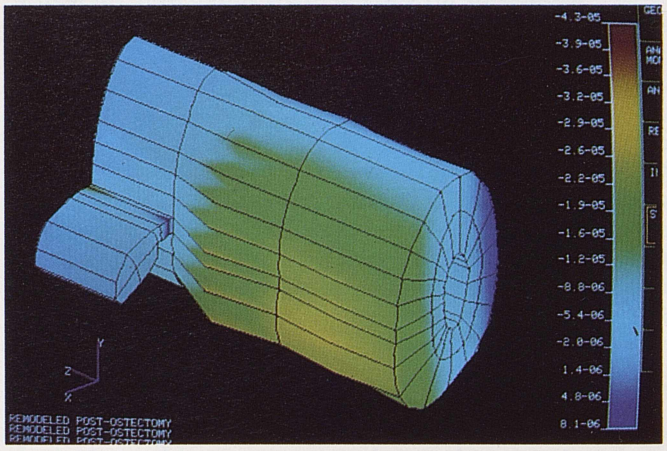
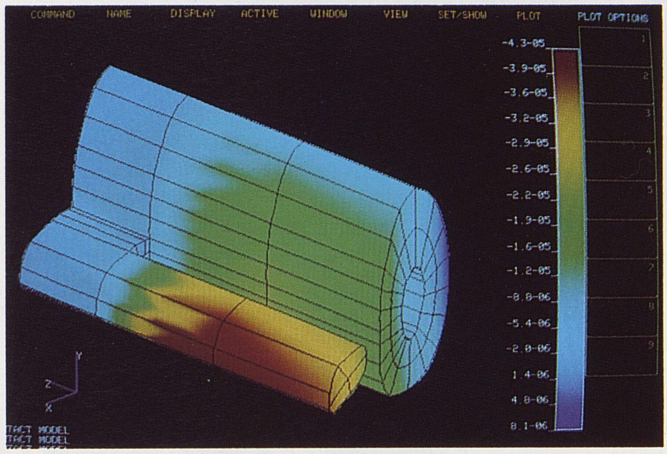
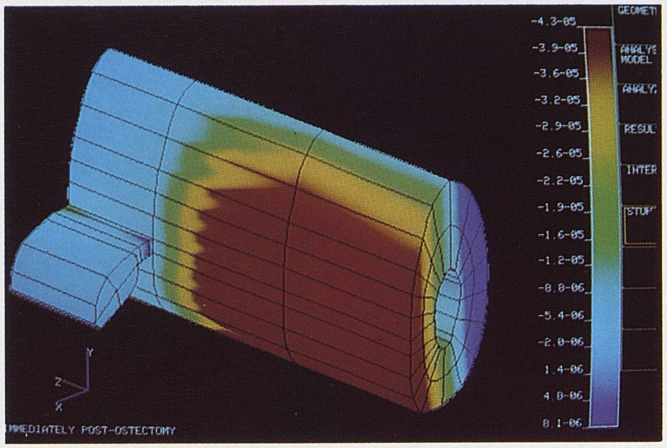


Figure 2a. Richard Hart's three-dimensional finite element model of a mature sheep foreleg. The model assumed a rigid cap below the cross-sectional point of ostectomy; strain contours are based on axial loading at that point. The contour lines represent gradations of strain. In this case, the leg is shown intact, and the strains represent the equilibrium strains.



2b. Sheep foreleg with a section of the ulna removed. The contour lines represent gradations of strain immediately after the ostectomy.



2c. The contour lines represent gradations of strain and correspond to the remodeled cross-section a year after ulnar ostectomy, as shown in Figure 1.

Supercomputing and Wall Street

Dimensionality of the Covariance Matrix for Stock Returns

Keith T. Poole and Scott F. Richard, Carnegie Mellon University

"The question we're trying to answer is how many risks are there underlying the market? If the structure underlying the changes in value of securities is one-dimensional, it might be possible to get rid of all unrewarded risks."

According to Scott Richard, it's not as far-fetched as it may sound. A professor of finance and economics at Carnegie-Mellon University's Graduate School of Industrial Administration, Richard is currently on leave, putting his theory into practice as vice-president for mortgage securities research at Goldman, Sachs and Company, a Wall Street investment firm: "If we can find quantitative measures of risks, then we can apply those measures and guide investment managers in choosing portfolios."

Keith Poole, a political economist and chair of the doctoral program at G. S. I. A., has worked extensively with advanced methods of statistical scaling and manipulation of data, experience he's needed in this project. "Our data matrix consists of daily returns on the New York Stock Exchange since 1962, 6157 trading days on 5019 stocks. That matrix in itself is huge, and as a data problem, this research is extremely difficult." With a massive data set requiring complex matrix maneuvering to arrive at a solution, Poole and Richard turned to the Pittsburgh Supercomputing Center's CRAY. "This study is possible only on a computer of the capacity of the CRAY," says Poole.

Theory: Stylized Facts and Two Kinds of Risk

Richard and Poole describe their project as aiming to generate a set of "stylized facts" which will structure the dynamics of securities' returns—the ratio of tomorrow's price plus dividend, if any, divided by today's price. "Stylized facts," says Poole, "is a term we started using at G.S.I.A. years ago, and it spread through economics and political economics. It's what natural science calls 'a fact.' An example from the American political process is that after the convention, nominees tend to moderate their positions." Why "stylized"? Because a statistical fact is "stylized" by probability as an inherent part of the statement, just as "tend to" makes Poole's statement about nominees essentially incontrovertible. Poole and Richard want to identify the factors, including their associated probabilities, that will give predictability to stock market returns.

Richard, who specializes in linear-algebra-based analysis and modeling, the sort of work sometimes referred to as econometrics, says that in recent years, theoretically-derived, predictive modeling of market phenomena has had a vast impact on investment practice. "It's revolutionized Wall Street." A model that predicts optimum buy-sell decisions for stock options is reliable in the high 90% range, and its areas of non-reliability are very well understood.

Scott F. Richard

The basic assumption, he says, referring to physics, is that stock prices move like Brownian particles. Though they exhibit a Brownian "perpetual dance" of randomness, it's this behavior in fact which validates the models. Their predictive capacity depends on the basic unpredictability of the market. This notion comes from "the efficient markets hypothesis," which says that competitive markets do not permit profitable opportunities to remain unexploited. "People keep changing the rules of the game," says Richard. "God is more reliable."

The Capital Asset Pricing Model. A model which has gained considerable authority in financial economics, the capital asset pricing model (CAPM), distinguishes between two kinds of risks; one—which follows the movements of the market as a whole—is called "systematic risk." "There are un-

necessary risks, which are diversifiable," says Richard, "but there are also 'systematic' risks, which are not. What we're looking for ultimately is how many kinds of systematic risks there are. We want to determine how many flavors rewarded risk comes in."

Rewarded and Unrewarded Risk. "It's the basis of much financial theory," explains Rob Byrne, assistant dean of the business school. "This systematic factor is usually represented as β, and there are low β stocks like utilities and high β stocks like bio-engineering firms, which may fluctuate drastically but tend nevertheless to fluctuate with the market." These risks can't be diversified away; they're inherent in the market. "Unsystematic risk" is the term for the risks inherent in owning a particular stock. This risk increases to the extent an investor fails to diversify. With one stock, you have the greatest chance for a large return, but you also have the risk of not diversifying, a risk for which there is no return. "No one pays you for unsystematic risk," says Byrne. For systematic risk, the assumption is you're rewarded by the basic structure of the market, which on a macroscopic scale moves always toward profitability.

Recent Models of Market Structure

Financial economists have developed several models recently, each of which explain stock returns as a function of these two kinds of risk. The models suggest that the cumulative effect of a small number of "common shocks" determines the value of returns. These common shocks, "common" because they affect more than one firm—though they may affect different firms in entirely different ways—are modeled as "factors." Both "common shocks" and "factors" are systematic. Returns for any particular firm depend on an unspecified number of factors—e.g., interest rates, the price of oil, value of the dollar. They also depend on what the theorists call "noise" or "random noise." Noise is unsystematic and idiosyncratic to a particular stock—the accidental death of a key executive is one example. By definition, it can't be explained in relation to a factor.

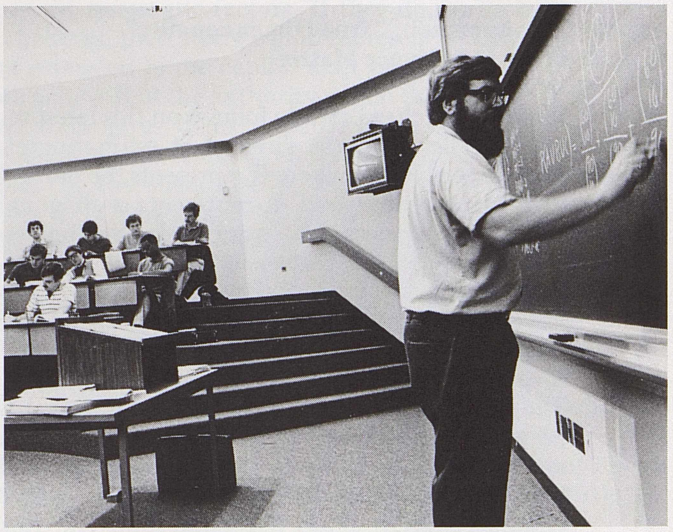
"In terms of algebra," says Poole, "this boils down to the classic factor analysis model":

$$X = BF + E$$

where X = the 5019×6157 matrix of returns
B = the $5019 \times k$ matrix of factor loadings, the matrix of βs
where k = the number of factors (βs).
F = a $k \times 6157$ matrix of common daily shocks (factors)
and E = a matrix of idiosyncratic noise.

Applying this generalized statement to Poole and Richard's project: the known quantity, market returns, is expressed as a function of unknown parameters, the factors, and a variable that accounts for noise.

Dimensionality. At issue is the "dimensionality" of the common shocks, which is a statistical way of saying you want to know how many systematic factors there are and the relative importance among them. What factors, βs, will express a relationship between today's price and tomorrow's in more than a random number of cases? Does one factor apply for every firm? Or do some factors apply only some of the time? Which ones will predict with the greatest accuracy, i.e., yield the least variance?



Keith Poole teaching an interesting problem in his MBA class in Probability Theory. Poole also teaches a course, using the Pittsburgh Supercomputer Center's CRAY X-MP/48, introducing graduate business students to supercomputer applications in financial economics.

The Covariance Matrix. The models of recent interest have proposed that this dimensionality can be determined by computing the eigenvalues of the covariance matrix. Eigenvalues characterize the size of a matrix. The covariance matrix represents a matrix statement of the "variances," the quantity by which a particular factor did or did not give the expected result when applied to the data in question. In this case, for any one firm, assuming it traded 6000 days, each factor would yield 6000 predictions of the next day's return. If it predicts correctly, the variance for that day is zero. For a wrong prediction, the variance is a numerical relation between the expected result and the prediction. Poole and Richard represent the covariance matrix of returns as R.

One of the recent models theorized that the dimensionality of systematic risks would equal the number of "unbounded" eigenvalues of R. "Unbounded" eigenvalues increase in magnitude with increasing size of R as more and more firms are considered. "What emerged from this empirical research," says Poole, "was quite remarkable—

only one unbounded eigenvalue was found." This meant that only one systematic factor affected market prices.

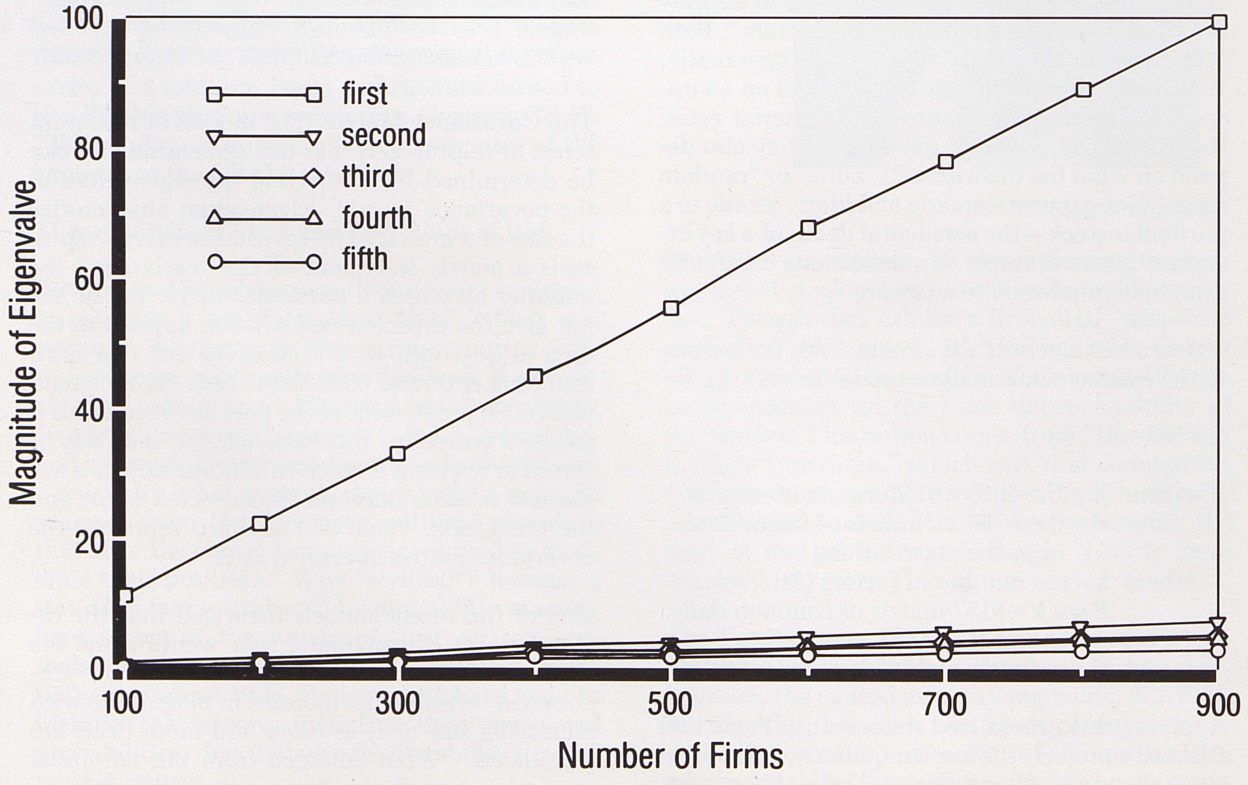
A subsequent theoretical critique of the earlier result, however, found that eigenvalues of R would not reveal the true dimensionality of the common shocks. This study, explains Poole, "showed that no matter what value of k resulted from the data, only one unbounded eigenvalue would be observed. It's not that the prior study was wrong, it's that ordinary principal components—that is simple eigenvalue-eigenvector decomposition—would not reveal the true dimensionality."

The Approach: "True Dimensionality" of the Covariance Matrix

Poole and Richard's study follows on this latest research. Their effort is directed at estimating eigenvalues of R - Φ, where R represents the normal covariance matrix and Φ represents a diagonal matrix of variances. The prior study, by Stephen Brown of New York University, proposed this correction as a way of overcoming the problem of observing only one unbounded eigenvalue and of arriving at the "true dimensionality." This revised covariance matrix is called the correlation matrix. The main problem Poole faced, however, is that R - Φ is "notoriously difficult to estimate. You have a matrix which is the sum of two matrices, and there are an infinite number of ways to add 2 numbers to arrive at any particular sum. The trick is how to partition. There is no agreement on the right way to do this."

Data Operations on the CRAY. To create the covariance matrix for his roughly 6000 × 5000 data set and to calculate eigenvalues, Poole had to overcome several difficult problems in data manipulation. "Not every security is traded every day, so the matrix has holes, a little like Swiss cheese. But you have to form a symmetric matrix. Suppose you can find 1000 firms that all trade on the same days. To do an eigenvalue on that, you have to form a 1000 × 1000 matrix. Normally, you would read a 6157 × 1000 matrix, and set up two 1000 × 1000 matrices for "scratch." But then you're already over the memory of the CRAY."

Poole's first maneuver was to transpose his giant matrix as he read it, letting rows = days and columns = firms instead of the other way around, as required by the B term of the factor analysis. Solution of the covariance matrix requires a transposed B term, designated as B′. By reading his data set as he transposed it, Poole was able to form the covariance matrix on first read. "To transpose a matrix of this size took three tries, and each try ran roughly three CPU hours." Poole used the SSD, high-speed disk solid-state storage device, to accomplish his matrix transposition. After reading the full data set into SSD, he read it out 30 times into CRAY X-MP memory. "I read it once to pick up the first 200 days, then wrote it off the disk and rewound to pick up days 201 through 400 . . . until I got through 6000 days."



Once his transposed matrix was in the CRAY, Poole had to contend with difficult, technical problems of matrix calculation. "The classic factor model assumes certain standards. It's very important for the validity of the result that the number of matching terms be very large. I formed the biggest matrix I could with the most complete data, taking an overlap of 1800 trading days to help get around the fact that firms have come and gone since 1962. I ended up with six matrices. There must be numbers in each slot of the matrix. If there are missing entries, you can get negative eigenvalues. I used 'pairwise deletion' to get rid of incomplete data. If there wasn't a match between firm A and firm B on a particular trading day, I didn't use that day."

Results: Market Movement and Underlying Dimensionality

The chart represents the results of Poole and Richard's calculations. For increasing size of each matrix, from 100 through 900 firms, the graph charts the first 5 eigenvalues. The first eigenvalue is clearly prominent, increasing in magnitude in an essentially linear relationship with the number of firms, unbounded through the quantity of data considered. "It's fundamentally one dimensional. The interesting thing is we did the calculation using the identity matrix term to correct the covariance, as Brown's study recommended, and we get the same result," explains Poole. "With this result, you can think of the distribution of data points for returns as having the shape of an ellipse with the first eigenvalue representing the long axis. The other dimension is small in proportion. This is a very long, narrow ellipse."

Does this mean that stock market returns depend on only one factor, and if so what is it? "We have found one overwhelming source of systematic risk," says Richard. "I have no doubt what it is. It's market wide movements or 'market movement,' as it's usually called. The tendency of the market to move together. It's what's involved in what's called indexation; this is what they do in pension fund management, for instance, where they'll tie buy-sell decisions to something like the Standard and Poor's index.

"The question now is this: after we remove the large risk, how many smaller risks are left? It's a little like having an elephant in your living room. It's easy to see the elephant, but there may be cats and dogs that you don't see. You have to take the elephant out before you can see the cats and dogs. That's where the real science begins." Poole and Richard both stress that the 4 secondary eigenvalues graphed on their chart, though more difficult to identify, may be very significant. The chart indicates that each of them, like the first value though much less sharply, expands in linear relationship with the number of firms. Preliminarily, this would tend to indicate that they also represent systematic risk and not random noise.

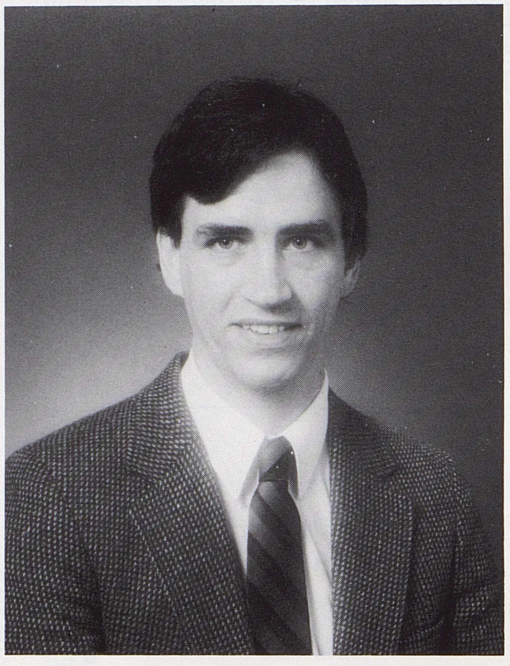
"There's a lot of debate about the underlying dimensionality," says Poole. "If you're a large investment firm and have tens of millions of dollars involved, these other four factors could be very important. Each eigenvalue represents a single process in the world. Eventually we'll be able to identify discrete phenomena with each of them."

The Role of Supercomputing and Future Work. For the next stage of their work, Poole and Richard will run Monte Carlo studies to determine definitely whether the four underlying dimensions represent systematic factors. If they do, as Richard believes, their study would represent a significant extension to the capital asset pricing model. Both believe that the authority which the extent of their data gives them should resolve questions raised by earlier work. "What's important," says Poole, "is the sheer size of the project. That's where the CRAY comes in. It's like having a microscope 10 times the power of an old one. You can't settle questions of structure and dimensionality unless you go up into the thousands in your dataset. It's basic empirical work to settle a theoretical question. This is research aimed at producing results that will last for decades."

Microtasking for Better Supercomputing
"Polite" Multitasking and Improved System Throughput

Michael Bieterman, Boeing Computer Services

The results are in: CRAY microtasking boosts supercomputing batch throughput. That means more return per dollar of supercomputing investment, more research results in the same wall-clock time, more science, more knowledge. These are the broad implications of a report from Dr. Michael Bieterman's National Science Foundation study conducted at PSC, "The Impact of Microtasked Applications in a Multiprogramming Environment."



Michael Bieterman

Multitasking and Microtasking

"Microtasking" is CRAY Research's term for its recently developed "small granular" form of multitasking. Like traditional multitasking, microtasking is a software feature which enables a program to be multiprocessed. For instance, on a system such as PSC's CRAY X-MP/48, which has four parallel processing units, a multitasked program can run on all four units at once, thereby finishing its work four times faster than a normal, one-processor-dedicated run. Think of going to the grocery store with your spouse: "You do eggs, milk, butter and cheese while I get the soap and Tender Vittles" is a domestic form of multitasking.

As you'd expect, 100% efficiency can't be realized. That's because multitasking code itself involves an increase in execution time, referred to as "overhead," time spent discussing who gets the milk, who gets the catfood. Efficiency is also hampered because traditional multitasking depends on the programmer's ability to organize and manipulate the program code, on the subroutine level, in such a way that large segments can run independently and simultaneously. Thus, so-called "critical" segments which pass updated data from one subroutine to another—"remember, don't pick out an avocado unless you let ME see it first!"—are seldom multitaskable; and perfect "synchronization" of segments running independently, which 100% multitasking would require, is impossible as a practical matter. CRAY says the actual upper limit on run-time improvement attainable with multitasking for the X-MP/48 is a factor of about 3.8 rather than 4.0, which would represent 100% efficient multitasking.

Microtasking differs from traditional multitasking in that it operates on small, discrete chunks of code—FORTRAN Do Loops being the most common example. The "granularity" of the task is smaller, thus the name, microtasking. CRAY analyst Lauren Radner, former PSC site analyst, gives the example of code for filling out an array in a mathematical application, "You just want it crammed with numbers," she says, "and you don't give a hoot which slot in that matrix gets filled when." This notion, not giving a hoot about when, describes CRAY's term for a microtaskable subroutine: a "fray," i.e, "willingness to let activity in a subroutine be this disordered."

Easy to Use and Understand. Belying the scientific aura of its name, microtasking—at least from the user point of view—is simple to understand and apply. Radner invokes the Peggy Lee hit song from a few years back, "Is that all there is?" "That's how you feel," she says, "when it's explained to you." Unlike "macrotasking," CRAY's new term for traditional multitasking, microtasking doesn't require precise code analysis and management. "It's a snap," says Radner. The programmer inserts specified directives at easily identifiable "frays" in the existing code. During compilation, a preprocessor program converts user FORTRAN into a substructure which the user never has to know about and never sees; during execution, this substructure does the necessary allocating among CPUs.

Voila! Simple to use, with faster results almost guaranteed; it's an ideal technique in many matrix-laden computation-dominant mathematical applications. Just a matter of identifying the appropriate Do Loops and sticking in some directive code. Surely all regular and most irregular X-MP/48 scientific users know about, use and love microtasking, right? . . . . Wrong.

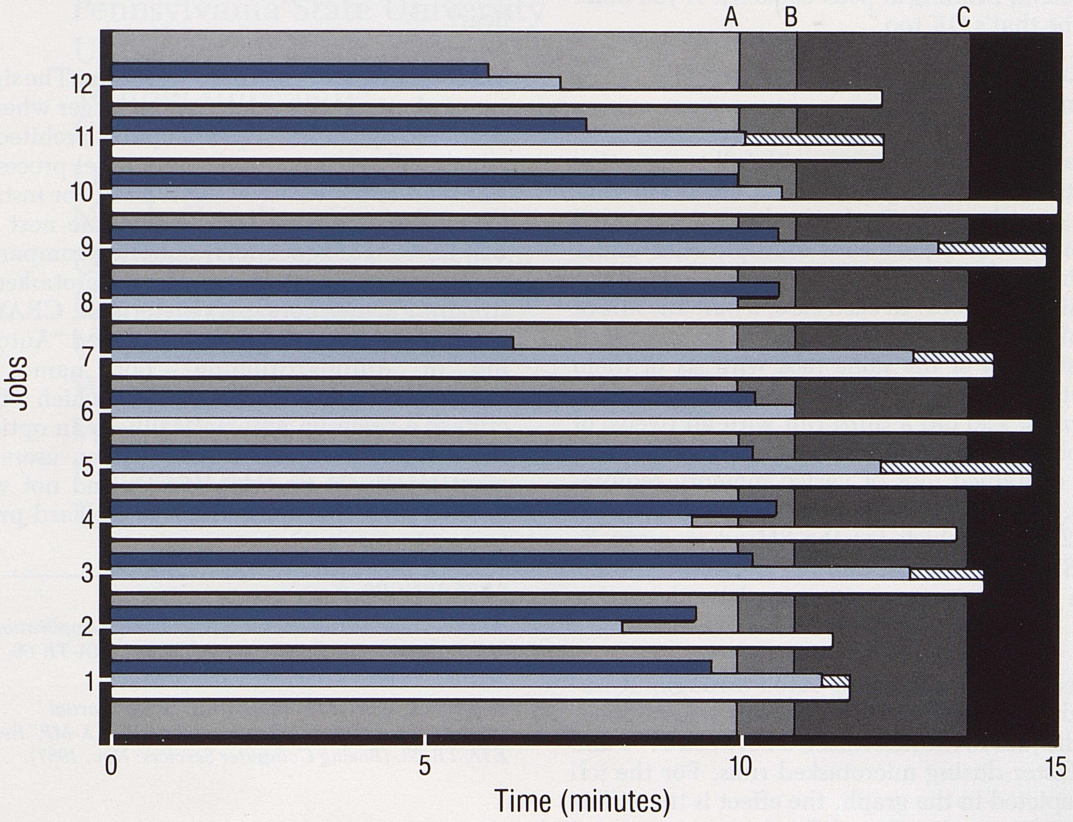
Bieterman's recently reported results, however, suggest strongly that microtasking will not for long sit quietly on the fringe of supercomputing. "No one had tried it," says Bieterman, a numerical analyst in the Engineering and Science Services Division of Boeing Computer Services, explaining the germination of his project. So I thought I'd do an experiment." At issue was the effect that microtasking could have on batch throughput by using otherwise idle CPU cycles.

Microtasking and Batch Processing

Microtasking "has the potential to increase total system throughput in a batch production environment by using otherwise idle CPU cycles," states CRAY's X-MP Multitasking Programmer's Reference Manual. Bieterman knew the manual statement and became interested in exploring this aspect of microtasking's potential. "I assumed there might be results somewhere, but no articles said here is the empirical data. This motivated me to go ahead."

Vectorization and Microtasking. There is no reason to doubt that microtasking can improve wall-clock run-time for individual programs, but in a sense this is not particularly thrilling news to CRAY users. As explained, these improvements are limited on the upper end by a factor of 3.8. The main factor improving run-time on the X-MP/48 has been "vectorizing," the application of programming techniques which optimize code and enable the program to use CRAY's architecture to best advantage. Improvements made possible by microtasking can seem trivial in comparison.

But vectorizing and microtasking are not mutually exclusive. CRAY's developing philosophy is that CRAY software does vectorizing for the programmer, and microtasking is then an option available. "To do microtasking without vectorizing first," says Radner, "is like polishing the handle of the passenger-side door instead of waxing the car." Microtasking, furthermore, appears to offer advantages for total system operation that don't apply to other run-time speedup techniques.



A base run (white) with no jobs microtasked; a second run (blue tint) in which the even-numbered jobs, half the mix, were microtasked; and a third run (full blue) with all 12 jobs microtasked. A, B, and C indicate when half the mix—six jobs—finished for the all microtasked run (A), the half-microtasked run (B), and the all unitasked run (C). The cross-hatched blue tint sections indicate speedup of non-microtasked jobs in a microtasked environment.

Idle Time and "Polite" Multitasking. Bieterman notes that processors operating in a batch environment can be idle under a number of circumstances, including:

- system load is temporarily low;
- less than N jobs simultaneously fit in memory on an N-processor system (i.e., memory is "oversubscribed");
- resident jobs use I/O, but not CPU, for some time;
- I/O occurs in rolling jobs in and out of memory;
- jobs wait for a critical I/O resource (e.g., solid-state disk storage).

Microtasking offers unique advantages for occupying these idle moments. It has relatively low "overhead," i.e., the CPU allocation process takes relatively little CPU time in its own right; and rather than demanding CPU time from idle processors, a microtasked program requests it. These good software manners may be likened to the difference between acting selfishly and acting with the best interest of a community in mind. "Microtasked jobs in memory," writes Bieterman, "get more than one processor only when no processor-less job in memory is waiting for CPU." Like a good utility infielder a microtasked program says, "I'll fill in wherever I'm needed to keep this team, this system, running at peak capacity. If you don't need me that's OK too."

The Experiment: Mixed and Unmixed Job Streams

To check it out, Bieterman—with full cooperation from the PSC—set up controlled batch environment experiments. Each involved a prepared stream of twelve jobs with multiple runs under differing mixes of microtasked versus non-micro or "unitasked" jobs. In each case, a baseline mix of all unitasked jobs was run and then compared against a run of the same jobs with six of them microtasked, six unitasked. For one job mix, Bieterman also did a third run with all twelve of the jobs microtasked (see graph). Each of the mixes contained jobs of varied memory requirements, with some small jobs and several larger than 2Mwds, which for the 8Mwd memory X-MP/48 creates a condition of "oversubscription" which is quite common in normal batch runs at PSC.

Results. In every instance, batch throughput improved when half the job mix was microtasked. Half the jobs in the mix finished between 17% and 38% faster during microtasked runs. For the job mix depicted in the graph, the effect is the difference between points B and C, an improvement of 20%. When the entire batch was microtasked, half the jobs finished at point A, an improvement of 26% over the same batch running with no microtasking. For total throughput of all twelve jobs, Bieterman's experiments showed an improvement with microtasking that ranged between 12% and 24%. The reduced improvement for total throughput reflects that fewer or no microtasked jobs remained in the mix over the latter half of run-time, decreasing the job stream's ability to use CPU idle time. This is also indicated in the graph by the difference between point A and point C compared to the difference between points B and C. Bieterman monitored recovery of CPU idle time, the results of which indicate that microtasked jobs have no more, and sometimes less, CPU idle time attributable to them when running in a batch environment than in a fully dedicated run.

Bieterman's empirical results, the first available on microtasking, show that it can improve overall system usage by a significant factor. A partially microtasked jobstream will give faster throughput overall and will speed up the non-microtasked jobs. This interesting result, that non-microtasked jobs gain a wall-clock speedup just by being in a partially microtasked batch—a free ride in effect—is indicated by the cross-hatched blue-tint sections of the graph. These results indicate that in a scientific supercomputing environment, the use of microtasking means faster research and greater return in science results per dollar of public investment—a point that PSC has taken to heart.

Microtasking and CRAY Architecture. The significance of microtasking looms even larger when one considers upcoming supercomputer architecture, which will rely even more on parallel processing. The soon to be released CRAY Y-MP, for instance, which PSC plans to install sometime next year, will have eight processors, doubling compared to the X-MP the advantages of a microtasked job stream. Furthermore, says Bieterman, CRAY will soon publish a compiler feature called "Autotasking" or "Autopartitioning"—both names have been used by CRAY spokespeople—which will microtask a program automatically, as an option of the compiler. At that juncture, even users who want merely to do their science and not worry about computing subtleties will be hard-pressed to avoid microtasking.

REFERENCES:

Bieterman, M., "The Impact of Microtasked Applications in a Multiprogramming Environment," Report ETA-TR-69, (Boeing Computer Services: Dec. 1987).

Bieterman, "Microtasking General Purpose Partial Differential Equation Software on the CRAY X-MP," Report ETA-TR-68, (Boeing Computer Services: Nov. 1987).

The Academic Affiliates of the Pittsburgh Supercomputing Center

Carnegie Mellon University
Case Western Reserve University
Drexel University
University of Florida
Georgia Institute of Technology
Johns Hopkins University
Lehigh University
University of Maryland
University of Michigan
Michigan State University
Northwestern University
Ohio State University
University of Pennsylvania
Pennsylvania State University
University of Pittsburgh
Temple University
University of Rochester
Duke University
North Carolina State
University of North Carolina
University of Tennessee
University of Virginia
Washington University
West Virginia University
University of Wisconsin
Yale University

Photographs:

Melissa Farlow: black and white photographs, pages i, iii, iv, v, viii
Joe Renckly: color photo of Pittsburgh, page 10
Mark Martin: pages vii, 29
Larry Rippel: color photo of CRAY, page ix

Editor/Writer: *Mike Schneider*
Production: *Vivian Benton*
Design: *Jeff Lunderstadt, Mellon Institute Graphics*
Dave Kent & Sharon Krinock, Chas. M. Henry Printing, Greensburg, PA

Thanks:

Science Consulting: *Martin Head-Gordon and Teresa Head-Gordon of the CMU Department of Chemistry. Also Hugh Nicholas, Doug Fox and others of the PSC staff.*

Copy and Editing: *Sally Flecker, University of Pittsburgh, Department of English; Tracy Muffett, University of Pittsburgh, PSC work-study; Hugh Nicholas, PSC; Cindy Gill, Health Publications, University of Pittsburgh; and to RR also — thanks.*

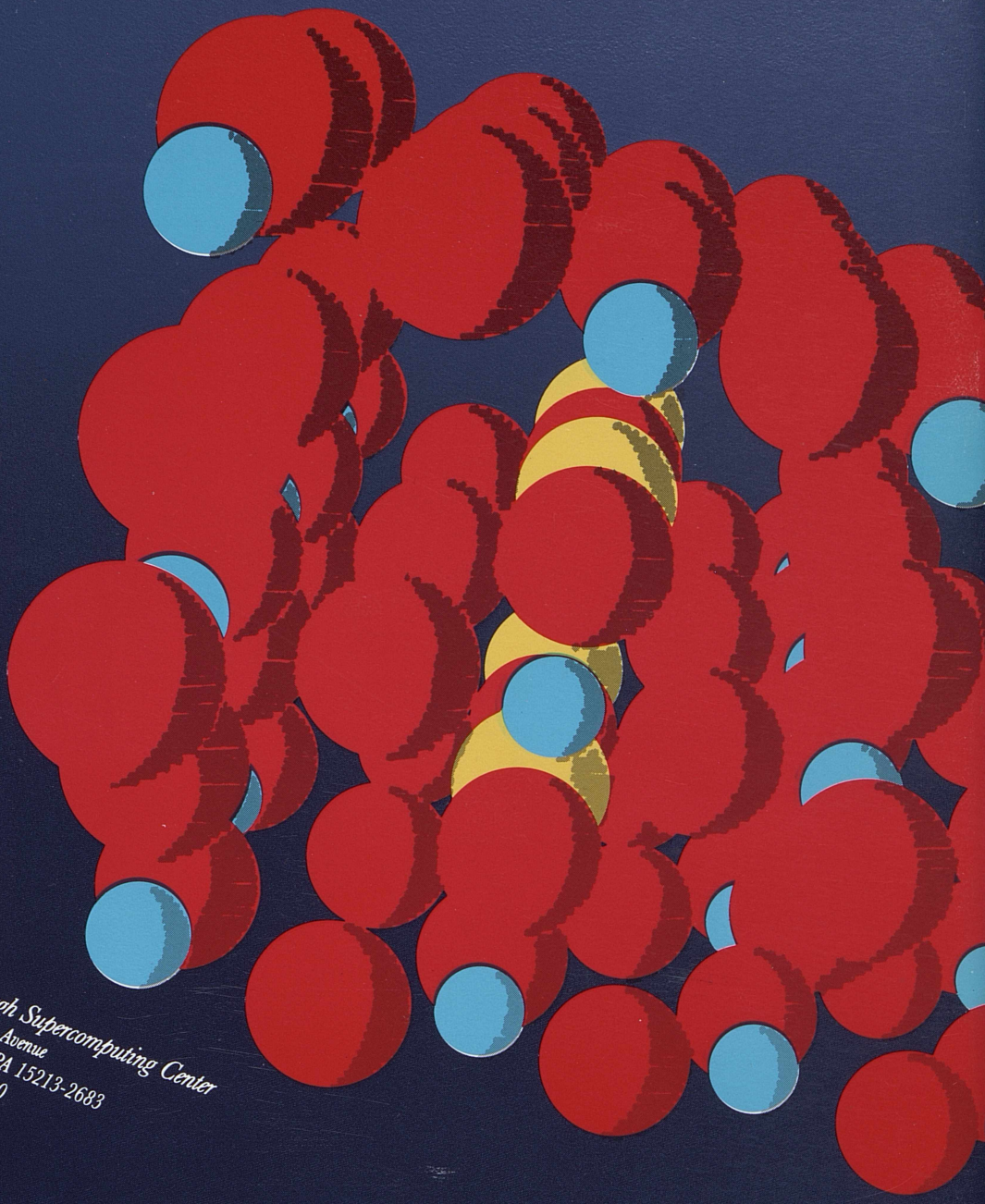
Photography on short-notice: *Mark Martin and the Mellon Institute Graphics staff.*

Research: *Lynn Labun, Kathy Bossick and the Mellon Institute Library staff.*

Special thanks to the contributing reseachers for their cooperation and assistance.

The PSC gratefully acknowledges support from:

National Science Foundation
Commonwealth of Pennsylvania
National Institutes of Health
Cray Research, Inc.
Digital Equipment Corporation
Western PA Advanced Technology Center through the Ben Franklin Partnership Fund



Pittsburgh Supercomputing Center
4400 Fifth Avenue
Pittsburgh, PA 15213-2683
412-268-4960