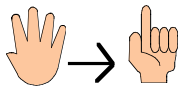
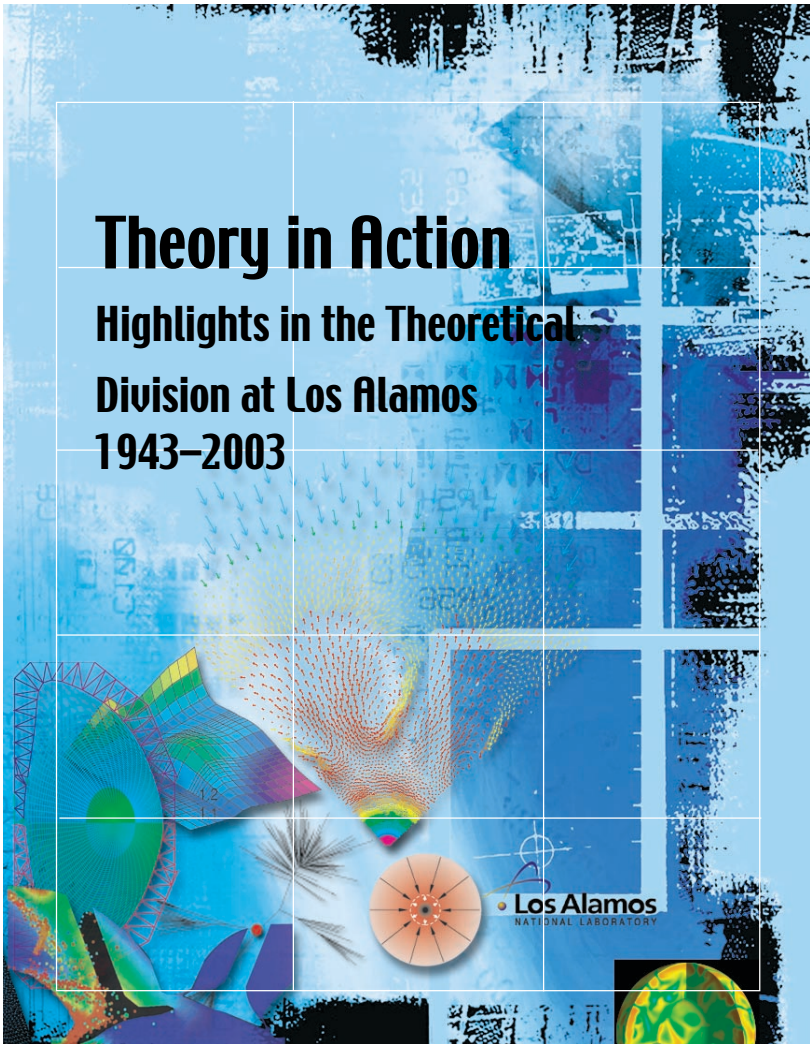
Note to Readers: To help you navigate within this electronic document, we have provided, at left, bookmarks for each chapter. By clicking on a topic, you can go to the chapter on that topic. If you wish to access a particular article, click on the arrow to the left of the appropriate chapter title. A list of articles in that chapter will appear.

To learn what articles are in the entire book, click the "Contents" bookmark. From "Contents," you may go directly to any desired article by clicking on the article title.

We have also placed electronic links between elements in each chapter. The linked elements, each of which shows up as an icon of a pointing hand, are in the following places throughout the document:

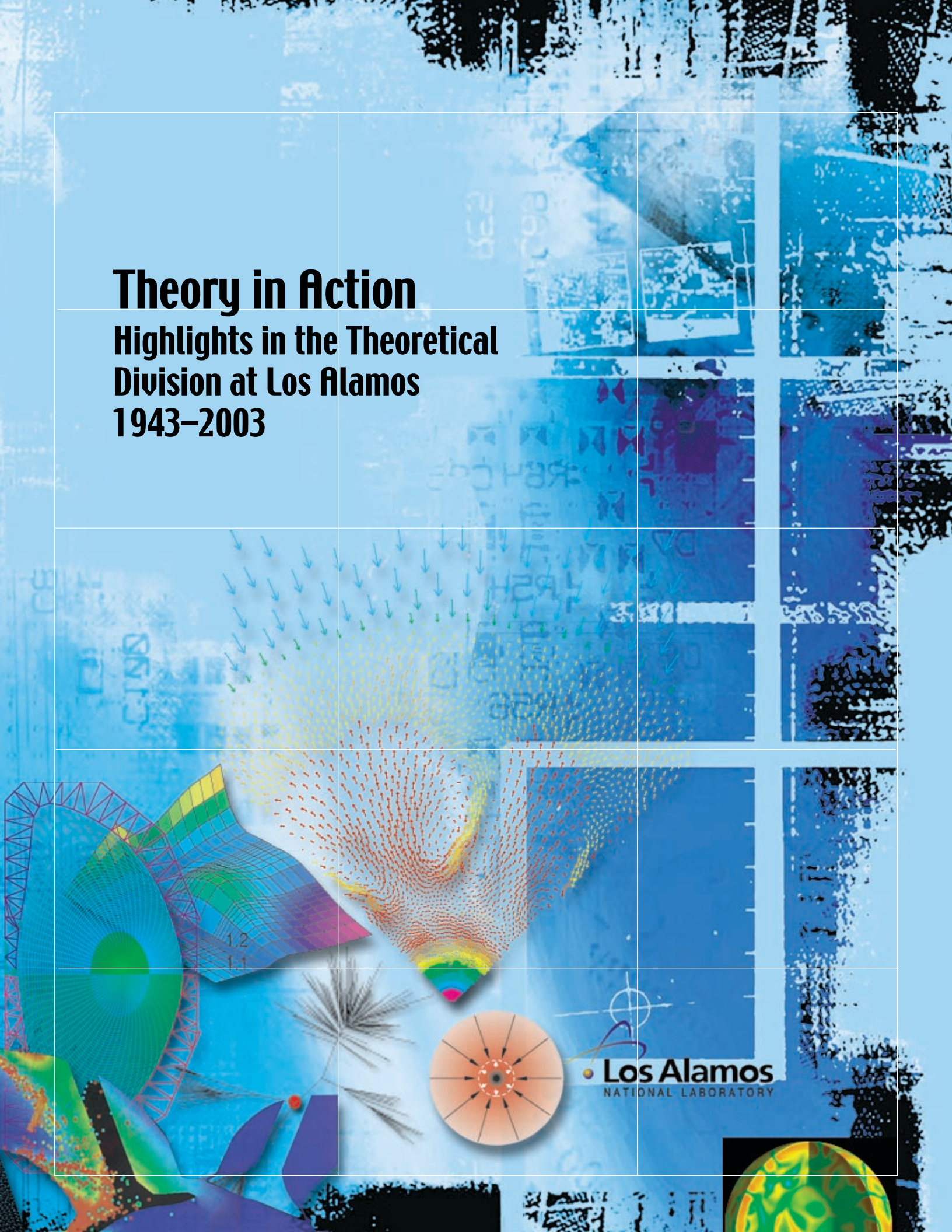


- Contents page (For example, on the "Contents" page, you may click on a specific article to go directly to that article. If you wish to return to the "Contents" page, click on the title of the article.);
- Figure callouts (For example, you may click on "Figure 1" or "Fig. 1" in the text to access the appropriate figure and on the figure captions themselves, such as "Fig. 1." [followed by the caption text], to return to the first figure callout in the article text.);
- References in the text (For example, you may click on "[1]" in the text to find the complete reference in "Appendix A: References" and "[1] H.-P. Bunge, . . ." in "Appendix A: References" to return to the reference callout in the article.);
- Researcher names (For example, you may click on "Bell" in the margin at the beginning of an article to see other pages on which that author appears in "Appendix C: Index of Researchers" and on the specific page number of "Bell 9, 95, 119, . . ." in "Appendix C: Index of Researchers" to access an article by that author or contributors.); and
- Chapter titles (For example, you may click on "Fluid Dynamics" on the title page [page 4 of this electronic file] to access a particular chapter. If you click on the chapter title in the upper-right hand corner, it returns you to the title page.).

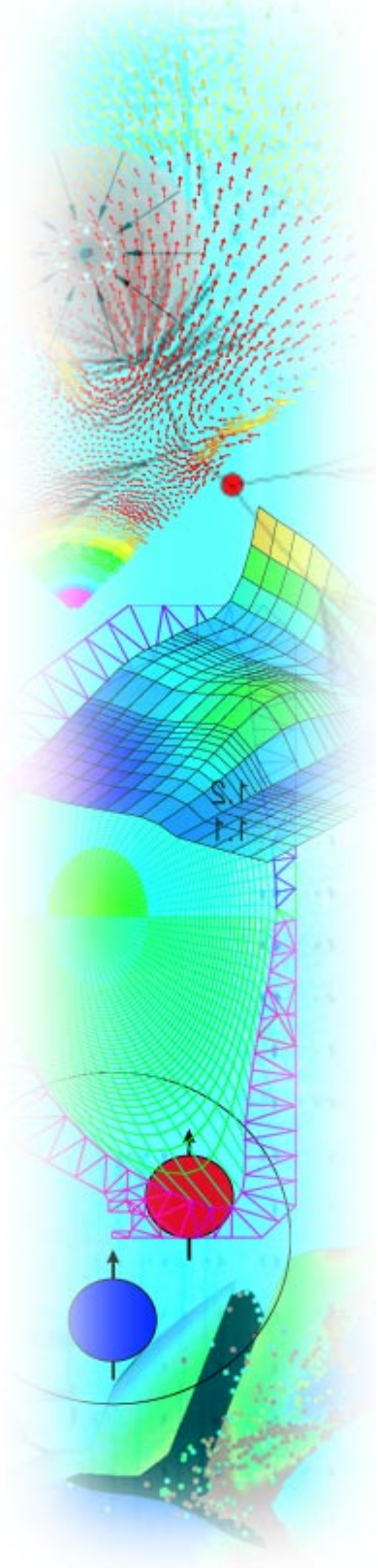


Theory in Action

Highlights in the Theoretical Division at Los Alamos 1943–2003



1943 - 2003
Los Alamos
NATIONAL LABORATORY
Ideas That Change the World



Published by Los Alamos National Laboratory, Theoretical Division

Theoretical Division Leader: **Alan R. Bishop**

Managing Editors: **Denise Sessions and Linda K. Wood** Communication Arts & Services (IM-1)

Editors: **Ryan DeMares, Heather Edwards, Elizabeth Varnedoe, and Mabel Amador** (IM-1)

Designer: **Susan Carlson** (IM-1)

Electronic Publication Specialist: **E. Katherine Valdez** (IM-1)

Additional EPS Support: **Deidré Plumlee, Joyce A. Martinez, and Jeanne Bowles** (IM-1)

Production Assistant: **Christine M. Hedquist** (IM-1)

Additional Publication Support: **Philip Parker** (T-DO)

Printing Coordination: **Lupe Archuleta and Jody Drinkard,** Imaging Services (IM-4)

LA-14000-H
History Report
Unclassified

Issued April 2003

Theory in Action

Highlights in the Theoretical Division at Los Alamos 1943–2003

Compiled by Francis H. Harlow and H. Jody Shepard

Contents

Contents

Contents

Contents

Theoretical Division, The Beginning

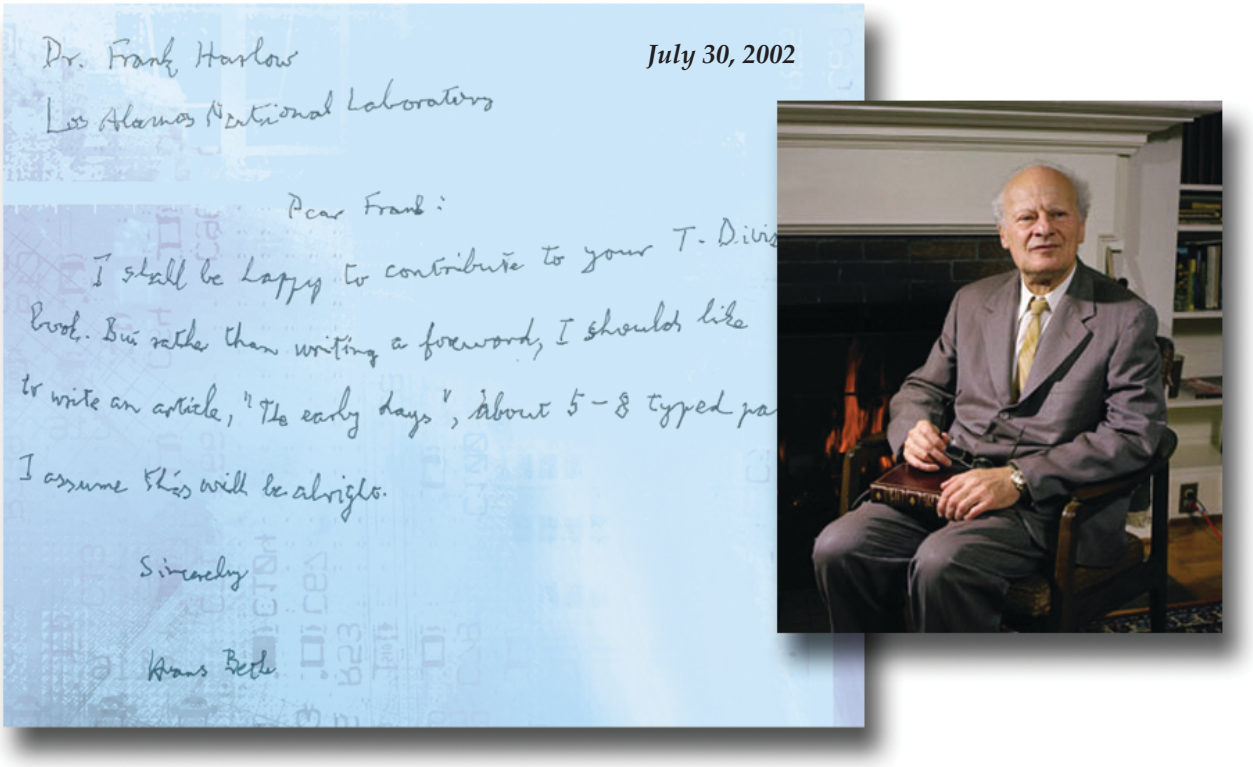
by Hans A. Bethe

Newman Laboratory for Elementary Particle Physics
Cornell University
Ithaca, New York 14853, USA

I

Los Alamos had its beginnings in October 1942. Robert Oppenheimer was to be the director, an unexpected, courageous, and wise decision by General Leslie Groves. It was to be organized on the pattern of the MIT Radiation Laboratory, in Division and Group. There was to be an experimental physics division (P), a chemistry division (C), an ordinance division (O) and several others. But no theoretical division.

Oppie wanted to direct theoretical physics himself. But his senior advisors from MIT Radiation, I. I. Rabi and Robert Bacher, told him, "You are crazy. As director, you will have your hands full. You cannot lead theoretical physics in addition." So Oppie agreed to have a T-Division. Again Rabi and Bacher recommended that I be the leader of T-Division. Oppie agreed, but he could not quite let go. Once we were at Los Alamos, senior members of T-Division had weekly meetings. Oppie came to nearly every one and contributed.



Dr. Frank Harlow
Los Alamos National Laboratory

July 30, 2002

Dear Frank:

I shall be happy to contribute to your T. Div
book. But rather than writing a foreword, I should like
to write an article, "The early days", about 5–8 typed pa
I assume this will be alright.

Sincerely

Hans Bethe

II

I had three obviously senior members: Serber, Teller, and Weisskopf, all in the their mid-thirties. Clearly, they would be group leaders. In addition, I had a mature mathematician, Flanders, above 40 years old, who directed a number of women (4 to 10), mostly wives of staff members who did numerical calculations on mechanical calculating machines: let's call them group T-5. Flanders was gentle and liked by everybody.

We also had a most important consultant, John von Neumann, a famous mathematician born in Hungary. He helped us with many mathematical problems, and through the fact that he also consulted for other military projects.

I had at least three younger (under 30 years old) but mature physicists, Christy, Feynman, and Metropolis. Each of them did important work later. Feynman showed outstanding ability in the discussions of the introductory lectures given by Serber, in which Serber explained to us what he and his assistants, Frankel and Nelson, had already calculated before Los Alamos.

III

It was clear that Serber, Teller, and Weisskopf should be group leaders. Serber had already a good understanding of neutron diffusion and deepened this during the war. Thus, Serber could make a prediction of the yield of the Trinity Test, 8 kilotons, which I adopted in a pre-Trinity colloquium. But one of his group members, Rarita, had gone into more detail and found 20 kilotons. The measurement at Trinity gave somewhat over 20 kilotons.

Weisskopf's group was officially charged with calculating the efficiency. Actually, the main feature was a sign at the entrance to his group, "The Los Alamos Oracle." Experimental physicists came in a constant stream, asking questions such as which nucleus are you interested in? Why? And what about nucleus B?

Victor Weisskopf had a distinguished career after Los Alamos. He was professor of physics at MIT, for many years department chairman. But most importantly, he was the most formative director of CERN at Geneva. They needed a man with deep experience in America and deep roots in Europe. He made CERN the prime lab in high-energy physics.

Edward Teller was dissatisfied, with the program of Los Alamos and with his position in the organization. He had wanted to be T-Division leader, since he had 2–3 years seniority in the uranium project over me. He wanted Los Alamos to put a lot of work into the H-bomb. He finally separated from T-Division to lead an independent group. In this, his senior collaborator, Konopinski, proved conclusively that neither A- nor H-bombs could set the atmosphere on fire by nuclear reactions such as $N + N \rightarrow C + O$.

While Teller was still in T-Division, the Lab adopted the implosion as its main line. Teller pointed out that, if successful, implosion would compress the fissile material and thus make one critical mass into several. He provided an equation of state that has been generally used. But he did not want to get into the detailed theory of implosion. This work was left to his successor, Rudolf Peierls, who joined us from the English project.

IV

Richard Feynman's group (T-4) had the problem of computing the behavior of uranium hydride as a weapon (UH_2). Feller had suggested this because it had a much smaller critical mass than uranium, perhaps one-half. But due to the hydrogen, the neutrons are slowed down severely, and we suspected that the efficiency would be low. This is what group T-4 found out by detailed computation.

The theoretical methods of group T-4, however, of keeping track of many velocity groups of neutrons simultaneously were later found useful in the design of water-based, peaceful nuclear reactors.

Group T-4 was very rambunctious, throwing each other against the flimsy walls of their two rooms. But it was the only group whose members later gathered two Nobel Prizes: Feynman for Quantum Electrodynamics and Fred Reines for measuring neutrinos from a large nuclear reactor.

V

Los Alamos developed two major technologies: implosion and the use of calculating machines for scientific purposes.

In Spring 1944, a Los Alamos experiment showed that ^{240}Pu had spontaneous fission and that therefore plutonium bombs needed to be assembled especially rapidly. A method to do this, implosion, had been proposed by Seth Neddermyer at the beginning of Los Alamos, but Neddermyer's experiments shattered rather than assembled the material.

The solution was given to us by James Tuck, who was sent to us by the British project. Tuck had worked at Oxford with explosive lenses to produce plane detonation waves. He reproduced these at Los Alamos. John von Neumann transformed them theoretically into spherically convergent waves. How did the British know we needed these?

To put the explosive lenses into practice, we got America's greatest expert in explosives, George Kistiakowsky, to head a new division that employed about 100 technicians. By the experiments of still another division, G, it was shown that the explosives division was successful.

But since this was a difficult technical program and we could not be absolutely sure, the Lab accepted happily the proposal by Robert Christy to start with the bomb fully assembled and rely on the compression by implosion to bring it from slightly subcritical to several times critical mass. (The Russians, commanded by Beria, precisely copied our design.)

VI

IBM machines came to us on the advice of Dana Mitchel, the leader of the Procurement Division. At Columbia, he had seen Professor Eckert calculate the orbits of planets over millions of years with IBM machines.

IBM machines then were electromechanical and operated with punch cards. There were some six machines that had to be used in succession. Cards had to be carried from one machine to the next. Later, the director of Brookhaven spoke of his "card-carrying Ph.D.s."

The IBM machines, each the size of a writing desk, arrived at Los Alamos in small parts, with the promise of an expert to get them up. He did not show immediately, so Feynman and Nelson unpacked them and put them together using the printed instructions. About a week later came the trained IBM operator, also in the Army. He found all the machines set up in a room, ready to work. "I have never seen that," he said, "that laymen put these machines together, correctly and ready to work."

I appointed Frankel and Nelson to be co-chairmen of the IBM group. It did not work. The chairmen were fascinated with the machines and made them do all sorts of numerical and algebraic games, but never anything useful for Los Alamos. So I deposed them.

I appointed Feynman instead. He soon devised a program to calculate the progression of the pressure wave into the uranium and plutonium using Teller's equation of state.

The SEDs, mostly graduate students drafted into the Army, loved the work on the IBMs. Once Feynman, or Nelson, or Metropolis—I do not remember whom—had written the program, they told Feynman to stay away, they would do it fastest alone. It took about a week for each model of a weapon.

VII

John von Neumann, on his consulting visits, observed the efficient work of the IBM machines. "We must use them more generally," he decided, "and they must work electronically." So he got a couple of engineers and mathematicians at Princeton to work on this problem and to actually construct electronic computers. It took a couple of years before workable machines were available. They were very useful for the design of the first H-bomb. It took several more years to persuade the IBM Company that this was the machine of the future.

Nicholas Metropolis at Los Alamos competed with the Princeton group and built the MANIAC (Mechanical And Numerical Integrator And Computer).

The successful work of T-Division during the War was the seed of modern computers.

I . . . climbed to the top floor of LeConte Hall where Robert [Oppenheimer, director of the newly established Los Alamos Scientific Laboratory] had his office, and pushed open the door . . . Oppie practically threw a piece of paper at me as I came in the door and said, "Here's your damned organization chart." Well, that was the way the lab was organized to start with and the initial organization sort of stuck that way with one exception. He had kept his own responsibility being head of the theoretical division. It didn't take him very long, especially with some pretty good pressure from I. I. Rabi and also from Bob Bacher, to realize that he couldn't manage the laboratory and also look after all of the theoretical work. He gave in and appointed Hans Bethe to take care of the theoretical division work at Los Alamos.

—John H. Manley [former associate director of Los Alamos Scientific Laboratory], "A New Laboratory is Born," in* Reminiscences of Los Alamos: 1943–1945*, (D. Reidel Publishing Company, Boston, Massachusetts, 1980), p. 30. (Various quotations are reprinted with permission from the publisher.)

Bethe
Bishop
Harlow
Shepard
Strottman

Theoretical Division, The Present

by Alan R. Bishop, Division Leader, and Daniel D. Strottman, Deputy Division Leader

From its inception and throughout its distinguished 60-year history, the Theoretical Division of Los Alamos National Laboratory has been a major intellectual resource to the Laboratory and to the United States of America, providing creative scientific and technological solutions in national security and in other problems of national and global importance. To meet those high expectations, the division has pursued a robust portfolio of fundamental research at the frontiers of science, thereby fostering a stimulating environment that attracts talented scientists and creates new skills and programs to benefit Los Alamos missions.

The celebration of the division's history recorded in this book cannot capture the full scope of people, ideas, and impacts over the remarkable 60-yr period: we would need several large volumes to do that. However, the dedicated and caring work of Frank Harlow and Jody Shepard; the expertise of committed editorial, composition, and design staff from the Laboratory's Communication Arts and Services Group; and the contributions of many distinguished staff and alumni have given us an archival window into six decades of creativity, major discipline and mission challenges, and successes. This book gives us a glimpse of the many efforts that altogether constitute a very special national treasure—the people and products of T-Division.

As the Laboratory faces the new national security and scientific challenges of this 21st Century, Theoretical Division is boldly exploring new interdisciplinary frontiers. As articles in this volume attest, at their core these frontiers demand the integration of theory, modeling, simulation, and visualization that are so richly represented in T-Division. We are now accepting the challenge of implementing a bold strategy to provide new tools to interpret and guide experiments and to establish *science-based predictive tools for complex systems*. A science-based stockpile stewardship program, the centerpiece of the United States Department of Energy weapons complex, depends critically on the viability of this approach. Likewise, the new technical challenges in homeland security and threat reduction that the nation faces can only be met by the same strategy of predictive tools for decision-makers who are equipped and experienced in understanding complex systems.

Indeed, the reliance of a fully integrated science-based prediction strategy extends to virtually every major initiative at Los Alamos. This reliance, of course, engenders its own huge challenge: *the coupling of theory, computational simulations, and experiments as a cornerstone of technical programs and organization requires new generations of ideas and concepts to qualitatively improve the usability, fidelity, certainty, reliability, and sustainability of the tools*—in weapons, homeland security, biology, nanomaterials, energy, infrastructure, frontier science, and more.

We were fortunate to be able to interview Hans Bethe on the occasion of his receiving the first Los Alamos Medal in 2002. As the first T-Division leader, he set truly world-class standards for developing and applying the best science. We are particularly grateful for his reminiscences in this book that will continue to motivate future generations of Los Alamos staff, as his regular visits inspired all previous staff.

Los Alamos is uniquely equipped to accept this challenge of a science-based methodology for truly complex systems. The staff of the Theoretical Division will continue their seminal roles in the formation and development of the Laboratory's core integrated capability by melding theory, modeling, and simulation with experiments to provide the necessary new conceptual frameworks and predictive tools. We look forward to an exciting and challenging future for which our luminous history has prepared us well!

Eventually, in 1951, a bright idea emerged jointly from Stan Ulam, a very well known mathematician, and Edward Teller. It's hard to say about ideas. The day, of course, is long gone in group research when one man sitting in one lab with his own string and sealing wax makes a great invention. The way things go in a laboratory is that somebody sits around the coffee table with somebody else and says, "Hey, I wonder if this would work?" Then some other chap says, "Ah, no, that won't work, but you've given me an idea. I wonder if this would work." This constant interplay between people, each one coming up with an unworkable idea, let's say, but fertilizing an idea in somebody else's mind is terribly important. Talk goes back and forth and finally somebody says, "Hey, maybe if we try it this way, it might work." And then the way is found. Well, who invented it? No single individual probably invented it, but the patent [for a device to be used on the fusion bomb] is at least in the name of Teller and Ulam.

—Norris Bradbury [director of Los Alamos National Laboratory, 1945–1970], "Los Alamos—The First 25 Years," in Reminiscences of Los Alamos: 1943–1945, pp. 168–169.

F.D. Roosevelt,
President of the United
White House
Washington, D.C.

Sir:

Some recent work b
unicated to me in manus
um may be turned into a
diate future. Certain

Albert Einstein
Old Grove Rd.
Nassau Point
Peconic, Long Island

August 2nd, 1939

F.D. Roosevelt,
President of the United States,
White House
Washington, D.C.

Sir:

Some recent work by E.Fermi and L. Szilard, which has been communicated to me in manuscript, leads me to expect that the element uranium may be turned into a new and important source of energy in the immediate future. Certain aspects of the situation which has arisen seem to call for watchfulness and, if necessary, quick action on the part of the Administration. I believe therefore that it is my duty to bring to your attention the following facts and recommendations:

In the course of the last four months it has been made probable - through the work of Joliot in France as well as Fermi and Szilard in America - that it may become possible to set up a nuclear chain reaction in a large mass of uranium,by which vast amounts of power and large quantities of new radium-like elements would be generated. Now it appears almost certain that this could be achieved in the immediate future.

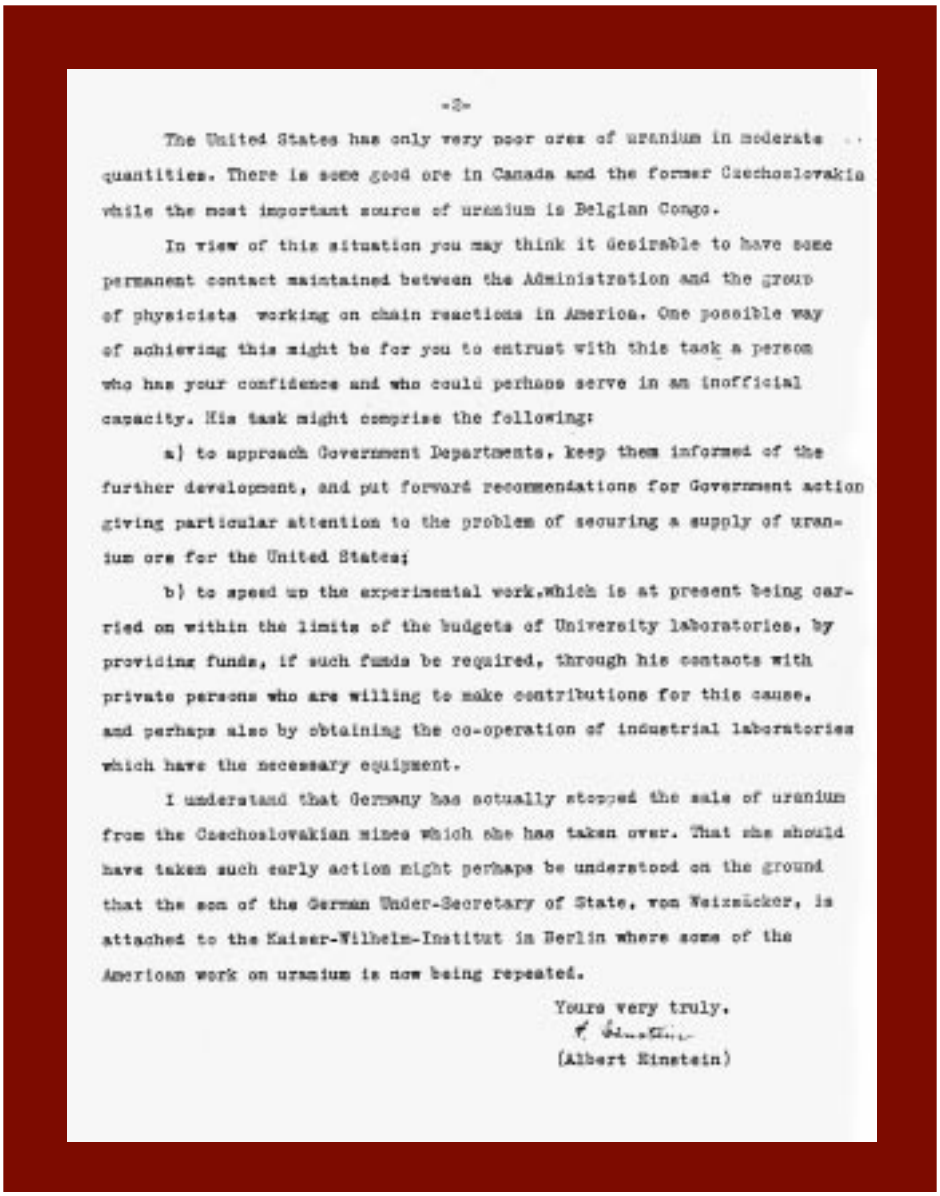
This new phenomenon would also lead to the construction of bombs, and it is conceivable - though much less certain - that extremely powerful bombs of a new type may thus be constructed. A single bomb of this type, carried by boat and exploded in a port, might very well destroy the whole port together with some of the surrounding territory. However, such bombs might very well prove to be too heavy for transportation by air.

Introduction

by Francis H. Harlow

Sixty years ago, the world was in turmoil. World War II was raging on two fronts, European and Western Pacific. The Soviet Union was an ally, but with uncertainties regarding the cordiality of the association.

Early in the war, a small group of scientists in the United States began to worry about a matter with horrendous implications beyond the ordinary battlefield strategies of concern at that time. They foresaw the potential for Germany to develop a super weapon with explosive potentiality much greater than anything imaginable, involving the



-2-

The United States has only very poor ores of uranium in moderate quantities. There is some good ore in Canada and the former Czechoslovakia while the most important source of uranium is Belgian Congo.

In view of this situation you may think it desirable to have some permanent contact maintained between the Administration and the group of physicists working on chain reactions in America. One possible way of achieving this might be for you to entrust with this task a person who has your confidence and who could perhaps serve in an inofficial capacity. His task might comprise the following:

a) to approach Government Departments, keep them informed of the further development, and put forward recommendations for Government action giving particular attention to the problem of securing a supply of uranium ore for the United States;

b) to speed up the experimental work, which is at present being carried on within the limits of the budgets of University laboratories, by providing funds, if such funds be required, through his contacts with private persons who are willing to make contributions for this cause, and perhaps also by obtaining the co-operation of industrial laboratories which have the necessary equipment.

I understand that Germany has actually stopped the sale of uranium from the Czechoslovakian mines which she has taken over. That she should have taken such early action might perhaps be understood on the ground that the son of the German Under-Secretary of State, von Weizsäcker, is attached to the Kaiser-Wilhelm-Institut in Berlin where some of the American work on uranium is now being repeated.

Yours very truly,

(Albert Einstein)

Fig. 1. The famous letter from Einstein to President Franklin D. Roosevelt. **(See p. 8 and text on p. 10.)**

Fig. 2. Black Mesa near Los Alamos.

production of energy from nuclear reactions. These scientists expressed their concerns to Albert Einstein, hoping to elicit his help in bringing this issue to the highest levels of government. The result was the now-famous letter to President Franklin D. Roosevelt. (See Fig. 1.) Einstein's prestige lent great urgency to the matter, to the extent that the President ordered an all-out effort to understand the nature of this threat and to develop our own version of this explosive device, if indeed it turned out to be feasible.

Thus Los Alamos was born. Situated in the remote and rugged terrain of northern New Mexico, the town and Laboratory quickly grew despite the constraints of great secrecy (Figs. 2–5). The country's foremost scientists were recruited to build upon Enrico Fermi's famous experiments in Chicago, which indicated that nuclear fission very likely could be harnessed in such a way as to produce an explosion many orders of magnitude greater than could be achieved with any known chemical explosives.

Experts in theoretical physics and chemistry were joined by the leading experimentalists in the country to form a team whose urgent mission was underscored by the tremendous loss of life taking place every day on land, sea, and in the air, with no end in sight.

Fig. 3. Prehistoric Native American ruins at Bandelier National Monument near Los Alamos.

Introduction

The Los Alamos Scientific Laboratory, led by J. Robert Oppenheimer, was organized into divisions and groups. Of the four technical divisions, Chemistry (C), Engineering (E), Physics (P), and Theoretical (T), only two (P and T) survive to the present day. The total number of divisions soon grew to 18 and changed little while Norris Bradbury was Director of the Laboratory (1945–1970), but proliferated considerably thereafter. (See Fig. 6.)

Much has been written about the historical sequence of events at Los Alamos during WWII (up through 1945) and on the subsequent evolution of the Laboratory. The expanding goal was toward an ever-broadening place where problems of national concern were and are being addressed. The Laboratory continues to the present day, however, to be a place where the design and stewardship of the United States nuclear weapons arsenal persist as its dominant mission.

This book is not a treatise on the fascinating history of the Laboratory and the town of Los Alamos. Our goal is to describe the most influential scientific discoveries and developments taking place over the last 60 years in one of the two divisions that have survived since the beginning, the Theoretical Division. Many great advances have taken place in the Physics Division and in all of the other technical divisions that have been organized in the ensuing years, but we describe here only those advances that involved people in T-Division.

This book is organized into chapters highlighting many different disciplines and activities. For each we consulted with experts currently employed by the Laboratory, working elsewhere, or retired. We asked these people to contribute articles that could be anywhere from a paragraph to ten or so printed pages in length.

Fig. 4. The road to Los Alamos, 1943.

Fig. 5. Early Laboratory buildings.

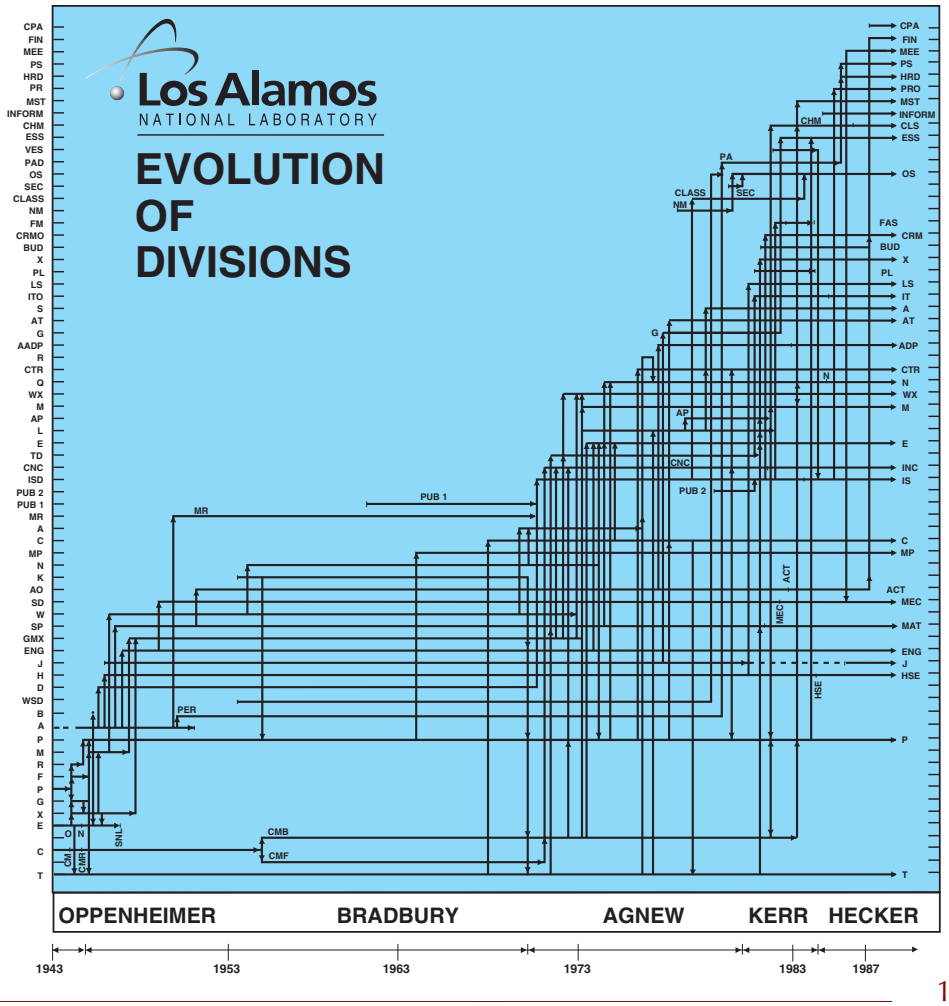


Fig. 6. This chart shows the evolution of the Laboratory's divisions from 1943 to 1987.

Fig. 7. The modern Laboratory site, TA-3.

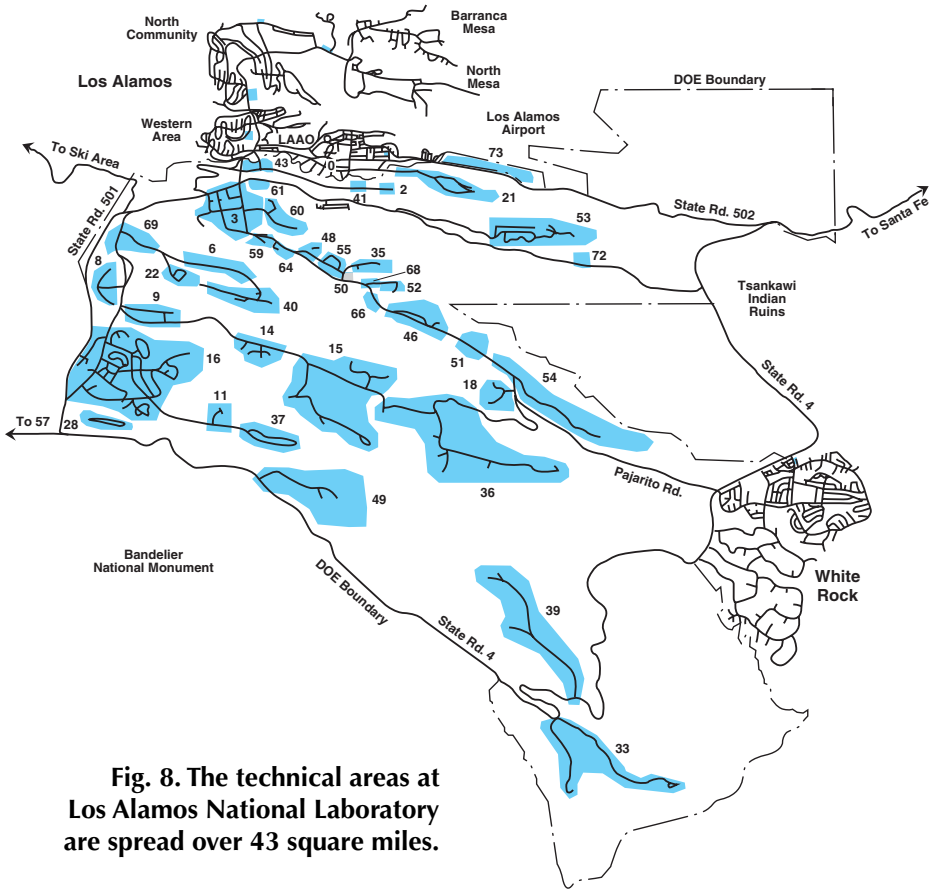


Fig. 8. The technical areas at Los Alamos National Laboratory are spread over 43 square miles.

The choice for inclusion in this book was based on the overall impact of each accomplishment to the nation, to science, to industry, and to the well-being of our citizens.

For each topic, we present the essence of the activity without describing the vast array of technical details that lie behind the work. We hope that interested readers will find meaningful discussions in disciplines that are not their field of specialization. T-Division has been (and is) a remarkable scientific and sociological entity from the beginning and through its evolving response to numerous national and international developments over the sixty years of its existence. In fact, the enormous interconnection among the disciplines gives T-Division much of its strength as an organization. Our wish is to give the reader a flavor of this organization through a discussion of its accomplishments.

Fig. 9. Los Alamos Ranch School, 1942.

The Setting

The Los Alamos National Laboratory, as it is now called, consists of approximately 12,000 technical and support employees working in buildings and laboratories spread over 43 square miles. (See Figs. 7 and 8.) The county of Los Alamos houses about 20,000 people; in addition, many employees commute from nearby towns and cities.

Sixty years ago (1943), the Laboratory and the town were much smaller, consisting of approximately 500 technical workers and 1,000 residents. Before that time there were a few homesteads and a famous boy's school at this location (Fig. 9).

The site for the Laboratory was chosen for the security that is enabled by the rugged terrain of mesas and steep-sided canyons. (See Fig. 10.) Access was restricted to one main road, which was winding and treacherous to maneuver, plus two others that were even more difficult to traverse.

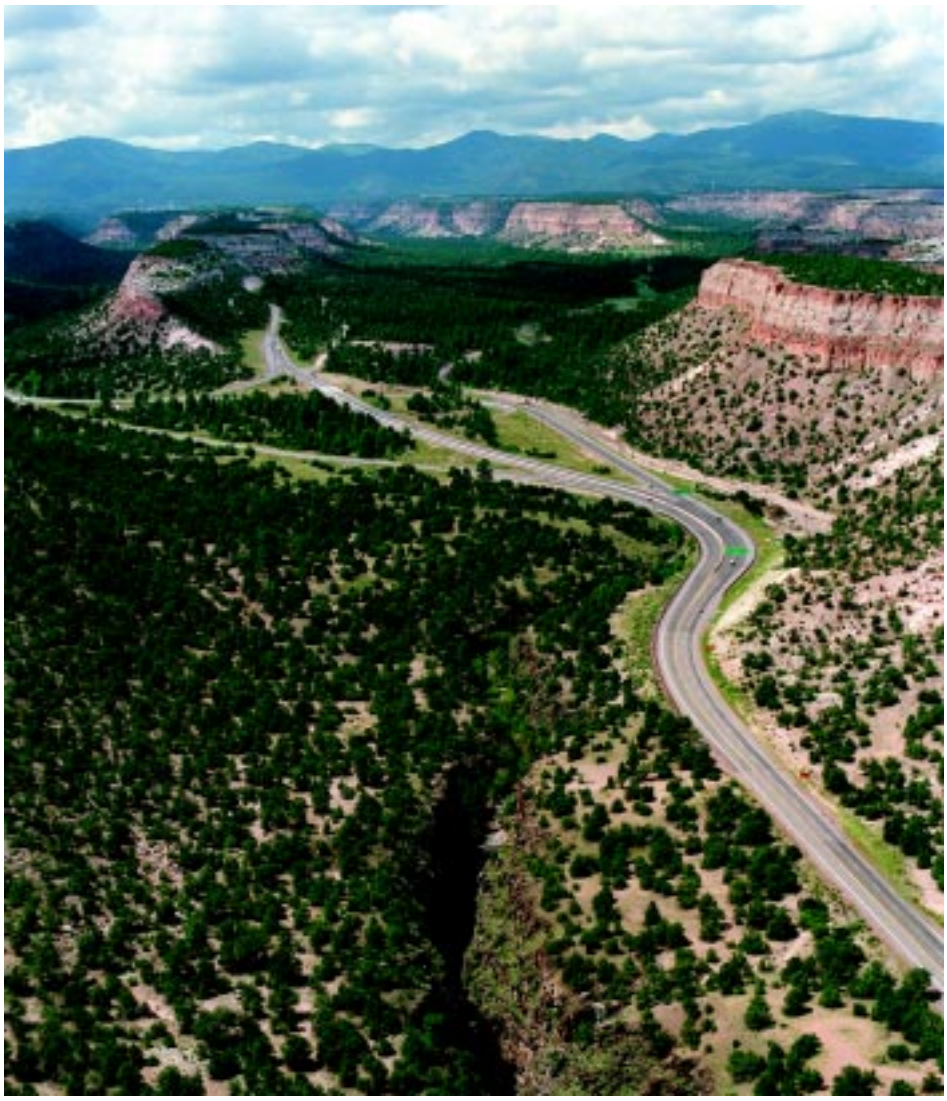


Fig. 10. Rugged terrain and steep-sided canyons characterize the location of Los Alamos.

The origin of this wild and beautiful terrain can be traced back to a million years ago, when volcanic eruptions piled up more than 500 feet of volcanic ash and lava at the base of a residual ring of mountains. In the succeeding millennia, the ash beds were dissected by rain and wind erosion into numerous canyons, typically up to half a mile wide and ten miles long. To travel from one mesa to its neighbor required a trip to the head of a canyon, where the crossing was relatively easy, or else a strenuous climb down and up through the canyon on one of several steep trails, some of which had been carved by the prehistoric Indians who lived here in many hundreds of dwellings, especially in the period 1250–1580.

Fig. 11. World War II Los Alamos entrance gate.

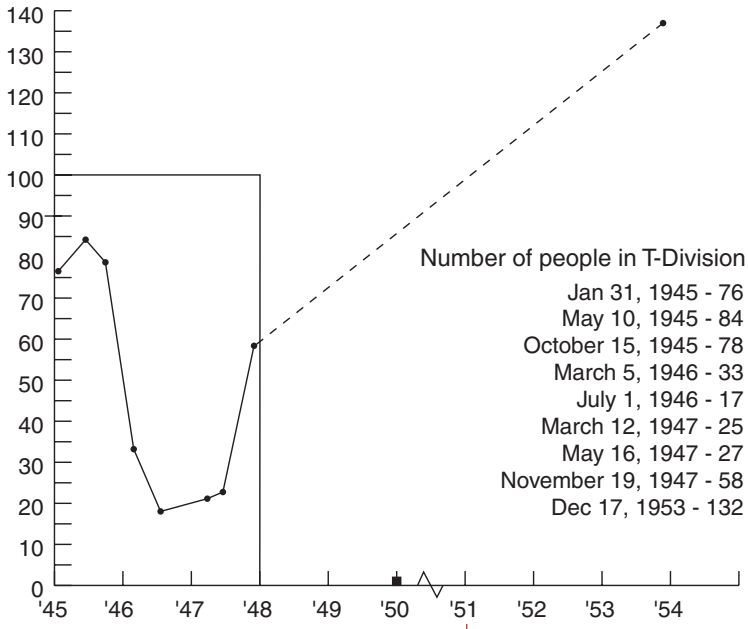
It was very difficult to approach Los Alamos through this rugged terrain by any route except those that were well guarded and, until 1957, required an entry pass for all residents over 14 years of age (Fig. 11). Today, there are two unguarded highways allowing easy access to Los Alamos from both the east and the west. A small airport also serves the town, but there is no railroad; everything has to be transported to Los Alamos by truck.

In its earliest days, the living in Los Alamos was relatively primitive. Only a few residences had bathtubs. Deep ruts and mud were inconveniences in the unpaved roads. The weather was pleasant, except for the infrequent deep snows. Shopping was mostly carried out in the Army Post Exchange. Security guards were everywhere.

Theoretical Division Personnel

The earliest personnel roster that we have been able to locate was printed on January 31, 1945, and shows a total of 76 people in the division. The population of the entire Laboratory dropped considerably in 1946 from approximately 3,000 (in May 1945) to 1,700 (late 1940s). Most people thought the job was done—it was time to return to peacetime pursuits. For the next 2 years, T-Division consisted of only 20–30 people. By the beginning of 1948, the number of people in the division had risen to 58 and continued to rise to 132 members by 1953. (See Fig. 12.) There are currently about 150 technical staff members in T-Division and another 150 support staff, graduate research assistants, postdoctural researchers, Lab fellows, and associates. The early rosters are fascinating to peruse, and many famous names (and one infamous name, Klaus Fuchs) can be recognized. Three examples of the rosters are illustrated in Figs. 13–15. The first T-Division roster to list an employee who is still present (2003) in the division is from December 17, 1953.

Fig. 12. The figure shows the number of people in T-Division between 1945 and 1954.



The sequence of T-Division leaders is as follows: Hans Bethe 1943–1945, George Placzek 1946, Robert Richtmyer 1947, Carson Mark 1948–1973, Peter Carruthers 1974–1979, George I. Bell 1980–1990, Richard A. Slansky 1990–1997, T. Daniel Butler (acting) 1997–1999, and Alan R. Bishop 1999–present.

THEORETICAL DIVISION
MAY 10, 1945

H.A. Bethe, Division Leader, E-208, Ext 71
V.F. Weisskopf, Deputy Division Leader, E-217, Ext 74
J. von Neumann, Consultant, E-205½, Ext 72 R3

GROUP T-1
R. Peierls, E-119, Ext 178
R.F. Christy, Section Leader E-120, Ext 178 R2
K. Fuchs, Section Leader, E-118, Ext 77
Baroody, E.M., E-121, Ext 178 R2
Calkin, J.W., E-117, Ext 77
Inglis, D.R., T-35, Ext 54
Keller, J., E-120, Ext 178 R2
Penny, W.G., E-101A, Ext 470
Podgor, T/5 S., E-116, Ext 469
Roberts, T/5 A.E., E-117, Ext 77
Skyrme, T.H.R., E-118, Ext 77
Stark, R.H., E-116, Ext 469
Stein, T/5 P.R., E-121, Ext 178

GROUP T-2
R. Serber, Leader, E-109, Ext 177
L.I. Schiff, Alt. Leader, E-108, Ext 76
Case, K.M., E-107, Ext 76
Glauber, R., E-107, Ext 76
Kurath, T/3 D., E-202, Ext 205
Rarita, W., E-111, Ext 177
Richman, C., E-110, Ext 177
Stehle, T/5 P., E-107, Ext 76

GROUP T-3
V. Weisskopf, Leader, E-217, Ext 74
R.E. Marshak, Alt. Leader, E-218, Ext 74 R2
Bellman, Pvt. R., E-222, Ext 468
Cohen, T/5 S., E-219, Ext 74 R2
Lennox, T/5 E., E-218, Ext 74 R2
Olum, Paul, E-216, Ext 74
Smith, J.H,, E-219, Ext 74
Wing, Milton, E-222, Ext 468
Bowers, W.A., E-216, Ext 74

GROUP T-4
R. Feynman, Leader, E-206, Ext 72
J. Ashkin, E-209, Ext 72 R2
Ehrlich, R., E-210, Ext. 72 R2
Peshkin, T/4 N., E-203, Ext. 79 R3
Reines, F., E-210, Ext 72 R2
Welton, T.A., E-209, Ext 72 R2

GROUP T-5
D. Flanders, Leader, E-205, Ext 69 R1
P. Whitman, Alt. Leader, E-204, Ext 69 R2
Atkins, A.L., E-211, Ext 73 R2
Davis, R.R., E-215, Ext 74
de le Vin, E., E-201, Ext 79
Elliott, J., E-213, Ext 73 R1
Hauser, T/5 F.H., E-214,
Huber, T/4 .C., E-120, Ext 70
Hudson, H., E-214, Ext 73 R1
Inglis, B., E-212, Ext 73 R2
Johnson, M., E-212, Ext 73 R2
Kellogg , T/5 H., E-203, Ext 69 R3
Langer, B., E-201, Ext 79
Page, T/3 W., E-214, Ext 73 R1
Rau, E., T/3, E-120, Ext 70
Staley, T/3 J., E-212, Ext 73 R2
Teller, M., E-214, Ext 73 R1
Vuletic, T/5 V., E-211. Ext 73 R2
Wilson, F., E-212, Ext 73 R2
Wright, T/5 E., E-213, Ext 73 R1
Young, T/Sgt., G., E-211, Ext 73 R2

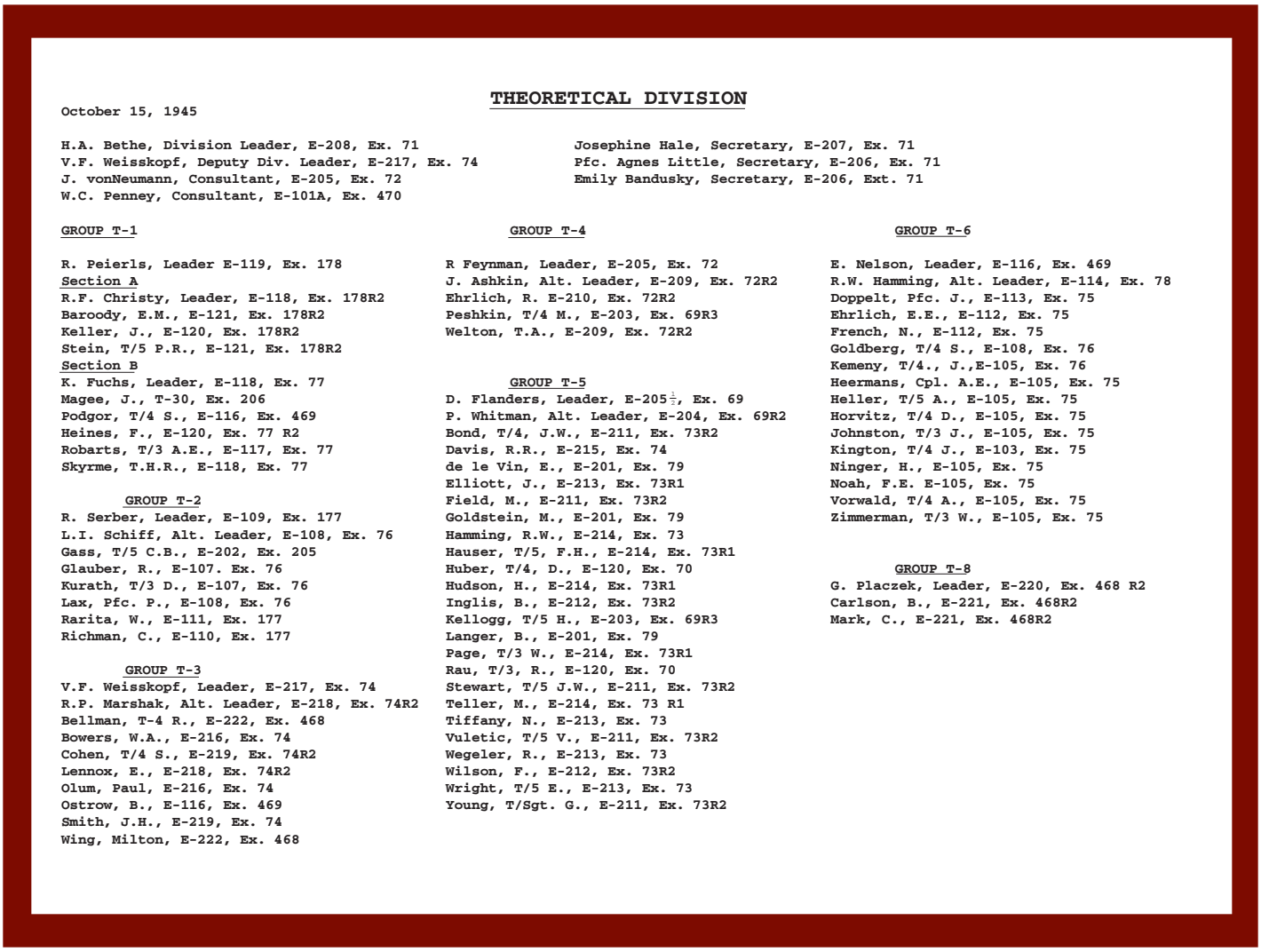
GROUP T-6
E. Nelson, Leader, E-116
N. Metropolis, Alt. Leader, E-115, Ext 78
R. Feynman, Consultant, E-206, Ext 72
Ewing, F.E., E-112, Ext 75
Goldberg, T/5., E-108, Ext 76
Kemeny, Pvt. J., E-105, Ext 75
Hamming, R.W., E-114, Ext 78
Heermans, Corp. A., E-105, Ext 75
Heller, T/5 A., E-105, Ext 75
Hurwitz, T/5 D., E-105, Ext 75
Johnston, T/3 J., E-103, Ext 75
Kington, T/4 J., E-105, Ext 75
Livesay, N., E-112, Ext 75
Ninger, H., E-105, Ext 75
Noah, F.E., E-105, Ext 75
Vorwald, T/5 A., E-105, Ext 75
Zimmerman, T/3 W., E-105, Ext 75

GROUP T-7
J. Hirschfelder, Leader, T-30, Ext 206
J. Magee, Alt. Leader, T-30, Ext 206
Brummer, T/5 E., T-28, Ext 206
Feckete, T/4 P., T-26, Ext 206
Larson, T/4 L., T-28, Ext 206
Ostrow, E., T-30, Ext 206
Schwartz, T/4 P., T-28, Ext 206

GROUP T-8
G. Placzek, Leader, E-220, Ext 468
Mark, C., E-221, Ext 468 R2
Carlson, B., E-221, Ext 468 R2
Day

Fig. 13. Early T-Division roster (May 10, 1945).

Each of these leaders has had a distinguished scientific career, in addition to serving as an administrator. Each has had to guide the division through periods of significant change, in the scientific challenges that needed to be addressed, in the processes for funding (from sources inside and outside of the Laboratory), in hiring policies and enforced reductions in personnel, and in accountability for safety and security.



THEORETICAL DIVISION

October 15, 1945

H.A. Bethe, Division Leader, E-208, Ex. 71
V.F. Weisskopf, Deputy Div. Leader, E-217, Ex. 74
J. vonNeumann, Consultant, E-205, Ex. 72
W.C. Penney, Consultant, E-101A, Ex. 470

Josephine Hale, Secretary, E-207, Ex. 71
Pfc. Agnes Little, Secretary, E-206, Ex. 71
Emily Bandusky, Secretary, E-206, Ext. 71

GROUP T-1

R. Peierls, Leader E-119, Ex. 178
Section A
R.F. Christy, Leader, E-118, Ex. 178R2
Baroody, E.M., E-121, Ex. 178R2
Keller, J., E-120, Ex. 178R2
Stein, T/5 P.R., E-121, Ex. 178R2
Section B
K. Fuchs, Leader, E-118, Ex. 77
Magee, J., T-30, Ex. 206
Podgor, T/4 S., E-116, Ex. 469
Heines, F., E-120, Ex. 77 R2
Robarts, T/3 A.E., E-117, Ex. 77
Skyrme, T.H.R., E-118, Ex. 77

GROUP T-2

R. Serber, Leader, E-109, Ex. 177
L.I. Schiff, Alt. Leader, E-108, Ex. 76
Gass, T/5 C.B., E-202, Ex. 205
Glauber, R., E-107. Ex. 76
Kurath, T/3 D., E-107, Ex. 76
Lax, Pfc. P., E-108, Ex. 76
Rarita, W., E-111, Ex. 177
Richman, C., E-110, Ex. 177

GROUP T-3

V.F. Weisskopf, Leader, E-217, Ex. 74
R.P. Marshak, Alt. Leader, E-218, Ex. 74R2
Bellman, T-4 R., E-222, Ex. 468
Bowers, W.A., E-216, Ex. 74
Cohen, T/4 S., E-219, Ex. 74R2
Lennox, E., E-218, Ex. 74R2
Olum, Paul, E-216, Ex. 74
Ostrow, B., E-116, Ex. 469
Smith, J.H., E-219, Ex. 74
Wing, Milton, E-222, Ex. 468

GROUP T-4

R Feynman, Leader, E-205, Ex. 72
J. Ashkin, Alt. Leader, E-209, Ex. 72R2
Ehrlich, R. E-210, Ex. 72R2
Peshkin, T/4 M., E-203, Ex. 69R3
Welton, T.A., E-209, Ex. 72R2

GROUP T-5

D. Flanders, Leader, E-205½, Ex. 69
P. Whitman, Alt. Leader, E-204, Ex. 69R2
Bond, T/4, J.W., E-211, Ex. 73R2
Davis, R.R., E-215, Ex. 74
de le Vin, E., E-201, Ex. 79
Elliott, J., E-213, Ex. 73R1
Field, M., E-211, Ex. 73R2
Goldstein, M., E-201, Ex. 79
Hamming, R.W., E-214, Ex. 73
Hauser, T/5, F.H., E-214, Ex. 73R1
Huber, T/4, D., E-120, Ex. 70
Hudson, H., E-214, Ex. 73R1
Inglis, B., E-212, Ex. 73R2
Kellogg, T/5 H., E-203, Ex. 69R3
Langer, B., E-201, Ex. 79
Page, T/3 W., E-214, Ex. 73R1
Rau, T/3, R., E-120, Ex. 70
Stewart, T/5 J.W., E-211, Ex. 73R2
Teller, M., E-214, Ex. 73 R1
Tiffany, N., E-213, Ex. 73
Vuletic, T/5 V., E-211, Ex. 73R2
Wegeler, R., E-213, Ex. 73
Wilson, F., E-212, Ex. 73R2
Wright, T/5 E., E-213, Ex. 73
Young, T/Sgt. G., E-211, Ex. 73R2

GROUP T-6

E. Nelson, Leader, E-116, Ex. 469
R.W. Hamming, Alt. Leader, E-114, Ex. 78
Doppelt, Pfc. J., E-113, Ex. 75
Ehrlich, E.E., E-112, Ex. 75
French, N., E-112, Ex. 75
Goldberg, T/4 S., E-108, Ex. 76
Kemeny, T/4., J.,E-105, Ex. 76
Heermans, Cpl. A.E., E-105, Ex. 75
Heller, T/5 A., E-105, Ex. 75
Horvitz, T/4 D., E-105, Ex. 75
Johnston, T/3 J., E-105, Ex. 75
Kington, T/4 J., E-103, Ex. 75
Ninger, H., E-105, Ex. 75
Noah, F.E. E-105, Ex. 75
Vorwald, T/4 A., E-105, Ex. 75
Zimmerman, T/3 W., E-105, Ex. 75

GROUP T-8

G. Placzek, Leader, E-220, Ex. 468 R2
Carlson, B., E-221, Ex. 468R2
Mark, C., E-221, Ex. 468R2

Fig. 14. Early T-Division roster (October 15, 1945).

Over the years, five new divisions have developed from parts of T-Division:

parts became C-Division (computers)—1968;

parts became TD-Division (nuclear weapons analysis)—1971;

parts became Q-Division (nuclear reactors for power)—1975;

parts became L-Division (Laser Research and Technology Division)—1976; and

parts of T- and TD-Divisions became X-Division (nuclear weapons analysis)—1981.

Activities of the Theoretical Division

The division carries on research for both fundamental and applied problems. The chapter titles show that these activities spread across virtually every aspect of the "hard" sciences (e.g., atomic physics and numerical analysis) and also include significant involvement in biology and epidemiology. A particular strength of the division is the multidisciplinary interactions that have taken place throughout the history of the Laboratory. The results of this research have been published in numerous papers, books, and review articles, and disseminated through many invited lectures, contributed papers, and conferences. In addition, significant information is recorded in classified

July 1, 1946

THEORETICAL DIVISION

G. PLACZEK, Division Leader, E-203, Ext. 71
R. D. Richtmyer. Alternate Division Leader
and Personnel Representative, E-203, Ext. 71
S. M. Ulam, E-203, Ext. 69R2

Charlotte Mullen, Secretary, E-207, Ext. 71
T/4 Agnes Little, Secretary, B-207, Ext. 71
Ruth Jourdan, Secretary, E-203, Ext. 71

Group T-1

FRED REINES, E-210, Ext. 72R2

T.H.R. Skyrme, E-113, Ext. 77
M. H. Hull, Jr., E-119, Ext 206
Pvt. Ernest Gray, E-121, Ext. 178

Group T-2
Section a

DONALD A. FLANDERS. E-205, Ext. 69

Josephine Elliott, E-213, Ext. 73R1
Marguerita Field, E-211, Ext. 73R2
Marie Goldstein, E-201, Ext. 79
Nell Tiffany, E-213, Ext. 73
Margarot Johnson, E-212, Ext. 73R2
Lillian Carson, E-201, Ext. 79

Section b

Dan Kervitz, (Section Leader), B-103, Ext. 75
John Johnson, B-103, Ext. 75
T/4 Jack Yoos, B-115, Ext. 75

Group T-3

ROLF LANDSHOFF, T-31, Ext. 215R2

Bernice Kelly, T-33a, Ext. 212
Joseph Mullanay, T-31, Ext. 215R2
Sam Cohen, T-33b, Ext. 212
David Judd, Lt. j.g., T-29, Ext. 215R2

Group T-4

CARSON MARK, E-220. Ext. 468R2
Bengt Carlson, E-221, Ext. 468R2
Max Goldstein, E-202, Ext. 205

Fig. 15. Early T-Division roster (July 1, 1946).

reports, which contain information that is confidential, secret, or top secret. We have extracted from recent T-Division self-assessment documents data from which we make the following conservative extrapolations to cover the full 60 years of T-Division publishing activities:

number of scientific papers in professional journals—25,000;

number of invited lectures—28,000; and

number of classified reports—600.

Another type of T-Division activity is its participation in summer programs for students at all levels from high-school senior to Ph.D. candidates. In addition to high-school students and Ph.D. candidates, T-Division has sponsored undergraduate and graduate students, military academy students, and staff with university teaching appointments. From the recent self-assessment documents we can make these extrapolations in regard to number of students sponsored,

undergraduate students—600; and

graduate students—2,300.

Division personnel have supervised about 50 Ph.D. dissertations. In addition, T-Division has an average each year of about 55 members of editorial boards for professional journals.

T-Division has been an especially active participant in the Laboratory's postdoctoral program, with many of these postdocs subsequently becoming members of the staff in one or another of the Laboratory's divisions. The estimated number of postdoctoral researchers in T-Division over the years is 2,700.

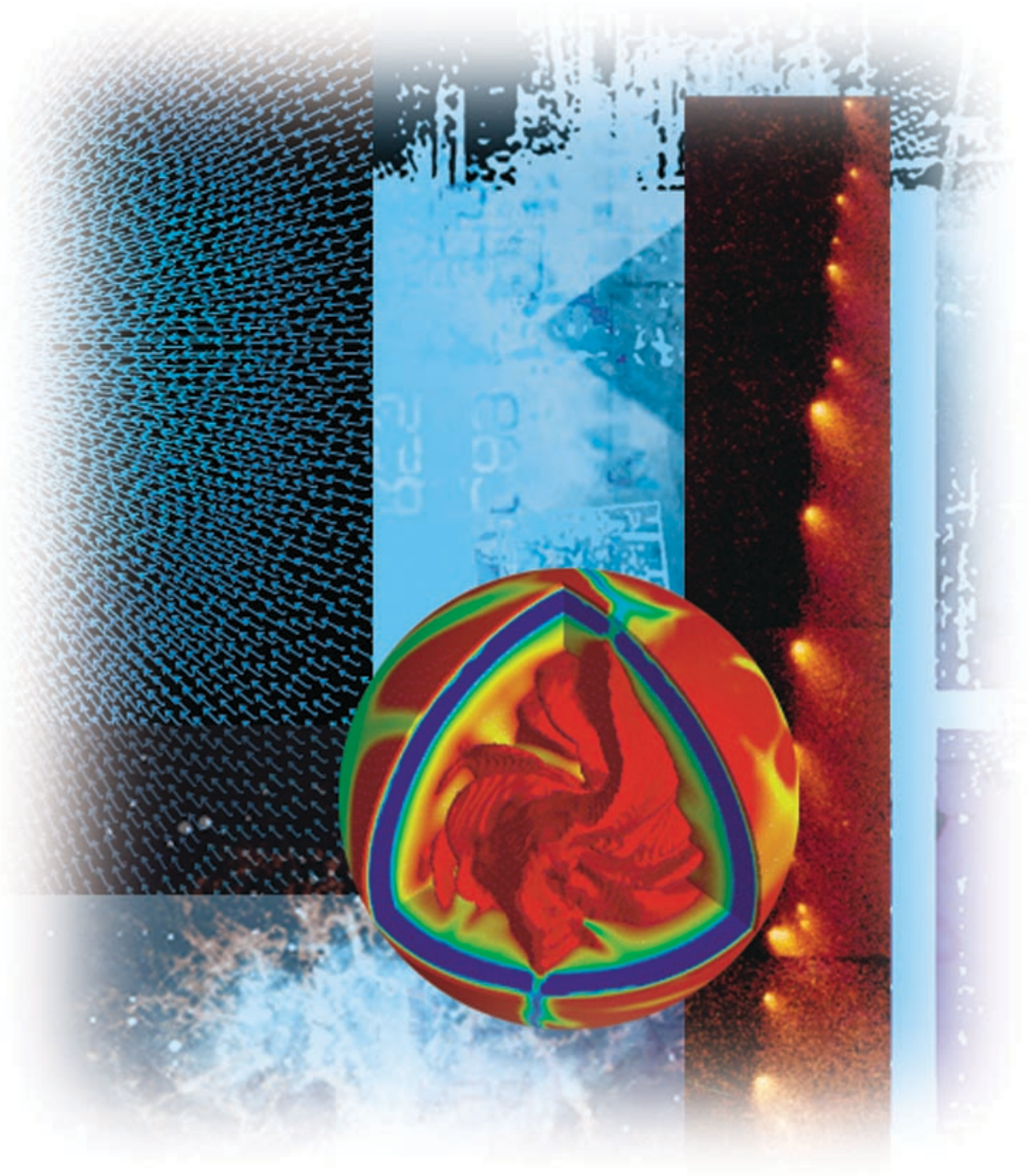
Another facet of T-Division participation in national and international activities occurs through the visitor's program. Many professional people from universities, governmental agencies, and industrial institutions come to T-Division as consultants or visiting staff members for frequent and/or extended visits to work in collaboration with the T-Division staff. The number of such visitors over the 60 years is estimated to be at least 40,000.

Thus, T-Division serves as a "Window on the World" for the Laboratory, through which many of the world's most prestigious scientists who have participated in the research of the division staff (Fig. 16). Nobel Prize Laureates who have had appreciable contact with the staff include such notable people as Enrico Fermi, Hans Bethe, Subrahmanyan Chandrasekhar, Richard Feynman, Murray Gell-Mann, and Frederick Reines.

International collaborations are also an important part of the division's activities, with people from many other countries in residence here. A notable example is the division's participation in the establishment of scientific interactions with Russia after the dissolution of the Soviet Union and the end of the cold war. In particular, there have been many visits and technical discussions with scientists at Chelyabinsk-17 and Sarov (Arzamas-16), where much of the Russian nuclear-weapons research takes place.

Even a casual perusal of the highlights described in this book shows that the Theoretical Division of the Los Alamos National Laboratory has been a beehive of diverse but interactive activities for the last 60 years. Foremost is our continuing response to Laboratory programs dedicated to solving problems of national (and international) concern. With its superb capabilities, flexibility, and dedicated personnel, the division is poised to continue contributing to the nation's needs for another 60 years and beyond.

Fig. 16. From beautiful Northern New Mexico, T-Division serves as "Window on the World."



Astrophysics, Space Physics, and Planetary Physics

Analyzing the behavior of planets, stars, and the universe brings together virtually every scientific discipline within the Los Alamos Theoretical Division. Radiation, gravitation, and nuclear reactions join with complex material dynamics to generate an astonishing set of phenomena. The study of these processes requires the same capabilities needed for earth-bound analyses, including the turbulent mix of materials, the opacities for radiation transport, and the nuclear reactions that generate both energy and new isotopes of elements. At the other extreme are the dynamics of materials inside a planet, ranging from the slow motion of continental plates to the violence of earthquakes and volcanic eruptions. Also, T-Division has investigated effects and mitigation questions associated with the interaction of space and planets during the bombardment of comets and asteroids.

In this "Astrophysics, Space Physics, and Planetary Physics" chapter, we describe a sampling of T-Division activities that combine intense microscopic processes with those that occur over vast distances.

Down-to-Earth Astronomy: When Astronomical Objects Hit Earth

Benz
Cameron
Colgate
Cox
Davies
Fryer
Goda
Hills
Hut
Leonard
Mader
Muller
Slattery
Solem

by Jack G. Hills

Impacts by astronomical objects are major threats to Earth. Astrophysicists in Theoretical Division have been concerned about such objects since 1980. T-Division researchers have studied impacts on Earth by comet showers and asteroids, the impact that led to the formation of the moon, and the effects such impacts can have on the people and elements of Earth. Also, these researchers have proposed preventive measures to deal with such impacts.

Impacts by astronomical objects are the ultimate threat to Earth. Astrophysicists in Theoretical Division have been concerned about such objects since 1980. When I interviewed, as a prospective staff member, with then group leader, Stirling Colgate, in August 1980, asteroid impacts and their mitigation was one of the topics brought up by Colgate. He had long-term interest in the problem going back to his days at Lawrence Livermore National Laboratory. Subsequent work in T-Division involved comet showers, the impact leading to the formation of the moon, and the effect of asteroid-comet impacts with Earth.

Comet Showers

During his interview process, I gave a talk on comet showers, a concept that I had just developed during the 2 months prior to the interview. This work was published shortly after I became a staff member in T-Division.

The idea for comet showers developed from my work on the tidal disruption of stars by black holes in galactic nuclei perturbations by fellow stars causing a fresh supply of stars to diffuse into orbits that take them near the black hole. Close to the black hole, the stars in a given volume have a nearly isotropic distribution of velocities except for an absence of objects with velocities that take them close enough to the black hole to be tidally disrupted. This void forms a loss cone towards the black hole. Close to the black hole, this loss cone is large, but it decreases steadily in size with increasing distance. The farther away from the black hole, the longer it takes a star to go from there to the black hole, so the more time that stellar perturbations can act to fill in the loss cone.

Beyond a certain distance from the black hole, the loss cone is so small and the time to reach the black hole is so long that perturbations by passing stars replenish the loss cone as fast as it is depleted. There is a steady supply of new stars approaching the black hole beyond this distance region. Inside this region, the loss cone is only filled if a star makes an unusually close approach to a fellow star, which perturbs one of the stars enough to deflect it into the black hole.

A similar thing happens to comets in orbit around the sun. Beyond a certain semimajor axis (about 20,000 astronomical units), there is a steady stream of comets that approach the sun. They constitute the classical Oort cloud of comets. Inside this semimajor axis, comets enter the inner solar system only after the close passage of a star to the sun, which refills the loss cone. This perturbation produces a shower of comets that is much more intense than the steady-state comet flux from the Oort comet cloud. This inner cloud is now sometimes called the Hills Cloud or the inner Oort Cloud (although Oort never proposed it). There is much greater danger of a comet hitting Earth during a comet shower.

The concept of comet showers attracted considerable attention in the scientific media. There were cartoons of a sky full of comets with dinosaurs watching it and wondering what it portended.

Even more media attention was given to the comet shower concept when Piet Hut at the Institute for Advanced Studies at Princeton University and Muller at the University of California, Berkeley, used it in their Nemesis theory to explain the apparent occurrence of periodic mass extinctions in the geological record of Earth. They invoked a companion of the sun, Nemesis, which periodically entered the inner comet cloud to produce a comet shower. The proposed companion had a mass less than that of a star but greater than that of a planet. Its orbital period, about 32 million years, would correspond to the interval between mass extinctions. Its orbit was assumed to be highly eccentric so that it passes well within the inner comet cloud at its closest approach to the sun, but the orbit is outside the inner cloud during most of its revolution around the sun.

There are problems with the Nemesis model. I showed that the orbit of Nemesis was large enough that it would likely have been torn from the sun by passing stars. I showed that, even if it had been able to survive, the orbital eccentricity and period would change greatly between subsequent orbits, which would not be compatible with a semiperiodic mass extinction. The Nemesis hypothesis has weathered away, but comet showers are now part of the astronomical lexicon and new papers appear each year that invoke them to explain other astrophysical puzzles.

A Cruel Blow to Earth Forms the Moon

A crucial ingredient in the Los Alamos work on the formation of the moon was the arrival of Willy Benz as a postdoctural candidate in T-Division. He did his Ph.D. thesis at the University of Geneva on star formation using an SPH code. I suggested that he use this code to study stellar collisions in three dimensions. These astronomical catastrophies are not common in the solar neighborhood, but they are in globular clusters and galactic nuclei. This work was starting to produce some interesting results when Harvard astrophysicist Al Cameron made one of his frequent visits to T-Division. I suggested to Willy that he show the results to Al Cameron.

Benz and I made the first simulations of stellar collisions in three dimensions, but Al Cameron had done such calculations in two dimensions. Al was interested in the results of the new simulations, but he was even more interested in the capability of the SPH code as demonstrated by the calculations. He suggested that Willy simulate a collision of a Mars-size object with Earth. This impact had been suggested as one of the scenarios for producing the moon, but no one had been able to simulate it. Willy made these simulations in collaboration with Wayne Slattery, a former student of Cameron, who was then in T-Division. The results were convincing enough that this hypothesis has now become the accepted one for the formation of the moon.

Benz and I produced three papers on stellar collisions. Melvyn Davies, Willy's Ph.D. student at Harvard, continued this work, extending it to stellar collisions involving encounters with binary stars. Several of Davies's Ph.D. students at Cambridge University came to T-Division during the summers to study stellar collisions. With the continued improvements in code (mostly SPH) and the discovery of objects in globular clusters that likely formed in stellar collisions, the study of stellar collisions has become a major astrophysical industry with entire conferences devoted it.

One of Willy's Ph.D. students from the University of Arizona, Chris Fryer, became a Feynman Fellow and later a staff member in T-Division. Willy Benz is now a professor at the University of Bern, Switzerland. Willy's SPH code was later adapted by various groups in the Los Alamos Applied Theoretical Physics Division.

Comets and Asteroids Impact Earth

In January, 1992, Los Alamos hosted a conference on the mitigation of asteroid-comet impacts with Earth. This conference was mandated by Congress. One result of this conference was a surge of work on this problem by Los Alamos. Some of this work was funded by Nuclear Weapons Technology Programs (NWT) because it was recognized that nuclear explosives could be used to deflect asteroids from Earth-impact orbits. Johndale Solem and I worked on these problems within the T-Division.

One of the most contentious arguments within the conference concerned the importance of smaller impactors.

A number of observers who were supported by the United States National Aeronautics and Space Administration (NASA) claimed that we should concern ourselves only with objects one kilometer (km) or larger in diameter. They argued that smaller objects would not cause global catastrophies, and only catastrophies of this amplitude were worth considering; although there were many more small impactors than large ones, so they hit Earth more frequently. Coincidentally, the Spaceguard Telescope that these observers were proposing to NASA would detect 90% of the Earth-crossing asteroids larger than 1 km in

diameter after a survey time of 25 yr. It would detect a much smaller fraction of the objects with diameters less than 1 km.

At the other extreme of the spectrum, colleagues at Livermore proposed that we concentrate our efforts on detecting objects about 10 meters (m) in diameter. They would use spears of tungsten to break them apart after they were detected by a constellation of space-based telescopes in Earth orbit.

This polarization of viewpoints in the research community led me and my undergraduate research assistant, Patrick Goda, to systematically quantify the amount and type of damage caused by asteroids of various sizes and compositions.

Goda and I published the results of our assessment in a 30-page paper in the *Astronomical Journal* in 1993. We determined the amount of damage from air blast, fire, cratering, earthquakes, and tsunamis as a function of asteroid size and composition. It was the first such systematic determination. The paper has been widely quoted since its publication. Goda and I were most alarmed by the great damage from tsunamis, which was largely unexpected. Even asteroids smaller than 1 km in diameter would produce tsunamis tens of meters high at thousands of kilometers from their impact points. The tsunami's height drops off only inversely with distance from the impact point because a tsunami is basically a two-dimensional disturbance. A tsunami is not significantly dissipated until it comes ashore, so a tsunami focuses much of the impact energy along the ocean shore, which has a disproportionate concentration of wealth. It was obvious that tsunamis generated by impacts should be studied much more carefully. The breakthrough came at a dinner party.

Just before our paper was sent in for publication, I attended a dinner party given by Art Cox of T-Division and his wife Joan. One of the participants at the party was Charles Mader, a Los Alamos Laboratory Fellow and retiree who was studying tsunamis at the University of Hawaii. He had developed an interest in tsunamis from his study of tsunamis generated by nuclear explosions. He gave me a number of literature references on tsunamis and agreed to collaborate on tsunamis generated by asteroid impacts. He had a computer code for determining the detailed run-up of tsunamis generated by earthquakes, which have about the same wavelength but much lower amplitudes than tsunamis generated by large asteroid impacts.

One of the first simulations involved the placement of a crater 150 km in diameter in the middle of the Atlantic Ocean. We expected this crater to be produced by an asteroid about 5 km across, which is about half the diameter of the one that killed the dinosaurs 65 million years ago. The wave from this crater swept all the way to the Appalachian Mountains. This work was published in several places and was partially the inspiration for Steven Spielburg's movie, *Deep Impact*. I was a technical consultant on this movie.

Goda and I improved the code used by Mader so it would run longer before encountering numerical difficulties. This work showed that Florida would be largely inundated by an impactor of this size hitting in the mid-Atlantic. It showed large inundations along the Iberian Peninsula with tsunamis more than 100 m high, and it showed deep penetration of tsunamis into the southern coast of France.

A similar study for the Pacific Ocean showed that the West Coast of the United States is particularly vulnerable to tsunamis as is Japan, particularly in the area around Tokyo. Much of this work was presented to the world tsunami experts at the Tsunami Society Meeting that was held in Honolulu in May 1999. It was well received, although it was very alarming to many of the tsunami experts.

Mitigation of Impacts

It was evident from our work that asteroids smaller than 1 km in diameter could do unacceptable damage. Can they be detected prior to Earth impact? My postdoctural fellow, Peter Leonard, and I addressed this question in a major paper published in 1995. We found that, except in the direction of the sun, even small asteroids such as the one that produced the Tunguska impact of 1908 could be detected several weeks prior to impact using telescopes with apertures less than 12. They do require a different detection algorithm than the ones currently used to detect those large Earth-crossing asteroids that are not yet on a collision course with Earth. The latter objects show large proper motion, a very high angular motion across the sky.

Objects that are a few weeks from impact are moving nearly straight at us, so they have little proper motion and may not be detected by existing surveys. Their small residual proper motion, which results from their orbiting the sun as they approach Earth, may be detected by comparing their current position to what it was on previous nights. The objects can also be detected by parallax, the slight displacement in their position relative to the background stars when they are observed simultaneously from several sites on Earth. The objects can be detected. Can we do anything to prevent their impact?

Johndale Solem of T-Division addressed the question of impact mitigation. He showed that nuclear explosives could be very effective in deflecting them before impact. A major trick is to avoid breaking them up in the process of deflecting them. Breaking them up may allow the impact of a number of pieces that are still large enough to cause unacceptable damage to Earth. Solem proposed using a strongly neutron-emitting explosive that is detonated about one asteroid radius away from the object. The explosion would flash off a layer about a meter thick from the hemisphere facing the explosive. This high-speed ejected material would produce a rocket effect that would propel the asteroid in the direction opposite to the position of the nuclear explosive. This process would provide much more control over the motion of the

asteroid compared to that gotten by setting off the explosive on the surface of the asteroid, but it would give the asteroid less kick for a given yield.

Mitigating against a comet impact is much harder than such mitigation for an asteroid impact. Comets can hit Earth with speeds exceeding 70 km/s while asteroids hit at an average speed of less than 20 km/s. It is very difficult to intercept comets far enough from Earth to provide sufficient leverage in distance so the small change in velocity imposed on the comet by a nuclear explosive provides enough accumulative displacement as it approaches Earth to prevent collision. It is very difficult to get this leverage in distance using chemical rockets. Their speed with respect to Earth is much less than that of the comet, so they intercept the comet relatively close to Earth. Solem has proposed a rocket that uses a series of small nuclear explosives rather than chemical burning. His rocket would attain a speed exceeding 100 km/s. A rocket of this type is the only feasible way of getting to a comet in time to prevent its impact with Earth. It would be an unmanned rocket, not the manned rocket based on Solem's design, that was featured in the movie, *Deep Impact*.

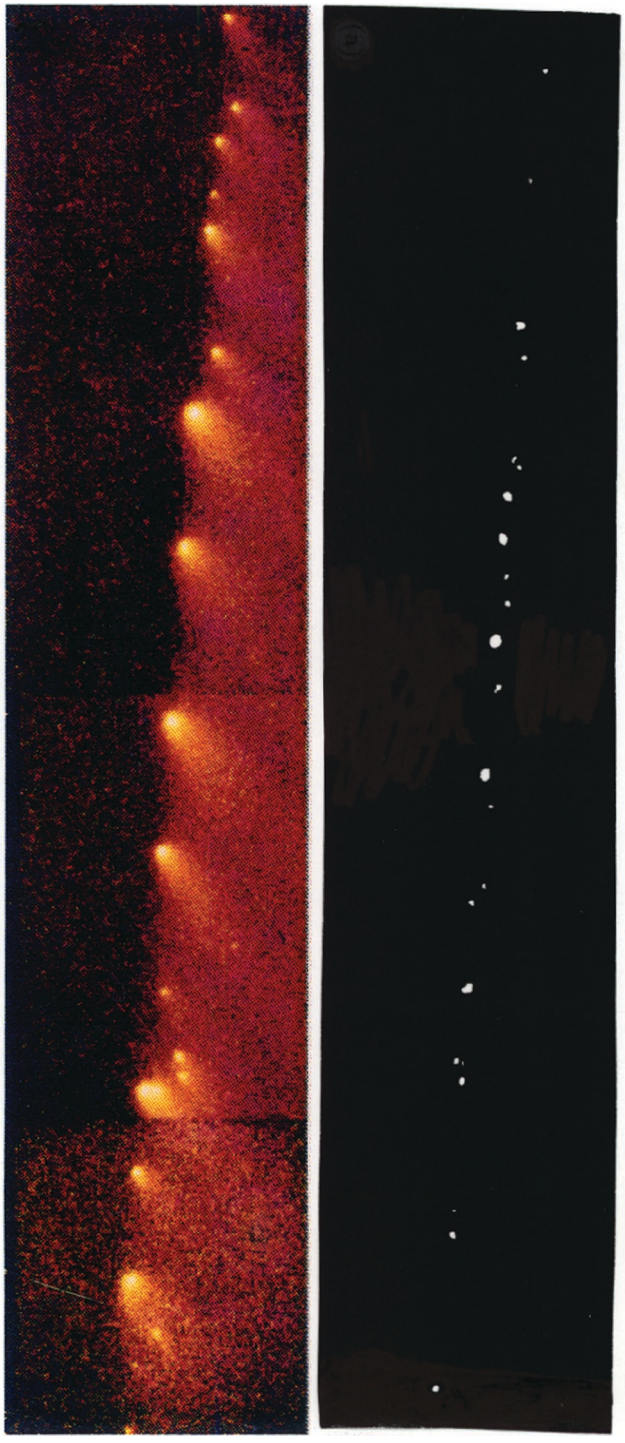
T-Division has shown the extent of the risk from asteroid-comet impacts and has shown ways to mitigate against this danger. We now need the political will to implement these plans.

Cometary Breakup Calculations Based on a Gravitationally Bound Agglomeration Model

by Johndale C. Solem

A Theoretical Division researcher calculated the size and the density of a comet captured by Jupiter, and the accuracy of these calculations was validated when the comet fell into Jupiter.

By modeling Comets as agglomerations of identical spherical components, "snowballs," bound together only by gravitation, I have obtained several qualitative and quantitative features of cometary breakup during encounters with strongly gravitating bodies. The model is contrary to the more widely accepted view of a comet as a single solid body, although the picture of comets and some chondritic asteroids as "flying rubble piles" has enjoyed increasing support from astronomers and astrophysicists alike. By studying the morphology of cases of cometary breakup, I can use the model to estimate both the comet's size and its density. A particularly interesting test case was Comet



(a) (b)

Fig. 1a. A calculation of the relative brightness and perspective of the fragments of the Comet Shoemaker-Levy 9, as they would be viewed from Earth. Figure 1b is an image of the comet from the Hubble Space Telescope.

Shoemaker-Levy 9 (1993e), which broke into a spectacular string of fragments owing to the tidal disruption as it passed closest to Jupiter on July 8, 1992. Figure 1a is a calculation of the relative brightness and perspective of the fragments as they would be viewed from Earth. It shows a total of 23 fragments with 7 or 8 major fragments, which is in good agreement with the general description given for Shoemaker-Levy 9. Figure 1b is an image of the comet from the Hubble Space Telescope. The pictures should be compared on the basis of the number, relative brightness, and relative position of the coalescent fragments.

No effort has been made to match the observation in detail, and the comparison is necessarily somewhat subjective. The best qualitative agreement occurs at a "snowball" density of 0.55 g/cm^{-3}, which corresponds to a comet average density of about 1/2 g/cm^{-3}, assuming a face-centered cubic (fcc) initial packing for the components. A much less dense comet breaks into fragments that do not gravitationally coalesce into anything resembling the observed string of objects. A much more dense comet may distort or even fracture near perijove but will quickly coalesce into a single object. Between these limits, the number of major coalescent fragments

(1) decreases rapidly with an increase in density,

(2) decreases only slightly with an increase in the assumed number of components, and

(3) is relatively insensitive to elasticity within the credible range of variation.

The breakup is sensitive to initial rotation of the comet, but the essential results remain the same as for a nonrotating comet, as long as the comet's angular velocity is less than 1/3 that required for centrifugal breakup in free space.

The model exhibits a remarkable scaling relation that allows a single calculation of the geometric and energetic features to be applied to a comet of any size. Consequently, the model can be used to estimate the size of the parent comet, which comes out to be about 1.8 km in diameter, assuming an initial 321-component fcc bumpy sphere. This diameter is roughly consistent with recent calculations based on the observed plume heights from the fragments after impact.

Shaping of Earth-Crossing Asteroids by Tidal Forces

by Johndale C. Solem and Jack G. Hills

Hills
Solem

Three near-Earth asteroids (NEAs)—Castalia, Toutatis, and Geographos have been of considerable interest to astronomers for sometime; and recently, images of the objects have been captured by Earth-based radar. All three exhibit a characteristic "potato" shape—in some quarters they have earned the sobriquet "spudnicks." We have undertaken an analytical study of the conjecture that such a shape may result from a close tidal encounter with Earth or other planets.

The population of Earth-crossing asteroids (ECAs) is in a nearly steady state—the number being removed is balanced by a similar number being added. The principal loss mechanisms for ECAs are

(1) impact with Earth, which is the fate of about 30% of ECAs; and

(2) perturbation by Earth into hyperbolic or Jupiter-crossing orbits.

For every object that hits Earth, another three objects pass within $1R_{\oplus}$ of Earth's surface without hitting it, so a sizable fraction of the current ECA population has passed near Earth ($R_{\oplus}$ = Earth's mean radius $\cong 6.378 \times 10^6$ m). This is particularly likely for the Atens, ECAs with semimajor axes less than 1 AU (AU = astronomical unit = mean radius of Earth's orbit $\cong 149.6 \times 10^9$ m). If the Atens originated in the asteroid belt, with semimajor axes of 2 to 3 AU, they should have made one or more close approaches to Earth in order for their semimajor axes to be reduced to their present values. These close passages expose ECAs to the tidal field of Earth, whose effect we have investigated.

Modeling Asteroids

We have addressed the tidal distortion problem by calculating a series of test encounters with Earth, modeling each asteroid as an assemblage of rocks that are bound together only by their mutual gravitation, which is contrary to the more traditional view of their being solid bodies. The picture of comets and asteroids as "rubble piles" has enjoyed increasing support, and comets with multiple nuclei are not uncommon.

Figure 1 shows a series of representative hyperbolic orbits. The trajectories are in the x-y plane with geocentric origin and initial asteroid velocity parallel to the x axis. They are found by integrating the Newtonian equations of motion, and they describe the time-dependent position of the asteroid's center of mass (COM). The trajectory for the last encounter passes about 250 km above the surface of Earth.

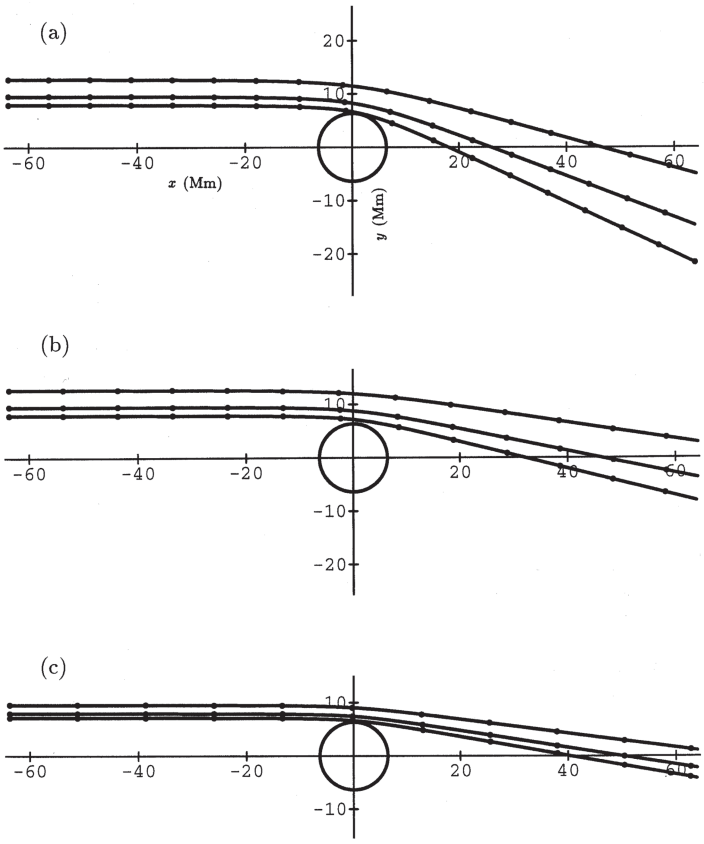


Fig. 1. A series of representative hyperbolic orbits.

We model the test asteroids as conglomerations of 135 identical, individually competent spherical rocks bound together only by their mutual gravitation. The actual asteroid components are not all spherical and they may have cohesive forces between them, but we assume these forces are much smaller than gravity. The collision of two components is treated as a nonadhesive frictionless scattering, i.e., the velocities are changed instantaneously in such a way that linear momentum is conserved, but some of the kinetic energy may be converted to heat. Because the spheres are frictionless, they receive no spin in a collision.

A further simplification is that radius r_0 and density ρ_0 of each component is the same, which endows the model with some remarkable scaling relationships. First, all distances scale with simple proportionality at a given time. Locations are described by the dimensionless vector r_i/r_0; e.g., if we increase the diameter of the asteroid by a factor of 2 and keep the same number of components, the geometrical arrangement of all components at anytime during the encounter will be exactly the same, but with the distance between them increased by a factor of 2. Second, a factor of 2 increase in the radius of each component (while keeping its density the same) increases all energies (kinetic energy, gravitational potential energy, and thermal energy generated in component collisions) by a factor of $2^5 = 32$. As a result of these scalings, we can treat asteroids of all sizes with a single calculation if all other encounter parameters are the same.

Effects of Tidally Deformed Asteroids

Figure 2 summarizes the final configuration geometry and state of an asteroid after passing by Earth for each of the hyperbolic orbits shown in Fig. 1. The maximum computed elongations exceed those required to explain the elongations found by radar. Geographos, which has the largest known elongation of 2.7, is difficult to explain as a collisional fragment. The mean elongation observed in fragmentation experiments is 1.4, and elongations as large as 2.7 occur less than 1% of the time.

Our analysis has shown deformations larger than those found for Geographos. However, our models of tidally deformed asteroids appear more symmetric than the observed asteroids (except perhaps for Geographos). This is not surprising since all of our model constituents have equal masses. If the constituents of the asteroid are of unequal size and shape, they might more closely resemble the observed objects. We conclude that the observed appearance of the elongated NEAs is consistent with their being formed by tidal deformation. The greatest elongation shown in Fig. 2 is for near-grazing incidence at 15 km s^{-1}.

We believe that if the velocity were a little less than this value, the object would be dissociated by the close encounter.

The stretching or pulling apart of an asteroid by the tidal field of Earth is slow but relentless compared to the near instantaneous impulse it suffers in a collision with a fellow asteroid. Rubble-pile asteroids that pass near Earth can be tidally pulled into elongated bimodal shapes that resemble the NEAs that have been imaged by radar.

The lower the velocity of an asteroid with respect to Earth, the more tidal deformation it suffers and the easier it is to break apart. This dependence of disruption on velocity may explain the overabundance of small asteroids (peaking at diameters of about 10 m) having low velocities with respect to Earth.

A tidal encounter gives the asteroid constituents a net angular momentum about the COM of the asteroid. If the asteroid is perturbed enough to be nearly pulled apart, this angular momentum may allow the fragments to orbit each other without coalescing into a single object. If they eventually coalesce into two objects that have sufficient angular momentum to orbit each other without collision, this orbit can be stable against tidal disruption by the sun. Binary asteroids formed by tidal encounters could explain the otherwise surprising propensity towards double craters on Earth.

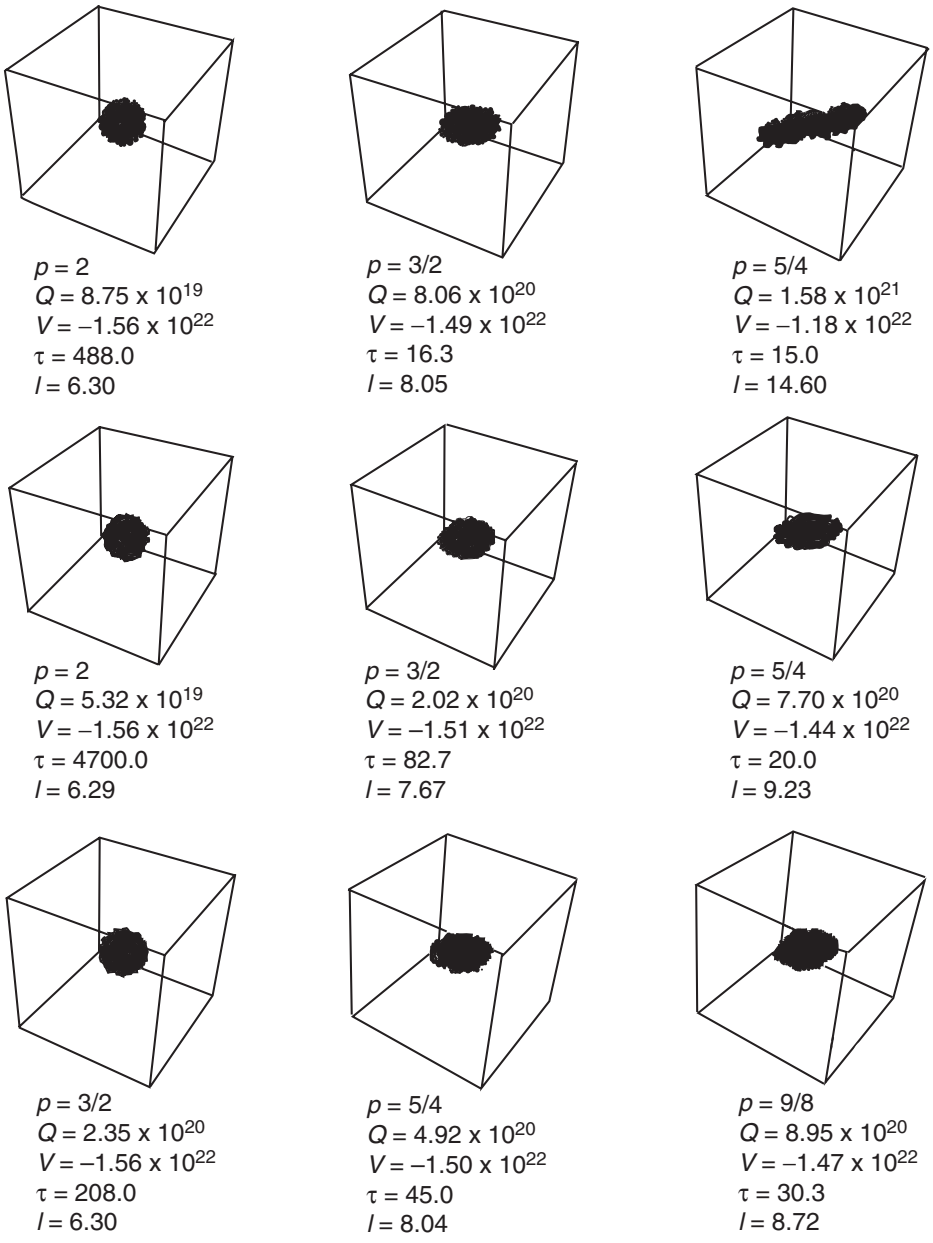


Fig. 2. The final configuration geometry and state of an asteroid after passing by Earth for each hyperbolic orbit in Fig. 1.

Unlocking the Secrets of the Rocky Planets

Baumgardner
Bunge
Frederickson
Harlow
Hsui
Reese
Richards
Ruppel
Solomatov
Stegman
Yang

by John R. Baumgardner

Theoretical Division researchers have applied supercomputers and cutting-edge numerical methods in computational fluid dynamics to explore the physical processes that have conspired to fashion the present state of the inner planets of our solar system, particularly our own planet Earth.

The inner planets of our solar system, namely, Mercury, Venus, Earth, and Mars, as well as most of the moons throughout the solar system, with a high level of certainty, have an interior structure composed of a mantle of silicate rock that surrounds a core mostly of iron, as illustrated for the case of the Earth in Fig. 1. Observations of the surfaces of these bodies reveal most have experienced an extremely dynamic and complex history involving significant melting and chemical differentiation of their silicate mantles. Both our own moon and Mars, for example, display a remarkable distribution of surface highlands composed of low-density rock in one hemisphere that is absent in the other. In the case of the Earth, the continents that cover about 40% of the planet are made of low-density rocks derived from dramatic multistage chemical differentiation involving partial melting, water, and vigorous thermal convection of the denser underlying mantle rock below. Such observations strongly suggest the earliest histories of these bodies involved large amounts of internal heat and vigorous convective overturn in their mantles.

Reconstructing the physical histories of these planetary bodies, from the observable present backward into the past, especially in the case of the Earth, although a formidable challenge, has long been deemed a grand challenge objective within the Earth and planetary science

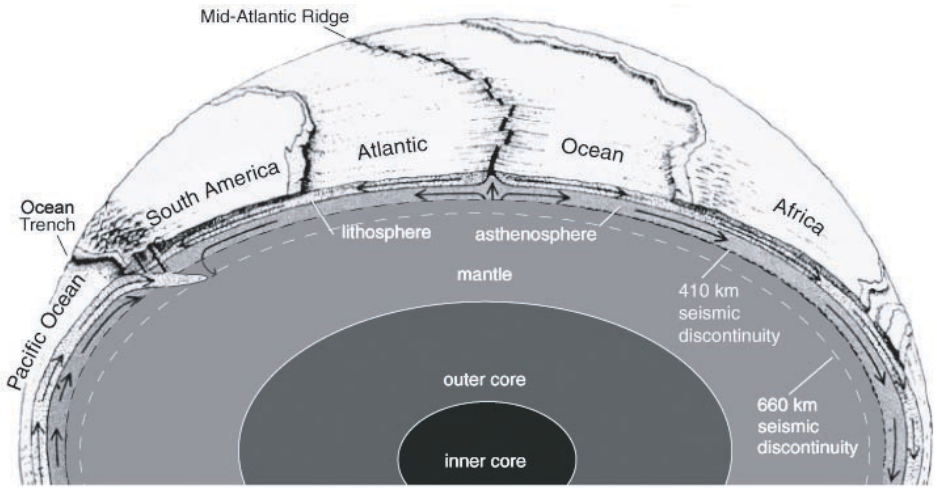


Fig. 1. Cross-sectional view through the Earth. The two major divisions of the planet are its mantle, made of silicate rock, and its core, composed mostly of iron. Portions of the surface covered with a low-density layer of continental crust represent the continents. Lithospheric plates at the surface, which include the crust and part of the upper mantle, move laterally over the asthenosphere. The asthenosphere is hot and also weak, probably because of the presence of water within its constituent minerals. Oceanic lithosphere, lacking the layer of continental crust, is chemically similar on average to the underlying mantle. Because oceanic lithosphere is substantially cooler, its density is higher, and it therefore has an ability to sink into the mantle below. The sliding of an oceanic plate into the mantle is known as subduction, as shown here beneath South America. As two plates pull apart at a midocean ridge, material from the asthenosphere rises to fill the gap, and some of this material melts to produce basaltic lava to form new oceanic crust on the ocean floor. The continental regions do not participate in the subduction process because of the buoyancy of the continental crust.

communities. Of particular interest has been obtaining a model that is able from first principles account for the large-scale motions of the Earth's tectonic plates as illustrated in Fig. 2.

Characterizing the Main Physical Processes

The prime strategy for achieving such a goal involves, first, obtaining accurate mathematical characterizations of the main physical processes, especially of the deformational properties of silicate materials, valid over a wide range of temperatures and stress regimes. Experimental measurement of these properties has been conducted in many laboratories around the world, including Los Alamos, for the past several decades. The second part of the strategy involves incorporating these mathematical characterizations of microscale processes into numerical models that can apply and integrate these processes to the planetary scale.

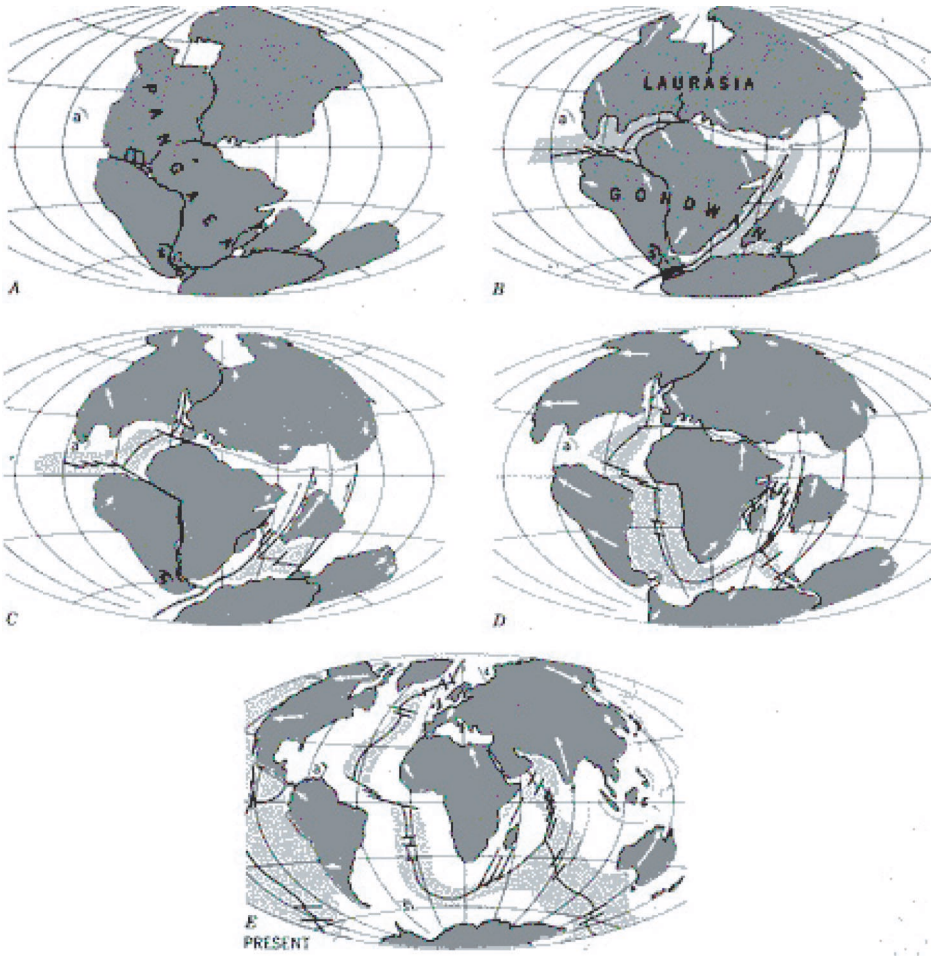


Fig. 2. Five snapshots in time of the Earth's continent locations from the early Mesozoic Era to the present.

A graduate student in the Earth and Space Sciences Department at the University of California at Los Angeles, in the early 1980s, I, who had experience in applying supercomputers to difficult problems, caught the vision of realizing the second part of this strategy. With prior experience in the U.S. Air Force in applying supercomputers to difficult problems, as a Ph.D. Student in the Earth and Space Sciences Department at the University of California at Los Angeles in the early 1980s, I caught a vision of realizing the second part of this strategy. Upon learning of the creation of a branch of the Institute of Geophysics and Planetary Physics (IGPP) of the University of California at Los Alamos National Laboratory in late 1980, I immediately wrote a proposal to develop a three-dimensional (3-D) spherical shell finite element model for treating planetary mantle dynamics, utilizing the unique computing resources available at Los Alamos. This proposal was funded with startup money provided for this new IGPP branch.

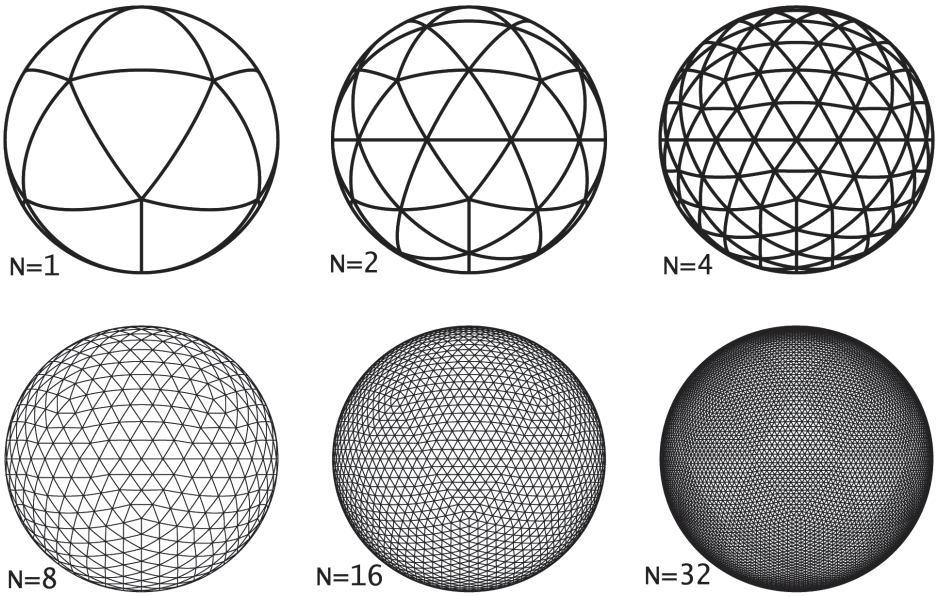


Fig. 3. Hierarchy of geodesic grids constructed by first projecting the regular icosahedron onto the sphere to divide the spherical surface into 20 equal spherical triangles and then connecting the midpoints of the triangle sides with great circle arcs to generate four times the number of spherical triangles in each successive refinement. Here *N* is the number of subdivisions of the triangle sides of the coarsest grid.

In the fall of 1981, I came to Los Alamos to collaborate with Paul Frederickson, an applied mathematician in the Computation Division, on this project. Paul had devised a powerful method for solving large systems of coupled partial differential equations in a remarkably efficient manner, a method known as multigrid. This technique was ideally suited to my application.

Paul, from previous work at the U.S. National Aeronautics and Space Administration (NASA) also offered the suggestion of using an almost uniform geodesic grid constructed from the regular icosahedron for discretizing the sphere as illustrated in Fig. 3. This type of grid avoids the singular points at the poles characteristic of the latitude-longitude grid commonly used for spherical problems. Paul's grid also provides the hierarchy of grids from coarse to fine, with points of all the coarser grids contained in the finer ones, required by the multigrid method.

TERRA

Working together closely, Paul and I, within about 8 months, assembled a working 3-D code for solving the equations that govern the creeping flow in a convecting planetary mantle in spherical geometry, a code that has since become known as TERRA. Using this code, I obtained the first fully 3-D solutions for creeping flow convection in spherical shell geometry. This work constituted the core of my Ph.D. thesis, which I completed in late 1983. Aware of this work, Hans Ruppel and Frank Harlow in T-Division interviewed me and offered me a staff position in T-Division. I accepted the offer and have been part of T-Division's technical staff since early 1984.

The T-Division environment encourages university collaborations, and I have developed several fruitful collaborative relationships in the area of planetary physics. The most successful have been with Prof. Mark Richards of the University of California, Berkeley (UC Berkeley); Prof. Albert Hsui of the University of Illinois at Champaign-Urbana; and Prof. Slava Solomatov of New Mexico State University. Funding from the Los Alamos branch of the IGPP has provided support for several graduate students to utilize the TERRA code for their Ph.D. thesis work and spend time at Los Alamos to collaborate more closely with me.

A student of Prof. Mark Richards, Hans-Peter Bunge, supported by an IGPP grant, came to Los Alamos in 1990 and modified TERRA to run efficiently on the newly emerging class of massively parallel computers consisting of hundreds of processors. With access to such powerful computing platforms, Peter proceeded to apply TERRA to several important scientific problems. He demonstrated that the radial viscosity stratification of a planet's mantle plays a dominant role in determining the lateral convection pattern, as suggested in Fig. 4. He showed that a postulated jump in viscosity at a depth of about 700 km in the Earth, by itself, explains most of long wavelength character of the Earth's style of mantle convection. [1] He also demonstrated, using TERRA, the internal density profile can account for most of the density structure presently observed in the Earth's mantle today by seismic tomography methods. [2] This internal density profile is generated by imposing the plate motions inferred from seafloor spreading data as surface velocity boundary conditions since mid-Cretaceous time. After completing his Ph.D. in 1996, Peter took a postdoctoral position at the Institut de Physique du Globe de Paris and then a faculty position at Princeton University, where he continues to apply TERRA, through his own graduate students, to cutting-edge problems.

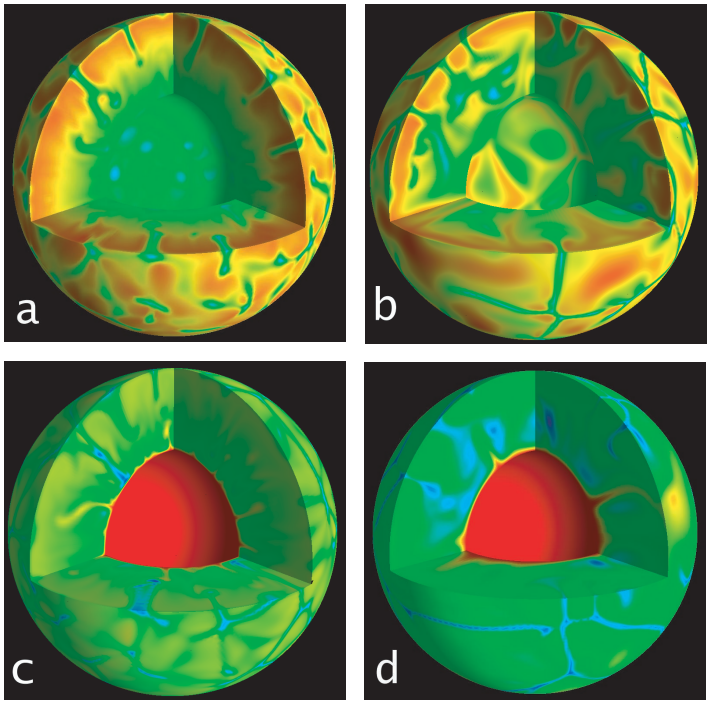


Fig. 4. Temperature fields for convection in a spherical shell with geometry of the Earth's mantle. Blue is cold and red is hot. The uppermost 200 km of the shell, representing the upper thermal boundary layer, has been removed to show temperatures beneath the boundary layer more clearly. The shell is heated by a uniform distribution of radioactive elements. In (a) and (b), the inner boundary is insulating, while in (c) and (d) the inner boundary has a fixed uniform temperature so that about 35% of the total heating is from heat conducted through this lower boundary. The shell has a uniform viscosity in (a) and (c), while in (b) and (d), viscosity increases by a factor of 30 below 700 km. Point-like downwellings characterize the uniform viscosity case, but linear sheet-like downwellings emerge in the case with stratified viscosity. Simply making the lower mantle more viscous, as supported by many lines of observational evidence for the Earth, results near the surface in this strikingly more Earth-like pattern of linear downwellings that correlate with observed subduction at plate edges. For the Earth, most estimates for the fraction of mantle heating from the lower boundary are in the range of 10%–30%.

Earth

In a similar manner a student of Prof. Albert Hsui, Woo-Sun Yang, also supported by IGPP funding and working much of his time at Los Alamos, made major improvements in TERRA's capabilities. Woo-Sun's focus was on finding a numerical method that would permit TERRA to handle extreme local variations in rock strength, or viscosity. Such extreme variations occur, for example, at edges of tectonic plates and also at the bottom boundary of these plates. Therefore, to capture important features of the dynamical behavior of plates on the Earth, the ability of TERRA to handle such strong variations in rock strength is essential.

After a diligent search of the literature on multigrid methods, Woo-Sun identified an approach that proved extremely effective in the TERRA context. [3] This new capability enabled TERRA to model the phenomenon of plates in spherical geometry from first-principle physics for the very first time. (See Fig. 5.) With the development of the new capability as the heart of his thesis, Woo-Sun completed his Ph.D. in 1997 and

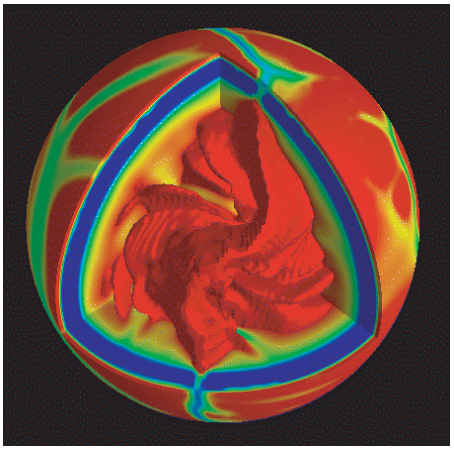


Fig. 5. Cutout view of a 3-D viscosity field from a case in which high-viscosity (red) plate-like patches occur at the outer shell boundary. This plate-like character appears when a zone of low viscosity (blue) is specified just below the top thermal boundary layer. In the Earth such a zone of reduced viscosity exists in this region because of the presence of water. The contorted red isosurface in the deeper portion of the shell corresponds to the high-viscosity cold material sinking from the surface where two plates are converging. Such a convergent region on the Earth is known as a subduction zone.

began a two-year postdoctural research position at Los Alamos funded by a NASA high-performance computing and communication (HPCC) Earth and space science grand challenge cooperative agreement. Woo-Sun's work had played a key role in our team's winning this $1.3 million award.

One of the impressive results of Woo-Sun's postdoctoral research was his demonstration of the critical role a zone of low viscosity beneath the tectonic plates plays in making a plate tectonics style of mantle convection possible. [4] The Earth is the only planet in the solar system with plate tectonics, and the reason is almost certainly the presence of liquid water at its surface. Water dramatically reduces rock strength, and its presence in the lattices of the minerals in the zone just beneath the plates leads to the remarkable weakness of this layer and hence to plate tectonics. Woo-Sun currently has a research position at Lawrence Berkeley National Laboratory.

Mars

Supported by an award from the New Mexico Universities Collaborative Research (NUCOR) initiative, Chris Reese, a student of Prof. Slava Solomatov at New Mexico State University, applied TERRA to investigating the physical history of Mars. Although small relative to Earth and Venus, Mars has the largest volcanic mountain in the solar system, Olympus Mons. An as yet unresolved question is what set of processes in the Mars mantle could have caused such persistent and localized volcanic activity over a large fraction of this planet's history. Chris explored the creative possibility that Olympus Mons is the expression of a large asteroid impact from moderately early in Martian history. The basic thesis is that because of its small size, the Martian mantle cooled much more quickly than the mantles of either Earth or Venus; and, when the impact occurred, convection in the Martian mantle was relatively sluggish and so the heat from the impact has persisted and remained relatively localized as suggested in Fig. 6. [5] TERRA's flexibility allowed Chris to explore this and several other potential explanations for this major feature on the red planet. Chris completed his Ph.D. in 2000 and is now on the faculty of the University of Minnesota.

The Moon

Another student of Prof. Mark Richards at UC Berkeley, Dave Stegman, has extended TERRA's capabilities in yet another direction. Supported by Los Alamos IGPP funding, Dave has added a particle treatment to TERRA to track chemical species and model chemical mixing and differentiation. Many lines of evidence indicate that, especially in their early histories, most of the rocky bodies in the solar system underwent significant melting and chemical differentiation. Surface features on many of these bodies today reflect these early processes.

Dave is currently applying this new capability for tracking chemical species in a convecting mantle to understand the early history of the Moon. Rock samples from the U.S. Apollo missions provided compelling evidence for massive melting of the lunar mantle and the differentiation of lighter rock that now forms the lunar highlands on the far side of the Moon. Dave is exploring the scenario that the heavier minerals from an early magma ocean settled downward to form a dense layer at the bottom. Based on known chemical affinities, a large fraction of the radioactive elements in the molten zone would have segregated into this dense layer. Because the dense layer was gravitationally unstable relative to the mantle below it, the dense layer rather quickly sank to the bottom and enveloped the small lunar core. Because of its high concentration of radioactive elements, this layer both heated itself and also the core, eventually reaching a temperature at which it had sufficient thermal buoyancy to rise in a plume-like manner back toward the surface of the lunar mantle. Its high temperature resulted in partial melting and generation of basalt that flooded the low-elevation maria on the Moon's front side. Removal of this "thermal blanket" allowed a net outward flux of heat from the lunar core that in turn was able to sustain dynamo action and generate a dipole magnetic field for a brief period. [6] The cooling basaltic rocks at the lunar surface, samples of which were collected by the Apollo astronauts, in turn recorded this magnetic field. TERRA's new particle tracking capabilities have allowed Dave to perform just this sort of numerical simulation. Dave is to complete his Ph.D. in 2003.

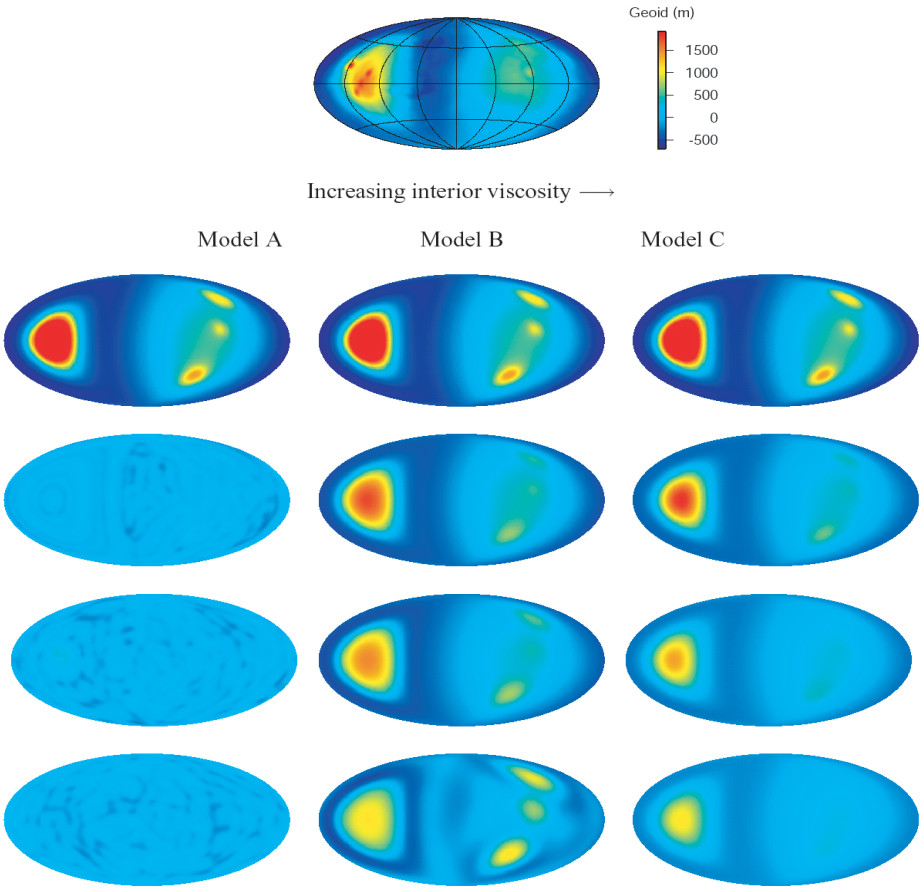


Fig. 6. The present-day Mars gravitational potential field (rows, top to bottom) as measured by the Mars Global Surveyor satellite (top) and the time history of the gravitational potential at 0.0, 1.5, 3.0, and 4.5 gigayears (Gyr) from three computational models for the Mars mantle. In a low-viscosity model (A), initial temperature differences are quickly eliminated by vigorous convection. For an intermediate mantle viscosity (B), convection is confined to impact-heated regions. Sufficiently high mantle viscosity (C) ensures a purely conductive thermal history and gives the best agreement with observation.

Additional Studies

Currently, TERRA is able, starting with relatively simple and plausible initial conditions, to model the breakup of the supercontinent Pangea and subsequent dispersal of the resulting continent blocks as illustrated in Fig. 7. I am eager to enhance TERRA's capabilities yet further.

The recently added particle treatment now makes possible the inclusion of rock models with history variables that can treat elastic failure and faults in the surface layer to allow faults to form, heal, and reactivate.

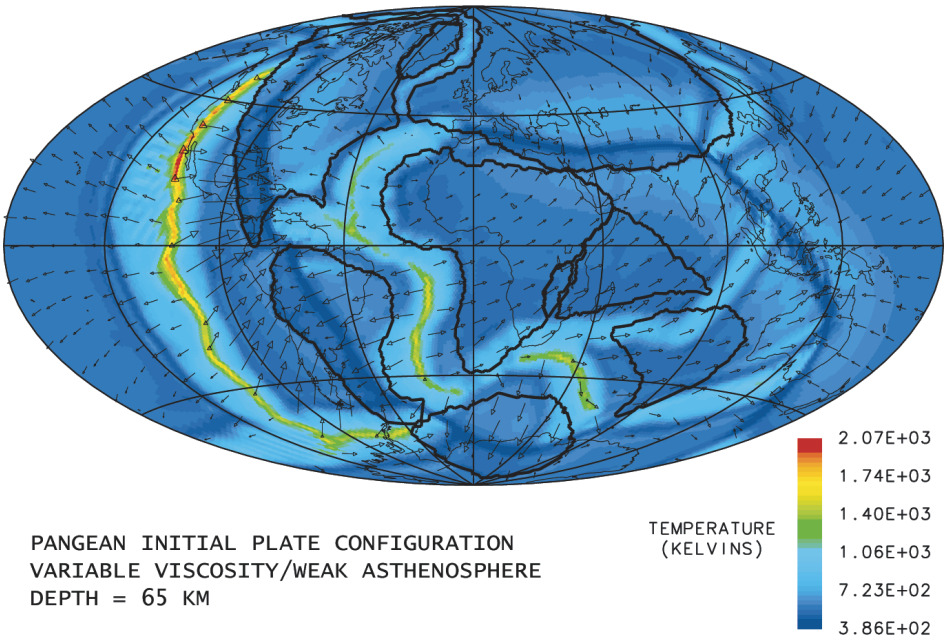


Fig. 7. Plot of temperatures, velocities, and continent boundaries from a TERRA calculation initialized with a Pangean plate configuration. Plate motions across the surface are tracked using particles.

The ability to treat faults in such a realistic manner opens the door to models in which details of surface geological processes such as mountain building, erosion, and sedimentation can be included. Such comprehensive models that link internal geophysical processes with surface geological processes offer the prospect of a much closer linkage between the fields of geophysics and geology. This would mean that the huge datasets associated with the Earth's geological history at the planet's surface could be applied to constrain models of the geophysical processes that have operated inside the Earth over time to drive the processes responsible for that surface record. In other words, global numerical models of the solid Earth are likely still in their infancy relative to what now seems possible.

One problem of particular interest to me involves understanding the regime of intermittent mantle instability identified in Woo-Sun's study of the effects of water in producing a low-viscosity zone in the upper part of the Earth's mantle. I suspect the mantle dynamics of the Earth may have been more unstable and violent than currently thought; and such episodes of instability may account for prominent features of the geological record that are as yet unexplained, such as the large thickness of sediments on the continents, the vast lateral extent of many of the Paleozoic and Mesozoic strata, and widespread evidence of very rapid sediment deposition. It seems likely that many important and exciting secrets of the Earth and the other rocky planets remain yet to be discovered.

On the Possibility of Extracting Energy from Gravitational Systems by Navigating Space Vehicles

by Stanislaw M. Ulam

This paper was originally published as Los Alamos Scientific Laboratory report LAMS 2219 (June 1958).

Everett
Ford
Ulam

In the 1950s, Theoretical Division scientists explored the possibility that space-traveling rockets could gain energy from the gravity of individual planets they came near. The findings impacted space exploration.

It is intended to outline in this brief report a number of problems of the following type: we assume an astronomical system composed of two or more stellar bodies and a space vehicle which, as an additional body of infinitely small mass compared to the celestial objects, forms part of a many- (e.g., 3-) body system. We assume that the "rocket" not only describes the trajectory under the action of the gravitational forces, but also that it has still a reserve energy available for steering by suitably emitted impulses. This energy in the discussion below will be assumed to be roughly of the order of the kinetic energy that the rocket already possesses. The problem, broadly speaking, involves the possibility of using this reserve energy in such a way as to acquire, by suitable near collisions with one or the other of the celestial bodies, much more kinetic energy than it possesses—more by an order of magnitude than the available reserve energy would allow it to acquire by itself.

As examples of the situation we have in mind; assume a rocket cruising between the sun and Jupiter, i.e., in an orbit approximately that of Mars, with an energy in reserve which would allow the kinetic energy of the vehicle to increase by a factor such as 2. The question is whether, by planning suitable approaches to Jupiter and then closer approaches to the sun, a rocket could acquire, say, 10 times more energy. Another example would be a space vehicle moving in a double-star system "half-way in between." Then the question is whether, by using additional impulses of its own, the rocket could acquire again a kinetic energy much greater than it already possesses.

As a purely mathematical problem we could consider the case of two mass points, each of mass 1 forming a Keplerian system, and a rocket of mass vanishingly small compared to 1 in an orbit which forms a curve between the two mass points. Suppose that the reserve power of the rocket is such that it could double its kinetic energy. Question: Can one, in this idealized condition, obtain a velocity arbitrarily large (i.e., close to light velocity)?

That this possibility exists seems extremely probable from the theorems on ergodic transformations. [1] It has been shown that arbitrarily near to any given transformation, like the one given by the Hamiltonian describing the *n*-body system above, there exist transformations which are metrically transitive, that is to say, in particular, Liouville flows such that the trajectory of the system will penetrate arbitrarily near any point on the phase space. The theorem has been proved for bounded phase spaces. This does not make our theorem inapplicable to the problem. We could put in cutoffs in the distance of approach and assume at a finite but very great distance from the gravitational bodies another cutoff. The theorem would imply that arbitrarily near the given dynamical motion there exists one which will make the rocket approach as close to the cutoff sphere surrounding any one of the given mass points as we please, which would in particular imply obtaining arbitrarily high velocities. The theorem asserts the existence of such motions arbitrarily near given ones. The question whether these can be obtained by changes effected through emitting additional impulses inside the rocket is not essentially answered, but, in view of the prevalence of ergodic motions near a given one, this seems extremely likely. Such an ergodic trajectory would, of course, in particular, provide arbitrarily high velocities. Nothing is said, however, about the times necessary for effecting this. They might be of super-astronomical lengths. It is clear, on the other hand, on general thermodynamical grounds that "in general" the equipartition of energy may take place. This implies that the body with the small mass of the rocket will acquire very high velocities. This is well known even in systems of a moderate number of particles. The energy distribution is Maxwellian, again tending to provide the small masses with high velocity. The problem is whether, by steering the rocket, one can to some modest extent, acquire the properties of a Maxwell demon, i.e., plan the changes in the trajectory in such a way as to shorten by many orders of magnitude the time necessary for acquisition of very high velocities.

As is well known, the perturbations of Jupiter on the motion of some comets provide them occasionally with velocities of escape from the solar system. It has been noticed also [2] that one can use the attraction of the moon to provide a rocket with additional kinetic energy, enabling it to escape from the Earth's gravitational field even if it did not have enough energy to do that to begin with.

Our problem is whether one can do it repeatedly to obtain essentially arbitrary kinetic energies by repeated and suitably timed approaches to the two or more celestial bodies.

The question is that of finding general recipes for a 3- or more) body problem to achieve that aim as quickly as possible. We propose to calculate some very schematic, simple, but perhaps instructive cases for a "strategy" of steering the rocket.

(1) The first case involves a problem in one dimension. Suppose two masses oscillate at the end of a segment with given amplitude, say harmonically; a point of vanishingly small mass is rocketed in between and, possessing some initial kinetic energy, collides elastically with the two oscillating end points. If these should be in phase, the calculation will show the increase of kinetic energy of the small mass. If the phases of the two oscillators should be randomly independent, the question arises how to plan the emission of additional small impulses by the middle point so as to make it increase the rocket's kinetic energy most efficiently. Obviously, one should plan to collide head on with the two oscillators as much as possible. In other words, through additional impulses, collisions that lead to a gain in energy for the "free" point should be maximized as far as possible. The ones that involve colliding by overtaking the receding end point should be diminished in their effect. Without a strategy of changes in the velocity of the "rocket," the gain of energy towards an eventual equipartition would be a very slow process—the rates at which this approach to equipartition takes place are unknown in statistical mechanics but certainly the gains in a random process increase with the square root of time or slower. With an operating intelligence perhaps this approach to near-equilibrium could be made vastly more rapid.

(2) This problem will involve two mass points describing quite elongated Keplerian ellipses around their center of mass. The rocket moves initially in a roughly circular orbit in between the two masses around their common center. Of course, the actual trajectory in this 3-body problem is very complicated. The question is again to plan a strategy of changing, by small amounts, the energy of the small object so as to approach one or the other of the large bodies to gain kinetic energy. If this is to take place, the approach to either of the two bodies must be increasingly closer. This involves great elongation and an increase in apastron of the rocket. Again the plan is to make near collisions head on. Presumably, the planned changes, that is to say the emitted impulses from the rocket, will be most efficient when the body is at maximum distance from the center of gravity of the two celestial points. It is there that a small increase in velocity will enable one to make changes in the time of the next approach.

The above discussion is, of course, intended for a purely theoretical mathematical question. Even so, during the next few decades, large objects may be constructed with a cruising velocity of 20 kilometers a second, and there will be still some additional energy left for changes in this velocity. It is obvious that the process of increasing the kinetic energy of the rocket by such extraction of gravitational energy from celestial motions is, at best, very slow. The computations required to plan changes in the trajectory might be of prohibitive length and complication. This little note is meant merely as an introduction to exploratory analyses and calculations undertaken with Kenneth W. Ford and C. J. Everett of LASL.

Two- and Three-Dimensional Supernova Calculations or Why a Star Explodes Nonsymmetrically

Benz
Bethe
Colgate
Epstein
Fryer
Herant
M. H. Johnson
Schramm
Tubbs
M. Warren
White
J. Wilson

by Stirling A. Colgate, Marc B. Herant, Will F. Benz, and Christopher L. Fryer

When giant stars die, they create spectacular explosions or supernovae. Scores of researchers, including many from Los Alamos, have constructed detailed theories of how and why these super-giant stars explode. Theoretical Division staff played a critical role in developing these theories by performing multidimensional calculations of supernova explosions. These calculations are now showing, for the first time, how these explosions really work.

Neutrinos and Supernovae

Although neutrinos clearly play a key role in nuclear physics through the weak interactions, their interactions with matter are just that: weak.

But there is an exception to that role. It occurs in the heart of massive stars, deep within the stellar core. When giant, heavy stars die, they exit spectacularly in the most powerful explosions, supernovae, known to occur within individual galaxies. (Only the formation of a galaxy, its central massive black hole, and the "big bang" of the universe are larger.) The supernova explosions depend upon neutrinos to transport the energy creating the explosion.

Over the years, scores of researchers, including many from Los Alamos, have had a particular interest in large explosions and have constructed detailed theories of how and why these supergiant stars explode. Theoretical Division has played a critical role in the development of these theories by being host to the first multidimensional calculations (two-dimensional [2-D] and now, 10 yr later, three-dimensional [3-D]) of supernova explosions. These calculations have now shown for the first time how these explosions really work.

More dramatic even than the explosion of the star itself is the remnant left behind, a neutron star, an object so tightly bound by gravity and so dense that its composition still baffles or, more politely, leads to speculation as to its composition. The greatest enigma of the whole event is that the available energy of the neutron star, its binding energy, is 300 times greater than the explosion energy itself. How is it that such a tiny fraction of the available energy makes its way reliably to create such a stupenduous explosion, 10^{51} ergs, a "foe" of energy?

We know that all stars emit light and shine because they "burn" fuel. But once a massive star has exhausted its fuel, its center—the stellar core—begins to collapse and fall in on itself. The resulting compression of the matter within the core is so severe and the collapse occurs so quickly that in less than 1 s the core reaches conditions of temperature and density virtually unparalleled in the physical universe. Under these

unique circumstances, theoretical physicists predict that enormous quantities of hot neutrinos are produced in the core; but, contrary to our normal expectations, those neutrinos become trapped in the ultradense matter and must slowly, over several seconds, diffuse to the surface. A fraction of those neutrinos are absorbed by material that is plummeting toward the compacted core; and, by a mechanism that will be described in this article, this infall becomes very hot, expands, and surges outward. Eventually, the star erupts in an explosion that violently ejects the star's outer layers into space. All that remains behind of the once enormous star is a tiny, incredibly compact, and remarkably novel object called a neutron star.

This pivotal function of the neutrino, so much in contrast with its usual marginal position, received triumphant vindication in February 1987, when a burst of neutrinos were detected in huge, underground detectors in coincidence with optical observation of a nearby supernova. That fortuitous event altered the course of supernova astronomy. For the first time in history, supernova theory could be tested against detailed observation and hard data.

That is not to say that prior to 1987 mankind had no information concerning supernovae. Far from it. Supernovae have been marveled at since early human history, calling attention to themselves as stars that suddenly appear in the sky where none were before. Such an exceptional star may remain visible for a month or more. Chinese astronomers recorded one such event as early as 185 A.D. But such sightings are rare, for supernovae are infrequent events that occur on average only once every 50 yr or so within a given galaxy. The inhabitants of the northern hemisphere have not been treated with a supernova visible to the naked eye since 1604.

On the other hand, there are billions of distant galaxies within the universe, and supernovae tend to be highly conspicuous. So much energy is released during the explosion that for a brief time the star may outshine the entire galaxy in which it resides. Thus, in the course of observing the sky with telescopes over the last 100 yr, astronomers have been able to monitor thousands of supernovae. They have been able to examine in great detail the expanding nebulae that linger for centuries as remnants of the explosions (Fig. 1).

The study of supernovae has advanced to the point that we can classify the explosions into two major catagories. Type I supernovae have no hydrogen in their emission spectrum and are generated (usually) by white dwarfs, which are old stars of small mass. In this article, we will concentrate on the more common Type II supernovae. These have for progenitors massive stars that are more than eight times the mass of the sun.

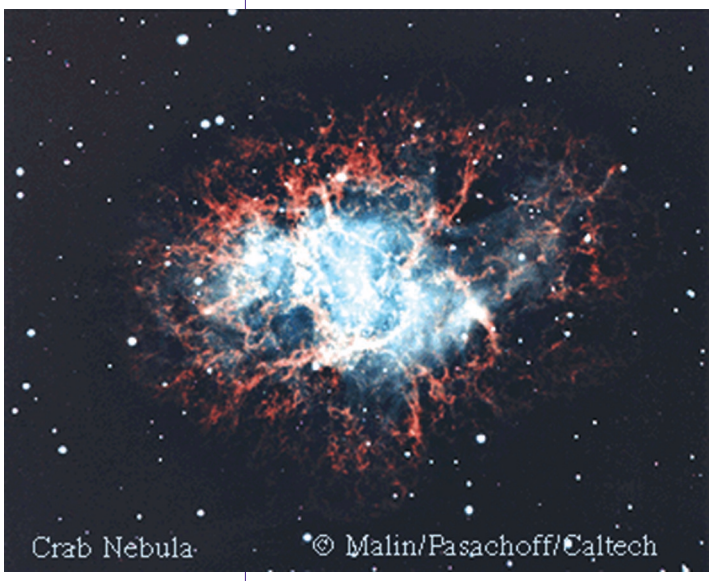


Fig. 1. The Crab Nebula. Located about 6,500 light years from Earth, the crab nebula is an expanding gaseous cloud that was hurled into space when a giant star exploded. That supernova was visible day and night for several weeks in July 1054. Even today, the visible emission from the nebula is still greater than 75,000 suns. At the center of this brilliantly glowing cloud lies a spinning neutron star, which is the core of the original star. (Copyright © Malin/Pasachoff/ CalTech, color photograph by David Malin)

The Evolution of a Star to the Point of Collapse

A star continually radiates energy away into space in the form of light, and is thus inexoribly cooling down. Nuclear fusion reactions liberate energy as light elements are fused into heavier ones. In the process, the light elements are used up, or "burned," while the heavier ones accumulate. All nuclear burning takes place in the core, which is the hottest and densest part of the star. Initially, in a multistep process, hydrogen is burned and fused into helium. Sooner or later (approximately 20 million yr for a massive star), all of the hydrogen at the core has been burned and fused into the next heavier element, helium. But while the core conditions were correct for accommodating hydrogen burning, helium needs a higher temperature and density in order to burn. Thus, fusion stops. With its energy source turned off, the core contracts under gravity with a rise in the core temperature and density. Conditions quickly surpass what is necessary to initiate the fusing of helium into carbon, and new energy is released. The star again restores itself as it begins a "helium burning" stage. The large star still consists almost entirely of hydrogen, but it now has a helium core that is burning and slowly turning into carbon.

Eventually, the helium fuel is exhausted, Once again, fusion halts and the star contracts until it is able to initiate the burning of carbon into still heavier elements, the alpha particle nuclei during stable burning progressing from carbon, to oxygen, then neon, and, finally, silicon. However, as the elements get heavier, the amount of energy released per reaction gets less. Therefore, the rate of burning must increase in order to produce enough heat to sustain the internal pressure. Incredibly, instead of millions of years, silicon burning lasts but 1 day. Because of the various burning stages, the star develops a layered structure consisting of many different elements, as seen in Fig. 2.

When a star of mass greater than approximately 10 solar masses (M) approaches the end of silicon burning, a process of adding alpha particles (i.e., helium nuclei), it reaches the most stable alpha-particle nuclei, nickel 56 or 14 alpha particles. This nucleus decays by beta and neutrino emission in a weird fashion, first to cobalt-56 in 6 days and a few million electron volts (MeV) of energy, and then to iron-56 in 72 days by means of a triply forbidden transition and releasing 6 MeV of energy. This long decay with such large energy is why supernova are visible to humans, as we discuss later in this article. Thus the core begins to accumulate iron that cannot "burn" and release further energy. Since iron is the most stable of all nuclei, both fusing it and fissioning it absorbs rather than releases energy. As a result, the star cannot gain further energy from nuclear reactions, and compressing the core will not eventually lead to a new burning phase. This is the beginning of the end of the star's life. As the core proceeds to cool, the omnipresent force of gravity will cause the core to begin collapsing. The events that follow happen very fast; and, remarkably, the end comes in less than a second, or literally within the blink of an eye.

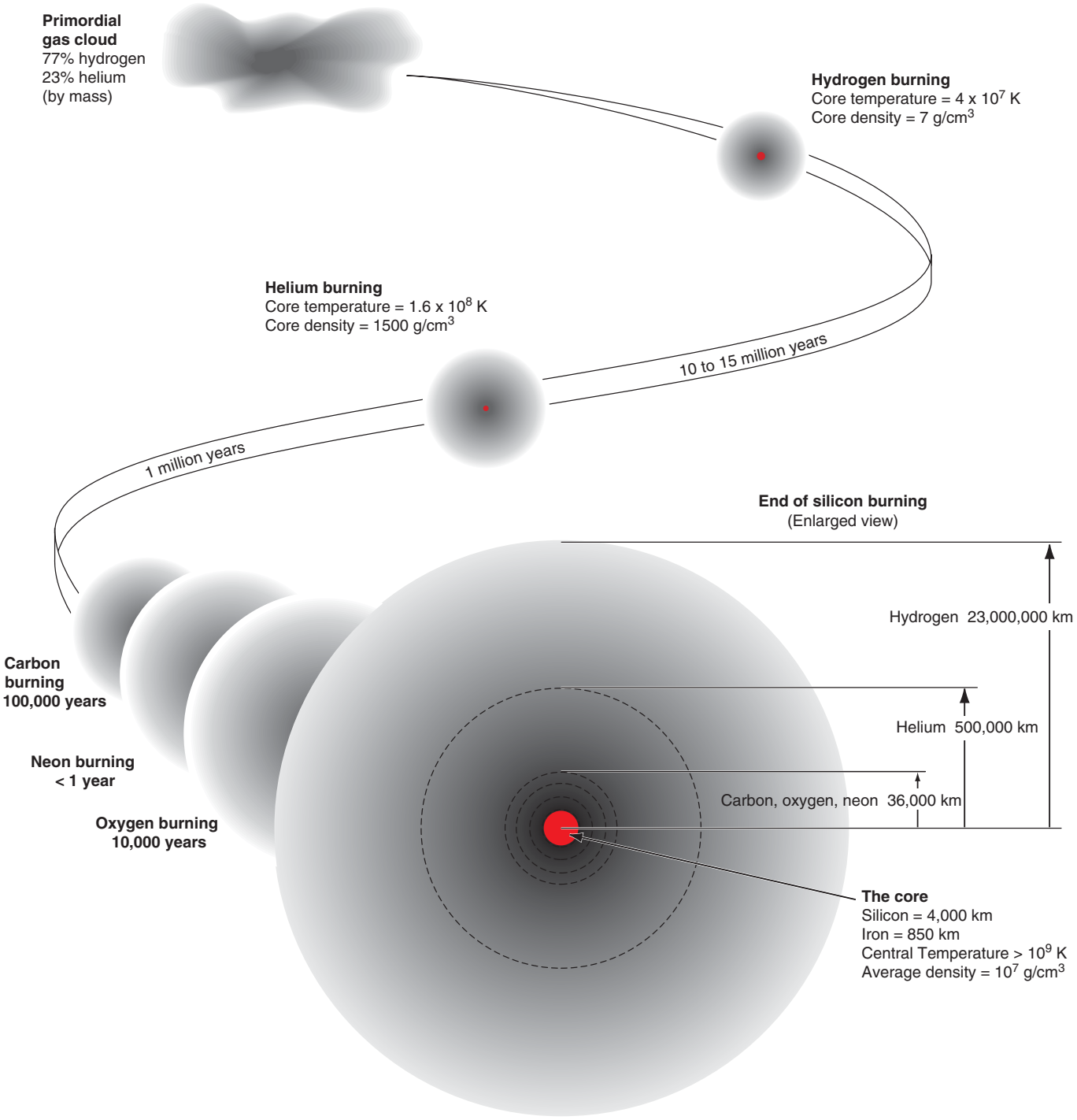


Fig. 2. The life of a massive star. A star is born when a huge cloud of primordial gas is compressed by gravity. The compression raises the density and temperature of the gas to the point that hydrogen nuclei can fuse into helium within the star's core. Both hydrostatic and thermal equilibria are quickly established. (See text.) The star will burn hydrogen for tens of millions of years, gradually accumulating helium in its core. Eventually, the core is fully depleted of hydrogen, and fusion stops. The core contracts, which leads to higher pressures, temperatures, and densities; and a new burning phase begins. Helium is fused into carbon within a hotter, denser, and much smaller core, even though the star itself has become larger during this phase. Over the course of its lifetime, the star's core will become smaller and much denser as it burns, in succession, carbon, neon, oxygen, and silicon. At the end of silicon burning, the star has developed a layered structure, shown here for an 18-M star. Note the tiny silicon and iron core. The core is 100 million times denser than the hydrogen layer.

The Core Collapse

Initially, the core, which at this point is about 5,000 km in diameter, begins to collapse relatively slowly. Two things quickly turn it into a nearly free-fall collapse. The first is that, because of compression, the temperature in the central regions of the core begins to rise above 5×10^9 K, and photons generated within this inner core become very energetic. When those photons bombard iron, the iron nuclei photodisintegrate, or break up into alpha particles and neutrons. As mentioned earlier, this reaction absorbs, rather than releases, energy. Meanwhile, the density has been steadily rising. Above approximately 10^{10} g/cm^3, electron capture reactions increase significantly. This reaction is a weak interaction process, in which a proton "captures" an electron and is transformed into a neutron, along with the emission of an electron neutrino. This process is called inverse beta decay.

At the still modest densities, before the formation of the neutron star, the neutrinos easily escape from the core and remove energy. As the collapse progresses and the density rises, encounters between protons and electrons become even more likely. In addition, the probability that the electron capture occurs depends on the square of the electron energy. Because the work done on the core compression renders the electrons more and more energetic, the rate of inverse beta decay increases tremendously during the infall. Neutrinos are produced in copious amounts. Despite the rising density and temperature, the rate of cooling within the inner core actually increases, and in fact increases in a runaway fashion. Thus, within a core diameter of perhaps 4,000 km, the collapse proceeds ever faster, much faster in fact than the remainder of the core region.

The collapse takes on a different character once the density of matter reaches 10^{11}–10^{12} g/cm^3. At those densities, the neutrinos begin to scatter off of nuclei and unbound nucleons. This means that the neutrinos no longer escape from the superdense inner core region; but instead they begin a slow, random walk toward the surface. The neutrinos are "trapped" and must diffuse through the core on a timescale that is comparable to, or longer than, the collapse itself.

Diffusing neutrinos? This is a particle that ordinarily passes through half a light-year of lead without scattering. But the conditions found in the center of a collapsing star are like none found anywhere else in the universe.

The diffusion has two main consequences.

The first is that the core can no longer cool down. Because "classical" heat transport by electromagnetic radiation or by matter conduction is totally negligible within the inner core (because of the extremely short mean free path of the photons and electrons), neutrinos are the only agents capable of removing energy, and so the inner core gets very hot, equivalent to tens of MeV, 10^{10} K, temperature.

The second consequence is a bit more arcane. Neutrinos of all types, (ν_e ν_μ ν_τ), come into thermal equilibrium with free electrons and

positrons, forming what is called a lepton gas. Because the leptons are fermions, they obey the Pauli exclusion principle, which states that any two fermions cannot occupy the same state, be it an energy state or a spatial overlap. The result is that the leptons resist being squeezed together as the core collapses and begin to exert what is called a "degeneracy" pressure that is independent of temperature. What this means is that a significant amount of nonthermal energy is stored within the lepton gas.

The pressure exerted by the lepton gas, however, is insufficient to halt the core collapse. This is because the response of this type of gas to further compression is soft, that is, the pressure does not increase very fast when the lepton gas is compressed.* But gravity, on the other hand, only increases in strength as the core shrinks in size. In the end, the increase in pressure caused by the neutrinos and electrons being squeezed is insufficient to balance the increasing pull of gravity and the collapse proceeds.

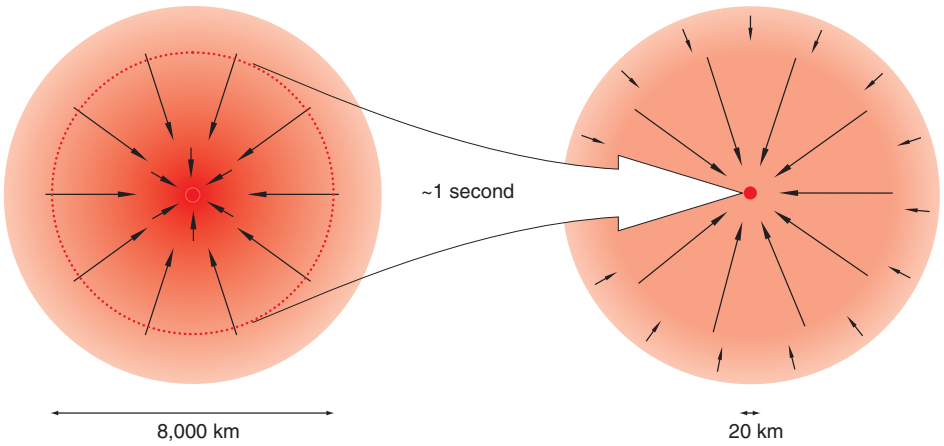
In fact, the core collapse would continue indefinitely and lead to the formation of a black hole were it not for strong nuclear forces. When the density begins to surpass 10^{14} g/cm^3, individual nuclei and nucleons are in effect squeezed to touching one another. Like the leptons, nucleons are also fermions, and so a substantial nucleon degeneracy pressure builds. In addition, short-range repulsive nuclear forces come into play that provide additional pressure support. Together with the "lepton degeneracy" pressure, the total internal pressure inside the core increases dramatically with density. Once past a density of about 3×10^{14} g/cm^3, the inner core stiffens and becomes more rigid. The collapse suddenly halts.

By the time the collapse stops, the inner core has been shrunk into a sphere that is only 40 km in diameter, although it has a mass of about 0.8 M. Much like a rubber ball which has been squeezed beyond its normal size, the core rebounds in an event called "core-bounce." The edge of the inner core is driven outward at roughly 10,000 km/s. These events are summarized in Fig. 3.

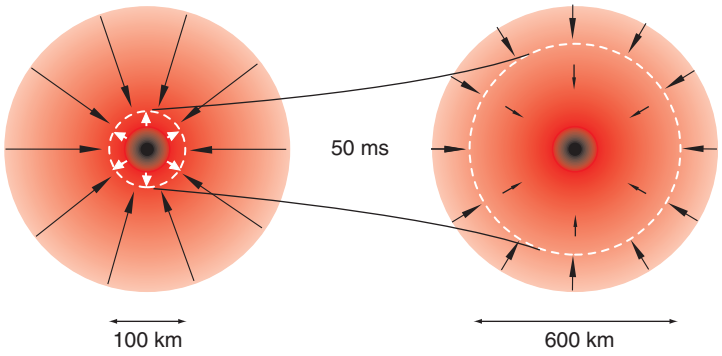
The barrage of details in the preceding paragraphs makes it easy to lose sight of just how incredible the events described really are. Before collapse, the stellar core is roughly the size of our planet Earth, although it is about 500,000 times heavier. In less than 1 s, half of that volume will collapse and be compacted into a sphere with a diameter that is only the size of Los Angeles, California. This tiny inner core will become extremely hot, with most of its energy bound up in neutrinos, but it will also become so dense that even those neutrinos have difficulty being radiated. This ball of nearly solid nuclear matter, which will become the neutron star once the supernova has occurred and is therefore called a proto-neutron star, will actually rebound and push its edge outward at 10,000 km/s.

*For those with a background in thermodynamics, the gamma coefficient of the lepton gas is only 4/3 because the neutrinos and electrons are relativistic.

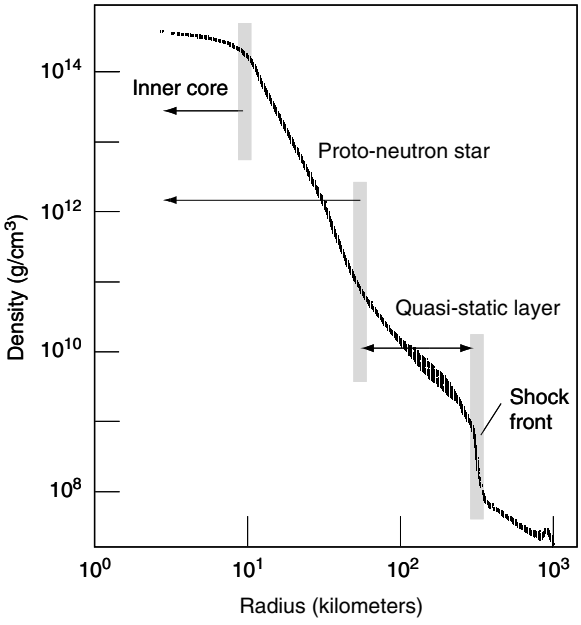
Fig. 3. Core collapse, bounce, and the postbounce structure of the new core. Once silicon burning ends, the star's core rapidly cools and loses pressure. As shown in (a), the entire 8,000-km core undergoes a nearly free-fall collapse and, in less than 1 s, about one-third of the mass is crushed into a sphere approximately 20 km in diameter. The new core, shown in (b) and (c), is vastly denser than the original. In these illustrations, arrows represent moving matter; and the length of the arrow is generally indicative of velocity.



(a) Core collapse. Core material (enclosed by red dotted line) races inward (long arrows), slowing down only when it runs into the dense matter in the center. The collapse happens so quickly that most of the material outside the core is unaware that the collapse has even occurred. The original core had an average density of about 10^7 g/cm^3, but the average density of the new core (red dot) exceeds 10^{14} g/cm^3, which is more than 10 million times greater than before.



(b) Core bounce. Once the central density of the new core reaches 3×10^{14} g/cm^3, the "inner" core effectively becomes incompressible and no longer collapses. Immediately, the inner core expands and surges outward (white arrows). The core quickly swells to a diameter of about 80 km (the protoneutron star, shown in grey) as a high-velocity shock front (white, dashed line) pushes its way against the infall. Within a few milliseconds, however, the shock loses energy, stalls, and stops its rapid outward movement.



(c) Postbounce structure and density profile of the core. As infall passes through the shock front, the material slows down, and its density increases dramatically (small black arrows). A dense layer, called the quasi-static layer, begins to accumulate outside the protoneutron star. The shock front slowly moves outward. The graph of density versus radius (both plotted on log scales) reveals the extent of the quasi-static layer, the protoneutron star, and the inner core (the very center of the protoneutron star) 50 ms after the bounce. These regions are typically defined in terms of density. At roughly 3×10^{14} g/cm^3, the inner core extends to a radius of about 10 km and encloses about 0.5 M. The protoneutron star decreases in density to about 10^{11} g/cm^3, encloses about 1.2 M, and extends to about 40 km. The shock front (evident as a sharp change in density) has by this time advanced to an approximate radius of 300 km.

The protoneutron star forms so quickly that the remainder of the original iron core, which still extends out to a radius of a few thousand kilometers, is left trying to catch up. The outer core is a shell of comparatively low-density matter with little internal pressure, and therefore falls inward freely. It is infalling toward the protoneutron star at the incredible velocity of 60,000 km/s.

Finally, the rest of the star is completely unaware of what is going on. This is because information about the core collapse can only travel as fast as the speed of sound. In the fraction of a second that it takes for the collapse to occur, information can only reach out to a few thousand kilometers, whereas the stellar envelope may extend hundreds of millions of kilometers into space. Thus, much like a cartoon character suspended in midair, most of the star has yet to learn that the rug has been pulled out from under it.

The processes just described are our current view of events leading up to core bounce. The physics has been fairly well understood for the last 25 yr, although the view underwent many revisions and modifications as more processes were considered and the role of neutrinos elucidated. But core bounce is not a violent supernova, merely the precursor to one.

Consensus regarding the postbounce physics, the real heart of supernova dynamics, has been much slower to emerge. The crux of the problem is how to turn what is essentially an implosion (the core collapse and the infalling of matter) into an explosion. What was largely unappreciated was the extent to which the details of protoneutron star formation would affect the ability to generate an explosion. As we shall see, neutrino diffusion will provide one key to the solution of that problem.

Making Stars Explode

The first modern model of supernovae was presented in 1960 by Colgate and Montgomery H. Johnson. It postulated that the core bounce was sufficient to produce an outwardly moving shock wave that would continue to move through the outer core and eventually expel the stellar envelope in a large explosion. This model later became known as the prompt mechanism, because the model shows that the explosion occurred immediately after the bounce. However, for the bounce shock to move out, it needs to beat back the infall of the envelope. Detailed calculations made in the 1960s by Colgate and Richard H. White and many similar calculations in the 1970s and 1980s showed that the postshock temperature is so high that many of the cooling processes that led to the core collapse, namely iron photodisintegration and intense neutrino emission, apply equally well to the shock front. Rapidly drained of energy, the shock stalls for all but the most extreme assumptions about the precollapse structure of the star.

It was also realized in the early 1960s that the amount of heat released by the core collapse and infall of the surrounding layers constitutes an immense reservoir of energy. The energy comes from the compression work done by gravity on the gas although most of that energy is tied up in the degeneracy energy of the lepton gas, as opposed to heat. The total energy available turns out to be much larger than what is typically observed as kinetic energy of the supernova debris, which is about 10^{51} ergs. (This amount of energy was coined a "foe" by Han Bethe: 1 foe = 10^{51} ergs.) The total energy released by the gravitional collapse is on the order of 300 foe. It was realized that most of this energy is carried by neutrinos liberated in the formation of the protoneutron star, and today it seems natural to expect that supernovae are powered by simply tapping a small fraction of that neutrino flux. However in the early 1960s, the idea that neutrinos might do anything dynamical, let alone power an explosion, seemed preposterous to many.

It was in this context that in 1965 Colgate and White put forth the first model that invoked heating by neutrinos as the mechanism responsible for supernovae, and they thus anticipated all modern supernova theories. They used a hydrodynamic code, originally designed for nuclear weapons modeling, to quantitatively analyze their theory. Theirs was the first attempt to simulate the hydrodynamics of a supernova with a computer and probably the first hydrodynamical simulation ever done in astrophysics. Although their model was substantially modified in the 1970s by new insights in nuclear and neutrino physics, some basic aspects remain current, including the collapse of an iron core, neutrino emission and absorption as the fundamental means of energy transport, and the heating of infalling matter outside of the protoneutron star.

Unfortunately, their model suffered from a lack of information. Neutral current scattering by nuclei was unknown at the time (It was discovered in 1975 by Tubbs of the Los Alamos Theoretical Physics Division and Schramm.), and along with other differences in understanding concerning the late stages of stellar evolution, Colgate and White were led to theorize that the protoneutron star was born cold. All of the heat had escaped by means of an unimpeded neutrino outflow. Their supernova was powered by a neutrino flux that was generated as matter crashed down onto the surface of the protoneutron star. That material heated up sufficiently to emit neutrinos, some of which would then deposit their energy in the outer core. A mighty explosion ensued.

However, it was known that outside of the protoneutron star, the fraction of neutrinos which interact with matter is small, and the pressure of the infalling envelope is large. Colgate and White used an inconsistant and somewhat artificial means of transferring energy to the infall. As models matured, numerous computer calculations carried out in the 1970s and 1980s showed that the neutrino energy deposition was insufficient to power an explosion. The model of Colgate and White fizzled.

The Colgate and White model is frequently believed to have failed because of an improper postbounce neutrino emission and absorption algorithm. In fact, the model fails because it neglects the effects of neutrino cooling during core collapse. In the model, a supernova develops when a high-temperature iron core collapses. The high temperature leads to a very rapid rate of electron capture, and the core becomes neutronized very quickly and at relatively low densities. A neutron star forms directly from the collapse. The collision energy of the infall onto the neutron star is high enough to generate a high-temperature accretion shock front, and high-energy neutrinos emitted from that front are readily absorbed in the falling matter. Once neutrino cooling was added to the model, the core temperature and the nucleon boiloff rate were reduced relative to what Colgate and White had originally considered. Neutronization proceeded much more slowly because there were fewer free protons. In addition, neutral-current interactions were not known at the time. They have the effect of enhancing the neutrino trapping rate, which further retarded core neutronization. Neutrino trapping also led to a large degenerate lepton pressure that supported matter at a radius some three-to-five times greater than the radius of a neutron star. Thus, it was eventually shown that the accretion shock front formed at larger radii, and matter that fell onto this shock front was not nearly energetic enough to produce the high-energy neutrinos needed to drive a supernova.

In 1982, James Wilson from Lawrence Livermore National Laboratory noticed that if he let his computer simulations run for a long time (~0.5 s after the bounce), successful supernovae occurred. The diminishing infall of matter from the envelope of the star would allow neutrinos slowly leaking from the protoneutron star to heat the external matter enough to generate an explosion. Apparent fizzles evolved into successful blowouts by what later became called the “delayed” (as opposed to prompt) mechanism.

Ironically, the very feature that causes the prompt mechanism to fail, namely the stalling of the bounce shock wave, helps the delayed mechanism to work. Although the bounce shock itself is unable to deliver an explosion and reverse the infall, the shock does manage to move out to a distance of 100 km so above the protoneutron star. This movement results in the creation of a relatively stagnant layer of material between the stalled shock front and the surface of the protoneutron star. This quasi-static layer is available for heating and can re-energize the shockwave, driving it outward (Fig. 4).

Although delayed mechanism models were able to produce explosions, less satisfying was the fact that all through the 1970s and 1980s computer simulations of supernovae seemed to be highly sensitive to the smallest details of the implementation of the physics. Where one group obtained explosions, delayed or prompt, another would get fizzles simply because the approximations used in the modeling were different. This was worrisome not only because it put any calculations at the mercy of a new wrinkle in the theory, but also because real supernovae

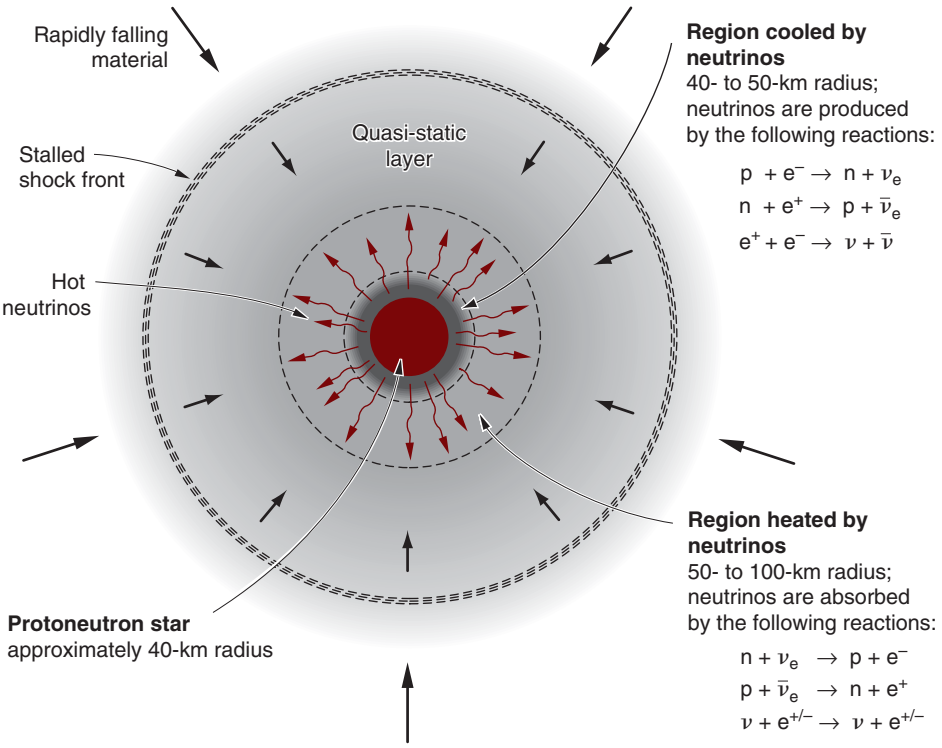


Fig. 4. Behind the front: heating matter with neutrinos. As a result of core bounce, a shock front moves beyond the ultradense surface of the protoneutron star. The shock loses energy as it propagates and stalls. The front is prevented from recollapsing by the pressure support of the degenerate lepton gas, and so it remains at a relatively stable radius, creating a quasi-static layer of dense matter. Some of the energetic neutrinos leaking from the protoneutron star deposit their energy in the quasi-static layer. The matter expands and becomes buoyant. The neutrinos, therefore, transfer energy out of the extremely high-temperature core and into a large mass of lower-temperature material.

do not seem to have such problems. Explosions of fairly uniform energies, always on the order of 1 foe, appear to be produced quite readily. Supernova theory seemed trapped in an endless cycle of successful and failed explosions; and in desperation, many alternative theories involving core rotation, nuclear burning, magnetic fields, etc., were considered. These met without much success. What astrophysicists really needed was some sort of lucky break, and on February 24, 1987, they got it.

Supernova 1987A

SN 1987A (the first supernova seen in 1987) owes its major impact on supernova theory to one reason: it occurred relatively nearby. It flared up a modest 170,000 light-years away in the Large Magellanic Clouds, which is a satellite galaxy of our own Milky Way. For the first time, it was possible to look back in photographic archives at the location of the explosion and find that the parent star of the supernova had been a blue supergiant of mass 20 M. Because of its proximity, and thus brightness, observations of unprecedented accuracy became possible.

However, the most important contribution from SN 1987A was that it was the first and, until now, the only supernova from which neutrinos were observed. Two underground detectors sensitive to electron antineutrinos, Kamiokande II in Japan, and Irvine-Michigan-Brookhaven (IMB) experiment in Ohio, detected bursts of 12 and 8 neutrinos respectively over a 10-s interval. The small number of neutrinos detected did not allow for detailed quantitative diagnostics of SN 1987A, but we did get qualitative estimates of what had happened. (Note that the Irvine neutrino physics group was set up by another T-Division member, Nobel Laureate Frederick Reines, when he moved there.)

First of all, the picture of the formation of a hot neutron star and its cooling by neutrino emission was given strong credence, in that the detected signal was entirely consistent with our current theories of core collapse. The energy spectrum of the neutrinos permitted an estimate of the total radiated energy, which was consistent with the genesis of a 1.4 M. protoneutron star with a 15-km radius. The energies of individual neutrinos corresponded to the expected initial temperature of a protoneutron star, while the duration of the bursts was in line with the cooling timescale of such an object.

At the same time, spectroscopic observations of SN 1987A unequivocally showed that the ejected envelope is stirred up considerably during the explosion. Especially puzzling was the presence of iron in the outer layers of the ejecta, indicating that a substantial amount of mixing of material from the core to the outer layers had taken place over large distance scales. Some of this mixing was explained by instabilities occurring in the period of hours when the shock wave runs from the core to the distant surface of the star, well after the explosion has been launched. Nevertheless, this circumstance promoted a new awareness of the possibility that violent instabilities might be involved in the explosion mechanism.

This was not an entirely new idea. Richard Epstein, Los Alamos physicist, had already proposed in 1979 that instabilities at the edge of the protoneutron star might be important to consider. Hans Bethe later pointed out that an explosion due to neutrino heating, as in the delayed mechanism, would necessarily lead to convection because matter is "heated from below." However, due to computer limitations and to the complexity of supernova physics, most work on supernova explosions had been done with the assumption of spherical symmetry, thus reducing the problem to one spatial dimension (the radius, as in Fig. 3). As a result of this focus, instabilities were mostly envisioned as mixing matter at microscopic scales rather than as turnovers leading to large-scale bulk flows.

Over the years that followed the advent of SN 1987A, we were convinced that to explain the observed churning of elements, one had to look into the explosion mechanism. We felt that "standard" one-dimensional modeling of the explosion mechanism was likely to miss some important qualitative aspects of the post core-collapse evolution which might be important.

In the late 1980s, in recognition of the need for a robust mechanism that would always lead to a supernova explosion, we sought a new mechanism of heat transport. Diffusive convection, the normal description of turbulent convection, would not be different from neutrino convection. What we needed was a transport mechanism with a scale that was large compared to the local density scale-height of the neutrino-heated atmosphere. Such a mechanism is familiar to us almost every day of summer. A thunderstorm is a plume that raises a cubic scale-height of hot moist air from the ground level and places it several scale-heights up above the tropapause and thus above the turbulent atmosphere. Would such a mechanism work for a supernova, and especially within the neutrino-heated "high-entropy atmosphere" in quasi-equilibrium, resting on top of the neutron star?

As a result, in 1991, we established a research effort at Los Alamos with the goal of simulating the explosion mechanism in multidimension. We were helped by the arrival of cheap, powerful desktop workstations on which 2-D, and sometimes 3-D simulations could be run. Even in

primitive initial calculations, we immediately noticed intense convective instabilities (akin to boiling) which were driven by neutrino heating of the matter in the vicinity of the protoneutron star. We were also encouraged by the fact that we obtained successful explosions in a way which seemed fairly insensitive to the details of the numerical implementation of the physics. Subsequent simulations carried out by ourselves and several other groups with an ever-increasing degree of realism (by tracking more physical processes) confirmed the key role of neutrino-driven convection, "The Colgate Mechanism," in the genesis of the explosion.

Supernovae and Convection

Figure 5 shows some of the results derived from our simulations. As in most models, our postbounce shockwave has stalled some 50 ms after the bounce, at a radius of about 300 km. Infalling material from the stellar envelope is slowed down by the stagnant shock and collected into downdrafts which flow toward the neutron star. As the matter descends, it experiences larger and larger neutrino fluxes which heat it. Some of the matter becomes buoyant and floats up in the form of large bubbles. As time goes on, more and more bubbles push out on the shock, finally succeeding in driving it out in a successful explosion.

The crucial element of this picture is the simultaneous existence of cold inflows and heated outflows, which was impossible in the context of

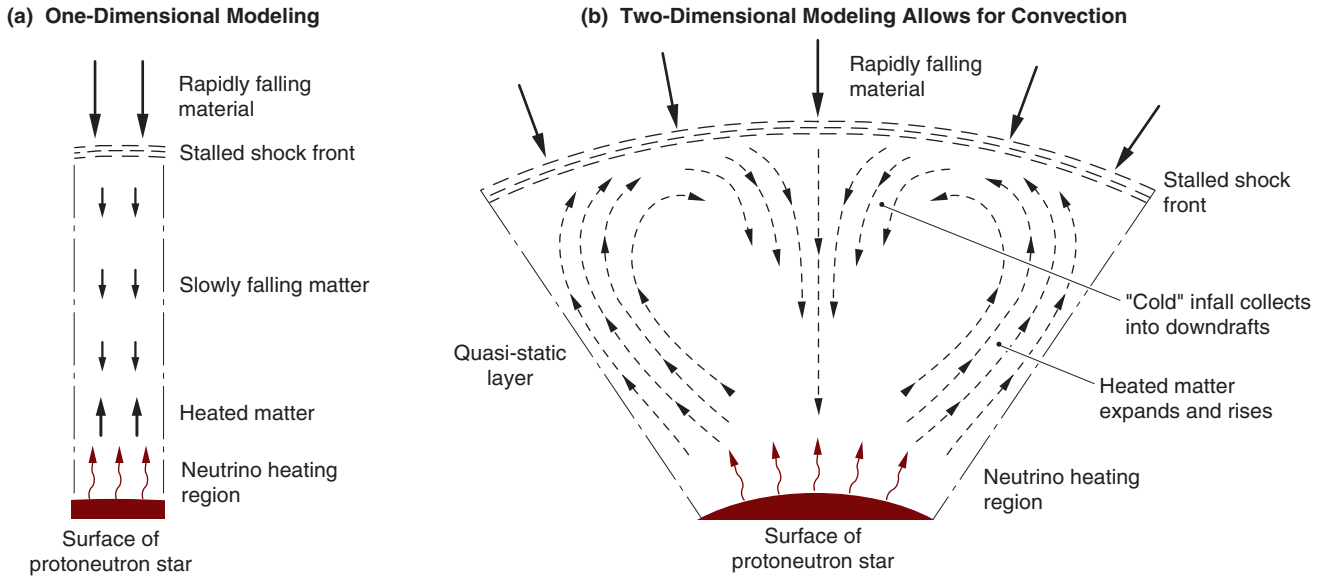


Fig. 5. A convective engine. (a) For simplicity, supernovae were often modeled in one dimension. A star was assumed to be spherically symmetric, its radius being the only spatial parameter that mattered. Doing simulations was, therefore, equivalent to doing physics in a long tube, even though the transfer of heat from one end of a pipe to the other is not very effective. (b) With the advent of multidimensional models, convection could occur. Hot, buoyant material could rise in one part of the star, to be replaced by cooler material falling from some other region. An in/out circuit is established that allows for the efficient and continuous transfer of heat out of the core and into the quasi-static layer. Energy from the gravitational collapse is thus converted into mechanical work, as heat is being transferred between hot and cold reservoirs. In this sense, supernovae can be thought of as being powered by a simple convective engine.

one-dimensional (1-D) models, which average quantities at a given radius. This has led us to elaborate a new paradigm in which the supernova is viewed as a convective engine, with the hot protoneutron star as a heat source radiating neutrinos and with the envelope of the star as a cold reservoir. The circulation of matter and heat exchange allows mechanical work to be extracted from the energy liberated by the gravitational collapse. This paradigm explains the failure or marginality of simulations in one dimension; heat transport with one pipe can hardly be effective, while in two dimensions, an in/out circuit can be established.

In effect, obtaining a supernova explosion is akin to blowing up a pressure cooker. The containing lid of the cooker is the stalled shock, while the ram pressure of the infalling envelope is what keeps the lid clamped tightly to the pot. To blow up the cooker demands a buildup of pressure against the lid, which in turn depends on a good transport of heat between the bottom and top. If the stirring is insufficient, then in cooking something like a heavy French sauce, one doesn't heat the lid but ends up with a burnt sauce with a cold top layer. It is convection that allows heat to be carried to the lid and build up a pressure to push against it.

Furthermore, once an explosion is obtained, the matter is ejected. Because there is nothing left to heat, the energy input stops. This self-regulating feature builds into the model a natural explanation to the general homogeneity of explosion energies for different supernovae. This tradition of evermore complicated dimensionality calculations at Los Alamos has recently been extended with the crowning achievement of Chris Fryer and Mike Warren of recalculating the explosion in 3-D with nearly exactly the same results. (See Fig. 6.) The large scale of the convective elements is preserved regardless of the dimensionality as a result of the geometry, neutrino heated equilibrium atmosphere, and the relativistic equation of state. Large-scale convection created by neutrino heating indeed seems to solve the mystery of a "robust" supernova mechanism.

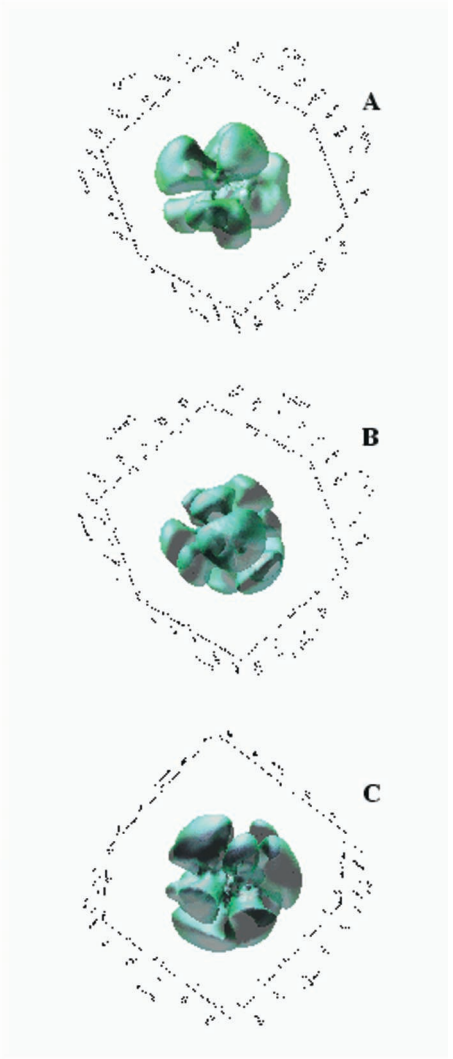


Fig. 6. The isosurface of material with radial velocities of 1,000 km/s for 3-D simulations with increasing resolution (300,000, 1 million, and 3 million particles). The isosurface outlines the outward-moving convective bubbles. The open spaces mark the downflows. Note that the up-dwelling bubbles are large and have very similar size scales to those of the 2-D simulations.

The Last Word

It should be clear by now that neutrino and supernova physics are intimately linked; and, therefore, it is not surprising that one of the dearest wishes of astronomers and neutrino physicists alike is for a supernova to occur within our own galaxy. If such an explosion were to take place, it is estimated that the new large neutrino detectors would register several thousands of events, thus providing us with a detailed picture of the events that accompany core collapse, a picture which is otherwise shielded from our view by the opaque envelope of the star. Moreover, an intense neutrino signal would provide clues and constraints on neutrino oscillations or other physical processes which we may not have imagined yet. It is in part the prospect of such serendipitous discoveries that promises to make the field of supernova neutrino astrophysics an exciting one for years to come.

Gravitational Condensate Stars: An Alternative to Black Holes

Einstein
Mazur
Mottola

by Emil Mottola

Many believe that rotating black holes are the inevitable endpoints of stellar collapse. Theoretical Division researchers have proposed another solution for the endpoint of gravitational collapse—an interior region with an equation of state that acts as a repulsive force to keep black holes from collapsing further and the separation of this interior region from an exterior region by a thin shell within which the speed of sound equals the speed of light.

The vacuum Einstein equations of classical general relativity possess a well-known solution for an isolated mass, M, the Schwarzschild or black hole solution. When the black hole is rotating, the appropriate solution of Einstein's equations is called the Kerr solution. It is generally believed that black holes of this kind are the inevitable endpoints of stellar collapse, when a star has burnt all its nuclear fuel and its mass is too large for its collapse to be prevented by any known force or matter equation of state (EOS).

We have proposed a new solution for the endpoint of gravitational collapse. By extending the concept of Bose-Einstein condensation (GBEC) and low-temperature phase transitions to gravitational systems, we have found a new solution to Einstein's equations with an interior de Sitter condensate phase and an exterior Schwarzschild geometry of arbitrary total mass, M, which is cold and compact. There are two new features about this solution that make it very different from black holes.

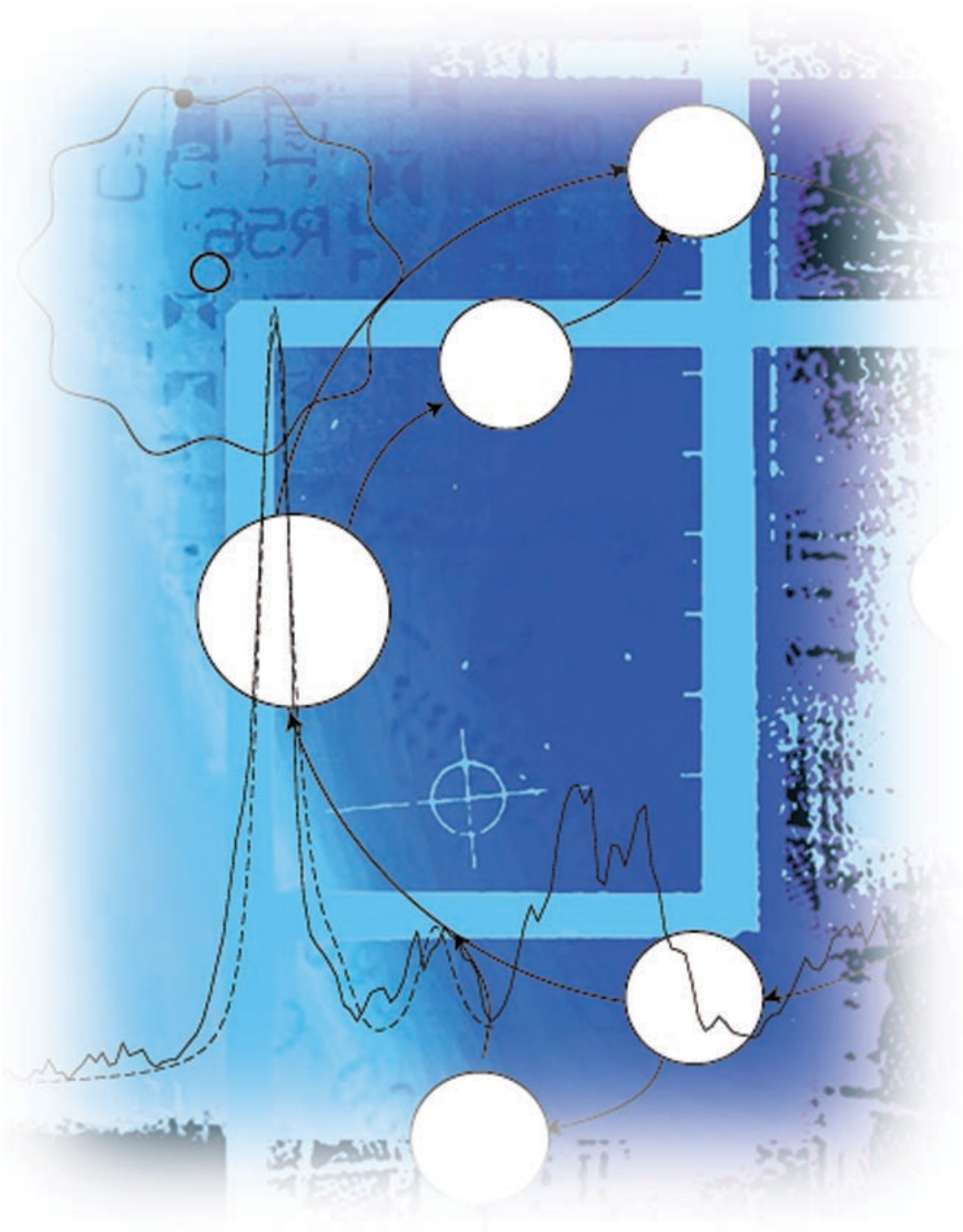
First, the interior region has an EOS that acts like a repulsive force to keep the object from collapsing any further. This is the gravitational vacuum condensate state, which behaves just like a cosmological constant, the same vacuum energy and negative pressure to which has been attributed the acceleration of the universe, recently reported by observations of distant supernovae.

Second, the interior is separated from the exterior in our solution by a thin shell of ultrarelativistic matter, which has the EOS, $p = r$, so that the speed of sound is the same as the speed of light in the shell. This shell is at the place that the event horizon of the classical black hole would be; but, because some definite matter exists there, our configuration has no event horizon (i.e., a place where light becomes trapped) or surface which makes a black hole black.

Also, unlike black holes, the new solution has no singularities, i.e., places where energy and pressure become infinitely large. The entropy of the new solution is maximized under small fluctuations and is given by the standard hydrodynamic entropy of the thin shell, separating the interior and exterior, so a collapsed star of this kind is thermodynamically stable and has no information paradox.

The assumption required for this solution to exist is that gravity undergoes a vacuum rearrangement-phase transition in the vicinity of the Schwarzschild radius. Specifically, in this region quantum fluctuations on the scale of $2\,GM/c^2$ dominate the stress tensor. These fluctuations grow so large that the EOS approaches the most extreme one allowed by causality, $p = r$. As this causal limit is reached, the interior space time becomes unstable to the formation of a new kind of gravitational GBEC described by a nonzero macroscopic-order parameter. If the square of this parameter is a constant scalar, it must couple to Einstein's equations in the same way as a cosmological term. A suggestion for the effective theory incorporating the effects of quantum anomalies that could give rise to both this interior GBEC phase and the $p = r$ shell has been presented in a previous paper with Pawel O. Mazur of the University of South Carolina.

We believe that the new GRAVitational VAcuum condensate STAR (gravastar) can be distinguished in principle from black holes by the existence of the shell which can eject matter, perhaps violently in energetic bursts, when struck. This may provide a rather different and more efficient central engine for energetic sources in the cosmos, such as active galactic nuclei, and gamma ray bursters, whose mechanisms are still unclear. The shell could also emit gravitational wave signatures which are quite different from black holes. We are working on both the astrophysical implications and the new directions this solution suggests for gravitational physics in the quantum domain.



Opacity Efforts

by John J. Keady

The interaction of various kinds of matter with radiation controls many physical phenomena, from microscopic to cosmological scales. Such interaction is referred to as radiative opacity or simply opacity. From Edward Teller in the 1940s to the present day, Theoretical Division scientists have done extensive research in opacity. Their work has served the international community in various ways, including the creation of an Astrophysical Opacity Library and the development of the Los Alamos Astrophysical Opacity Web Page: http://www.t4.lanl.gov/.

Introduction

The impedance offered by matter to radiation is often referred to as radiative opacity or simply opacity. This interaction of matter and radiation controls many physical phenomena from microscopic to cosmological scales. Hence, for many decades opacity has been of great interest to astrophysicists.

For example, in the center of stars where hydrogen and helium are consumed in nuclear reactions, the radiative opacity controls the transport of radiation and is crucially important in determining the structure throughout most of the stars.

Opacities have other astrophysical applications and applications to other science and technology issues:

- star formation from dense interstellar clouds;
- the solar neutrino problem;
- solar system and planet formation;
- modeling of planetary atmospheres;
- analysis of ground-based remote sensing and spacecraft data;
- energy balance in the Earth's atmosphere (the greenhouse effect, climatology, meteorology, and urban heat island effect);
- perturbations of planetary atmospheres (the collision of Comet Shoemaker-Levy 9 with Jupiter, asteroid mitigation, and nuclear winter);
- design of nuclear weapons;
- weapons effects;

Abdallah
Argo
Barfield
Barnes
Bauche
Carson
Collins
R. Cowan
Cox
Eilers
M. Geoppert-Mayer
Green
Huebner
Iglesias
Jacobsohn
Karzas
Keady
Kilcrease
Latter
Liberman
Magee
Mark
H. Mayer
J. Mayer
Merts
Sakharov
Sharp
Stewert
E. Teller

- inertial confinement fusion (ICF) and magnetic confinement fusion (modeling and plasma diagnostics); and
- uranium-fueled gas-core and plasma-core reactors (nuclear rocketry).

The matter can be in any of a number of forms—molecules, when the gas is cold enough; neutral atoms at higher temperatures; and ions (plasma) at still higher temperatures. These forms can coexist, as they do in and around sunspots. They can also be in particulate form ("dust"). All have been of interest over the years, and each presents special challenges.

In the gas phase, radiation (a photon) interacts with matter by means of a number of processes. There is the photo-ionization process (bound-free), where a photon removes an electron from the molecule, atom, or ion (ionizing radiation). In brehmstrahlung (free-free), a free electron in the vicinity of an ion absorbs a photon. In another photoabsorption process (bound-bound), an ion absorbs a photon that is insufficiently energetic to ionize the ion. There are other related more exotic processes. All these processes and their inverses compete. Initially, the importance of the bound-bound process was underappreciated for many years, until just after WWII.

Theoretical Division Begins Opacity Work

While at Los Alamos, Edward Teller was among the first to recognize the long-term implications of opacity to nuclear weapons. During the war, Teller persuaded Maria Mayer to initiate an off-site opacity project at Columbia University. This project required special approval from General Groves. Teller would make occasional trips from Los Alamos to New York to collaborate on the project. Harris Mayer (no relation to Maria and her husband, Joe) was recruited into this effort. Shortly after the war ended, Maria Geoppert-Mayer and Joe Mayer and the Columbia opacity project migrated to Chicago. This left Harris Mayer in a bit of a bind at Columbia, since Teller had already moved to Chicago, so that Harris was essentially sponsorless at Columbia. Eventually an accommodation was found, and Harris Mayer successfully defended his dissertation at Chicago. Boris Jacobsohn also completed a dissertation on high-temperature uranium opacity calculations under Teller at about this time. For many years both an unclassified and classified version of his dissertation could be found in the appropriate parts of the Los Alamos library. From Chicago, Mayer made the transition to Los Alamos, arriving in March 1947.

Prior to this point in time, essentially all the interest in opacities was in astrophysics. The physical processes thought to be of greatest interest were bound-free, free-free, and scattering. It was thought that the spectral lines associated with the bound-bound transitions were relatively unimportant for two reasons: sparsity and narrowness. Simple atoms and ions (e.g., hydrogen, helium, lithium, etc.) have few bound electrons, resulting in a simple spectrum of narrow lines that would not

absorb much radiation. Teller was among the first to realize this is not true, in general, for multielectron systems. With more than several bound electrons (e.g., iron), the resulting atomic configurations become both numerous and complicated. New quantum mechanical calculations of spectral-line broadening indicated that spectral lines were not always narrow at densities of interest. All this was also realized by Andrei Sakharov, at about the same time, in the Soviet Union. This had major implications in astrophysics and immensely complicated the accurate calculation of opacity in general. So it was Teller who prevailed on Mayer and Jacobsohn to investigate statistical approaches for modeling line absorption.

After arriving at Los Alamos, Mayer systematically surveyed and re-examined the existing methods of opacity calculations and continued investigating statistical approaches to calculating the bound-bound absorption. Beyond realizing the importance of the bound-bound absorption, it was also quickly realized that any straightforward approach to dealing with it was computationally intractable. Thus statistical approaches were a necessity. Harris Mayer's report, which included the first comprehensive iron opacity calculation, was published in 1948, [1] and has been a much referenced standard for many years. Mayer's approach is sometimes referred to as the "average ion model." The alternative approach is referred to as the "explicit ion model" or, in more modern terminology, as "detailed configuration accounting" (DCA), where one attempts to explicitly enumerate the atomic configurations existing in the plasma.

It should be mentioned that a considerable opacity effort subsequently developed in the Los Alamos Field Test Division (J-Division) under Art Cox in the 1960s. Among those working with Art on light element opacities were Al Merts, John Stewart, Don Eilers, and Norman Magee. Walter Huebner was in T-Division developing the light-element opacity code (LEO). The bound-bound absorption was being included in the opacity calculations, and astrophysical mixture opacities were being computed and distributed internationally. For a number of years, the effort of Art Cox and his associates was the primary source for astrophysical opacities for the international community.

There were also other opacity efforts within the defense community. R. Latter, W. Karzas, and Joe Green (affectionately known locally as Giuseppe Verdi) at the RAND Corporation, did extensive opacity calculations for elements heavier than iron. Some of their algorithms were eventually provided to Los Alamos, and were incorporated into the heavy element opacity code HEO.

Formation of New Group

The T-Division Equation of State and Opacity Group was formed in November 1971, at the instigation of then T-Division Leader Carson Mark. Walter Huebner, Al Merts, and Norm Magee moved to this new group in late 1971. Jack Barnes was group leader, working on

equation-of-state issues, while Walter Huebner was deputy group leader in charge of the opacity effort. Mary Argo, who had been an assistant of Edward Teller, and Bob Cowan, also joined the group, along with Dave Barfield. It was at this time that the first systematic attempts to generate an atomic database for opacity calculations using Cowan's atomic structure code were initiated. Even though this continued the significant improvement in the opacity calculations, computer limitations meant that some cruder hydrogenic data were still being used.

The number of requests for astrophysical opacities continued to rise, which led to the development of the MOOP code, which had a limited molecular opacity capability at low temperatures. Dave Barfield worked on molecular (air) opacities for J-Division field tests. Interest in molecular opacities continued through the 1980s, with Chris Sharp working on molecular equation of state, including solids, and molecular opacities. By the mid-1970s, interest in magnetic fusion had also risen dramatically. MOOP was modified from a thermodynamic equilibrium (TE) code to a coronal equilibrium (CE) code appropriate to model tokamak plasmas. Subsequent calculations of the plasma emission in tokamaks indicated very significant power loss caused by the iron and other materials ablated from various tokamak components, mostly injector tubes and walls. This was one indication that the road to successful magnetic fusion would be more difficult and expensive than initially imagined. This work was summarized in what became a classic report. [2]

By this time, in addition to the average ion capability, MOOP had a capability to do detailed configuration accounting using some experimental and calculated LS-coupled atomic data. The relevant breakthrough was the mixing of precalculated pure-element opacity tables on a temperature-electron degeneracy (electron pressure) grid. [3] This grid immensely simplified the calculation of mixture opacities. A new series of opacity calculations, the Los Alamos Astrophysical Opacity Library, was completed and used to generate a large number of astrophysical mixtures and mixtures of programmatic interest for the SESAME Library. Subsequently, both the Opacity and SESAME Libraries were extensively distributed internationally.

Also by this time, circa 1975, extensive opacity tables from other groups were available and these tables occasionally produced opacities that were quite different from the Los Alamos results. T. R. Carson's (the University of St. Andrews, Scotland) opacities for carbon, nitrogen, and oxygen were a factor 10 higher than the Los Alamos results at a temperature of 1 million degrees ("the CNO bump"), and had noticeably different effects on stellar structure calculations. In 1984, Carson spent a sabbatical at Los Alamos and collaborated with Norm Magee to find the source of the discrepancy. The difference in the opacities was traced to a bad renormalization in the St. Andrew's opacity code. Thus, the CNO opacity bump really did not exist.

Computer Contributions

In the early 1980s, the Cray supercomputers were online; and this development started to have a profound effect on our ability to generate large amounts of atomic structure data and subsequently calculate opacity. Computer hardware improvements, combined with Joe Abdallah's efforts in revamping the Cowan atomic structure code, led to, by the mid-to-late 1980s, to the ability to generate huge amounts of very detailed atomic data.

This led to the conception by Norm Magee of the LEDCOP opacity code, which was designed to fully exploit recent advances. Much, if not all, of the hydrogenic approximations of the MOOP opacity code would be replaced with very detailed atomic data from the revamped Cowan code. The amount of calculated atomic data was huge, and data compression was achieved by means of quantum defect fitting of energy levels and transition dipole moments. In 1965, Art Cox pointed out that one consequence of the hydrogenic approximations was that intrashell transitions (e.g., 3*p*-3*d*), were ignored, and that these were probably important. [4] The new approach would automatically include these transitions.

Circa 1985, Lawrence Livermore National Laboratory decided to start a new light-element opacity effort. Consequently, Carlos Iglesias was dispatched to T-Division for a few months for an apprenticeship with Norm Magee. The Livermore effort subsequently focused on using a parametric potential approach to generating the atomic data, in lieu of extensive atomic structure calculations. The resulting Livermore opacity code was called OPAL.

OPAL opacities preceded the LEDCOP opacities, demonstrating that the intrashell transitions were important under some circumstances (e.g., iron at about 200,000°), and resolved a long standing theoretical difference with the observed pulsational behavior of Cepheid stars.

Extending the earlier work of Walter Huebner and Mary Argo on the HEO opacity code, I incorporated the atom-in-jellium work of David Liberman into the calculation of heavy-element opacities.

Other Advances

The latter part of the 1980s and the 1990s brought experimental advances that permitted the first direct measurements of plasma absorption, at such facilities as the Rutherford Appleton Laboratory (RAL), the Livermore Nova laser, and the Sandia National Laboratories Saturn pulsed-power machine. Prior to this time, opacities were tested in an integrated fashion against experiments with outcome that depended, not only on opacity, but on other things with their own uncertainties, such as hydrodynamics, radiation transport, etc. These new experiments directly measured the monochromatic absorption. The indirect experiments were sensitive to the gray (averaged) opacity. [5] Thus, the new experiments represented an enormous improvement in sensitivity.

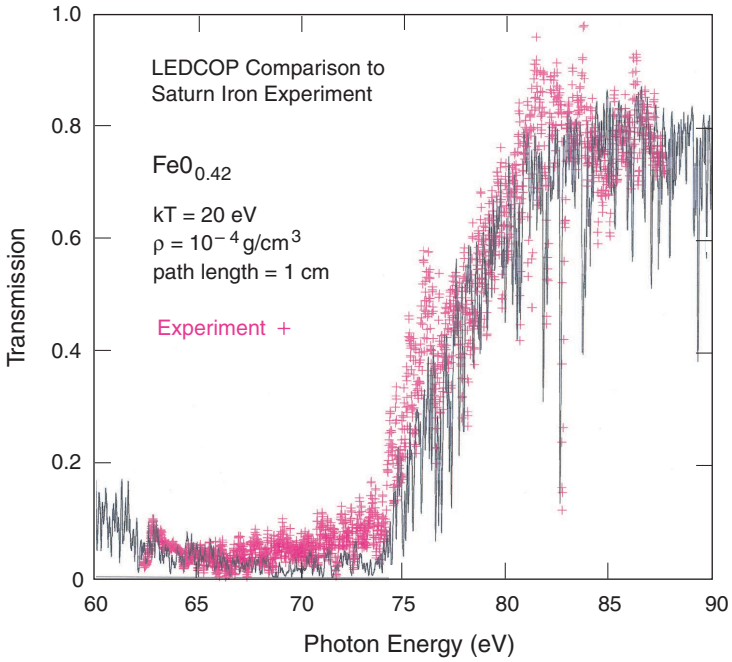


Fig. 1. Result of Springer's Saturn pulsed-power experiment and the Los Alamos calculation.

Among the first experiments were the RAL measurements on aluminium at a temperature of 40 eV. Using Cowan's atomic structure data, Joe Abdallah provided the first high-quality calculation of this experiment. Paul Springer (Livermore) and associates devised an experiment on the Saturn pulsed-power machine (in the 1990s) relevant for the Cepheid pulsation problem and a good test of the modern approaches to calculating opacity. [5] Their experimental result and the Los Alamos LEDCOP calculation can be seen in Fig. 1. The agreement is quite good and is even better if the nominal experimentally determined density and temperature used in the calculation are adjusted slightly (within the experimental uncertainties). This class of experiments has enormously increased our confidence about opacity calculations.

Complicated atomic configurations still require statistical treatments. Many transitions between atomic configurations (transition arrays [TAs]) are composed of millions of individual spectral lines, which still cannot be treated individually.

During the late 1980s and through the 1990s, Prof. Bauche (Paris) and his associates extensively published new results on the statistical representation of transition arrays. Prof. Bauche indicated that part of his inspiration came from chapter 21, "Statistical Distributions," of Bob Cowan's monograph. [6] Dave Kilcrease implemented this methodology for use in Los Alamos.

Los Alamos opacity codes now make extensive use of atomic data generated this way. While enhancements in computational capability continue unabated and permit direct use of more and more detailed atomic information, the statistical approaches will also be required in the foreseeable future. Ongoing work on the next generation of opacity code will use a variety of representations of the atomic structure: from the direct use of atomic structure data sets, as well as statistical representations of more complicated data sets.

Another frontier is in plasma effects on atomic structure. Most opacity calculations use isolated atomic data. Plasma effects are incorporated perturbatively. However, in many regimes of interest, the density is high enough to directly affect atomic structure. This has been largely ignored in the past, and may have a noticeable effect on opacity. The dense plasma work of Lee Collins and associates will have a bearing on the next generation of opacity calculations.

Muon-Catalyzed Fusion

by James S. Cohen

Alvarez
Armour
Cohen
Frank
Gershstein
Hale
S. Jones
Leon
Martin
Pack
Ponomarev
Sakharov
Struensee
Zeldovich
Zeman

Fusion of two nuclei can release a large amount of energy, but this reaction is difficult to achieve because of the electrostatic repulsion between nuclei. Theoretical Division scientists have worked with others to use muons (unstable elementary particles with a negative charge) to make the process easier. This work has helped bridge a gap between atomic and nuclear physics, has enabled nuclear reactions to take place at room temperature, and has allowed precise studies under unusual physical conditions.

Fusion of two nuclei can release a large amount of energy, but this reaction is difficult to achieve because of the electrostatic repulsion between nuclei.

In the most favorable case, two isotopes of the element hydrogen—deuterium (D with mass-2 nucleus d) and tritium (T with mass-3 nucleus t)—must be brought together about 10^4 times closer than they reside in a normal molecule. This can be accomplished in an unusual way by replacing one electron of the DT molecule by a negative muon (μ^-), which is similar to an electron but is unstable and 207 times heavier. The resulting bond distance shrinks by approximately this mass ratio, and the binding energy increases by a similar factor. The resulting small molecular ion is denoted $dt\mu$ (analogously, DT^+ could be written dte). Fusion then occurs in a time much shorter than the 2×10^{-6} s lifetime of the muon. The reaction produces an alpha particle (the nucleus of a helium atom) and a neutron. Because the muon accelerates the reaction and is usually released afterwards, the reaction is catalytic with multiple cycles.

The subject of muon-catalyzed fusion (μCF) has a fascinating history, beginning with its prediction by British theorist Sir Charles Frank in 1947. His paper in *Nature* [1] attracted the attention of Russian luminaries Andrei Sakharov and Yakov Zeldovich, who independently considered the prospects. [2,3] This work was unknown to Nobel Laureate Luis Alvarez in 1956 when he serendipitously discovered the reaction induced by cosmic rays in his bubble chamber at Berkeley. [4] The reaction he saw was fusion of a proton and deuteron but was too slow to be of practical interest.

Los Alamos Work Begins

In 1983, this situation dramatically changed and interest was rekindled by the observation, at the Los Alamos Meson Physics Facility (LAMPF) in a collaboration led by Steven Jones, of the μCF reaction between a *triton* and a deuteron. [5] As had been predicted by Russian theorists Semion Gershtein and Leonid Ponomarev in 1977, the *d-t* reaction is much more efficient than the *d-p* reaction, but the triton is radioactive and not ordinarily found in nature. [6] Los Alamos National Laboratory was uniquely suited for this experiment because of the concomitance of an intense muon source and tritium-handling capability. Laboratory theorists Melvin Leon and I were immediately called upon to help sort out the observations and further possibilities. The nation was just emerging from an energy crisis, and the Division of Advanced Energy Projects at the U.S. Department of Energy took notice and supported both the experimental and theoretical efforts, with the objective of determining whether μCF could be used to produce energy. The answer to this question rested on basic physics questions, which could be answered only by doing the experiments and testing the theory.

A number of Los Alamos theorists, including Gerald Hale, Richard Martin, Russell Pack, Michael Struensee, and me in the Theoretical Division, and Leon in the Meson Physics (MP) Division, contributed to the understanding of μCF. Though the emphasis in this narrative is on theory at Los Alamos, the work was very much an international collaboration. Theorists in the Union of Soviet Socialist Republics, Japan, Sweden, England, and at several American universities also made important contributions. Close ties were maintained with experimental collaborations, which worked at muon facilities in Russia (Joint Institute for Nuclear Research [JINR] and Leningrad Nuclear Physics Institute [LNPI]), Switzerland (Schweizerisches Institut für Nuklearforschung/Paul Scherrer Institut [SIN/PSI]), Canada (Tri-University Meson Facility [TRIUMF]), Japan (National Laboratory for High-Energy Physics [KEK]), and England (Rutherford Appleton Laboratory [RAL]), as well as at LAMPF.

The *d-t* μCF cycle (Fig. 1) starts with a free muon (directly from the accelerator or recycled from a previous fusion) being captured by one of the target molecules, which immediately dissociates into a normal atom and a muonic atom. With coworkers, I did the first dynamical calculations of capture for the *atomic* hydrogen target, starting in 1981 before the resurgence of μCF. [7–11] These calculations showed that the μ^- is slowed by collisions to a kinetic energy comparable to the target ionization energy and then captured into an excited orbital typically about the same size as the electron it displaces. The capture by an atom was found to be almost independent of the isotope. Later my results showed that capture by a *molecular* hydrogen target occurs before the μ^- is slowed down as much and has significant dependence on the target isotope. [12–13]

If initially captured by deuterium, the muon transfers to tritium, where it is more strongly bound. Postdoctoral student Struensee, together with Pack and me developed an improved adiabatic method for calculating cross sections for transfer between isotopes. [14–17] An unusual mechanism then enables the $t\mu$ atom to combine rapidly with a D_2 or DT molecule to form $dt\mu$. This reaction is resonant, due to a coincidence of energy levels made possible by the fortuitous existence of a state of $dt\mu$ so weakly bound that its binding energy can be taken up by vibrations and rotations of the host electronic molecule. In this process, the positively charged $dt\mu$ is so small that it acts like a mass-5 hydrogenic nucleus in the much larger electronic molecule to form a compound molecule (Fig. 1). Martin and I [18] corrected an error in early calculations, which had caused the calculated rates to be substantially too large. These were rather simplified calculations. The rigorous treatment of the formation of the compound molecule was finally carried out in 2000 by Pack, using his hyperspherical-coordinate formulation of reactive scattering (see the "Chemistry" chapter in this book), in collaboration with Vlado Zeman and Edward Armour of the University of Nottingham in England. [19]

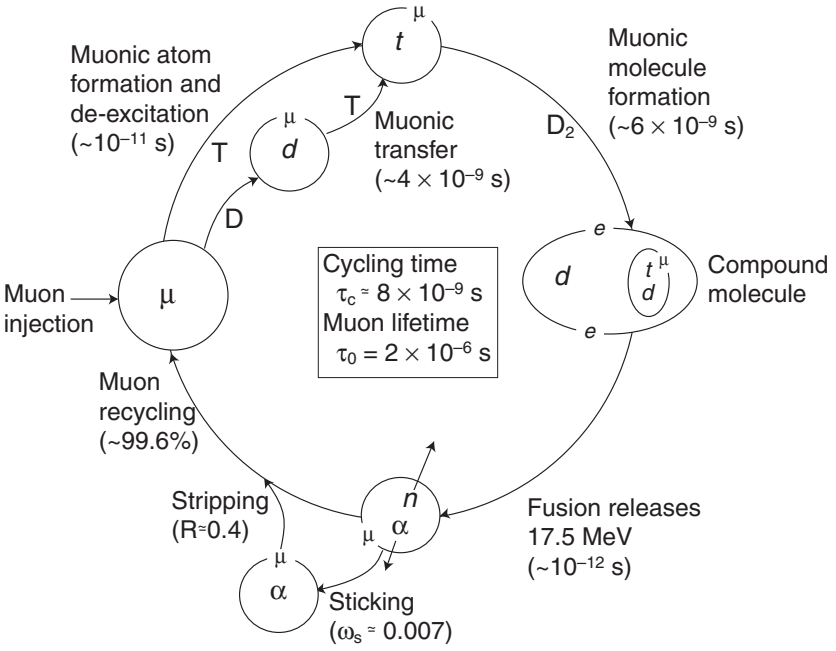


Fig. 1. Muon-catalyzed fusion cycle with reaction times in liquid deuterium (60%) and tritium (40%).

One of the big surprises in the μCF *d-t* experiments was a large transient cycling rate, i.e., fusion neutrons appeared at a more rapid rate at early times than in the later steady state. This effect was shown by Leon and me [20] to be due to the initially epithermal $t\mu$ atoms having a substantially larger resonant molecular-formation rate. Much later (in 2000 at TRIUMF), this enhanced rate would be directly observed in an experiment. [21] Though not yet experimentally proven, this phenomenon may provide a basis for increasing the fusion yield.

After the $dt\mu$ molecule is formed, nuclear fusion rapidly ensues. The improved adiabatic method was combined with the R-matrix method of Hale (see the "Nuclear Physics" chapter in this book) to do the first accurate calculation of nuclear fusion in the $dt\mu$ molecule. [22] With the total time required for the entire cycle designated τ_c, during the lifetime τ_0 of the muon there is time for τ_0/τ_c cycles, each cycle yielding one fusion. Usually the muon is left behind by the fusion products to begin another catalytic cycle. However, sometimes the muon is lost as a catalyst before it decays. This loss is primarily due to the probability ω_s that the muon sticks to a fusion alpha particle to form a muonic helium atom $\alpha\mu$. The effective ω_s is given by $\omega_s^0(1 - R)$, where ω_s^0 is the initial sticking probability and R is the conditional probability that the μ^- will be stripped off before the muonic helium ion is slowed down. Early experiments found ω_s to be smaller than calculated and to

have unexpected density dependence. The R-matrix method was adapted to calculate ω_s^0 and determine whether short-range nuclear forces could be responsible for the discrepancy, but the effect was found to be relatively small. [23] Calculations of the muonic helium kinetics [24] found some density dependence in *R* but not nearly as much as in the original experiment. Subsequent experiments have shown that the actual density dependence is consistent with theory, but a statistically significant discrepancy in the magnitude persists to this day.

Hence, the average number of fusions (neutron yield) Y_n that can be catalyzed by a single muon is given by $Y_n = 1/[(\tau_c/\tau_0) + \omega_s]$. The optimum value of τ_0/τ_c thus far attained experimentally at room temperature is about 250 with a 60:40 *d:t* mixture at liquid-hydrogen density. The measured value of ω_s is 0.004. Thus, we obtain $Y_n \approx 125$. The energy released by this number of fusions is not yet sufficient to recover the energy expended to make the muon. Each fusion yields 17.6 MeV of energy so that the total energy release per muon is 2,200 MeV. This is considerably greater than the rest of the mass energy of a muon, which is 106 MeV, but is still short of the estimated 8,000 MeV required to make the muon using known technology. Significant further improvement would require reduction of both τ_c and ω_s. We can only speculate on means of doing so, but might use the aforementioned epithermal molecular-formation rate to reduce τ_c and cyclotron acceleration of $\alpha\mu$ to enhance stripping and reduce ω_s.

Additional Applications

In addition to possible energy applications, the study of μCF has been fruitful for a number of reasons including the following:

(1) bridging the gap between atomic and nuclear physics;

(2) enabling nuclear reactions at room temperature;

(3) allowing precise studies under unusual physical conditions, e.g., violation of "mirror symmetry," which suggests that equal numbers of tritons and helium-3 should be produced in the *d-d* fusion reaction, and determination of fusion rates at very low collision energies; [25]

(4) observing a compound electronic-muonic molecular environment; and

(5) exhibiting phenomena spanning nine orders of magnitude in distance and energy.

Opacity Theory and Stellar Stability

by Arthur N. Cox

Cox
Huebner
H. Mayer
Oppenheimer
Sampson

In the 1940s, J. Robert Oppenheimer began stellar opacity and stellar stability research in the Theoretical Division. Since 1975, together with dozens of collaborators at Los Alamos and at many other research institutions worldwide, I have been involved in studies of stellar stability against pulsations. Such research has been appropriate for Los Alamos because it ties in with nuclear weapon design through the use of materials at high temperatures and pressures and because it involves studies of hydrodynamic instabilities. Observations of pulsations allow us to draw inferences about the internal structure of stars; and, fortunately, the many observed variable star classes allow astrophysicists to learn how stars evolve. The goal is to predict pulsation mode periods and their stability in stellar structures calculated from stellar evolution research. In many cases, pulsations can refine our understanding of what happens in deep layers and even at the stellar center.

This work continues research I started in 1960 in the Los Alamos Field Test Division (J-Division) and discussed, as a summer student in Physics Division, as early as 1947. "Nuclear physics of the muon-catalyzed d+d reactions."

In the 1960s, my work in J-Division was supported by many T-Division scientists, such as Harris Mayer, Doug Samson, and Walter Huebner.

More than a dozen variable star classes can be observed by recording their spectral and luminosity variations. In occasional cases, with the stellar distance known, at least approximately, even the actual mean luminosity can be determined. Three classes are discussed here:

(1) Mira variable stars with pulsation periods typically of years;

(2) our Sun, with the largest variations showing periods of about 5 min; and

(3) the white dwarf stars that pulsate with periods between 100 and 1,000 s.

Mira stars expand and contract, retaining their sphericity, whereas our Sun and the white dwarfs exhibit nonspherical, or so-called nonradial, modes. The very low-amplitude solar oscillations (only a few centimeters in the solar radius of 700,000 km) are driven by the random action of the Sun's turbulent convection to drive pulsation modes that involve acoustic or pressure forces. The white dwarfs display motions involving gravity forces. The solar p-mode and the white dwarf g-mode structures are not observed directly, but the understanding of their internal

structure and the observed period range of the oscillations lead astrophysicists to identify them using hydrodynamic theory.

Mira stars, known since 1596 for Mira itself, present a currently unsolved problem. Is the radial pulsation mode the fundamental with no internal zero-motion nodes, or does it pulsate in the first overtone with a static shell in the envelope layers with motions on both sides in opposite radial directions? Hydrodynamic calculations at Los Alamos and a few other institutions are difficult because these very cool (3,000 K) stars are vigorously convecting in most of their radii. It is unfortunate that the pulsation mode is not known for sure, because the Mira stars are very luminous and can be identified in many remote galaxies. The Cepheid stars have long been known as reasonably standard "candles" because they reveal their intrinsic luminosity by their pulsation periods. Thus, when a faint Cepheid varying with a period of, say, 20 days is observed, astrophysicists know its actual brightness or luminosity; and this, along with a Cepheid's observed brightness in the sky, can be used to determine rather accurately its real distance. Experts know that corrections must always be made for any dimming of the variable star's light due to light's absorption along its long path to us. Work continues to understand Mira star luminosities and even the detailed mechanisms that make them spontaneously expand and contract.

For decades, the convection-driven solar pulsation modes have been exciting to study to see how they restrict the solar internal structure. The Sun produces its radiative energy by central thermonuclear processes that also produce copious neutrinos. Observations, over many years, of these solar electron neutrinos have shown that a deficit of much more than half of them is expected by standard solar structure and evolution. Work at Los Alamos on both aspects of this problem has revealed that oscillation periods confirm the standard solar structure; and the neutrino deficiency problem needs to be, and is being, solved by advanced particle physics theory.

White dwarf pulsations reveal the state of very dense matter. These stars, which have burned all their hydrogen and helium fuel throughout their approximately 10-billion-year lives, are dead and are just cooling further. The cause of their g-mode pulsations has been debated for over 20 yr, with our Los Alamos efforts being at the forefront. Like the very cool Mira stars, the hotter (10,000 K) white dwarfs contain very active convection in their surface layers. Thus, understanding the effects of convection on white dwarf evolution and pulsation is most interesting. This unsolved problem is currently at the pulsation star research frontier.

To illustrate just part of the internal pulsation structure, Fig. 1 shows the so-called linear-theory normalized-radius amplitude and its phase as a function of the internal temperature. Note the displacement phase lag near the surface as the deep standing waves convert to running waves. From the surface at the right side of the figure to the center about 1,000 times hotter, there are six nodes for this g_6 mode. How the thermodynamics of the carbon and oxygen material in the convective envelope modulates the emergent luminosity to cause observed light variations remains uncertain. One approximation for how this operates has the plotted positive and negative work produce motions as opposed to position in the stellar envelope.

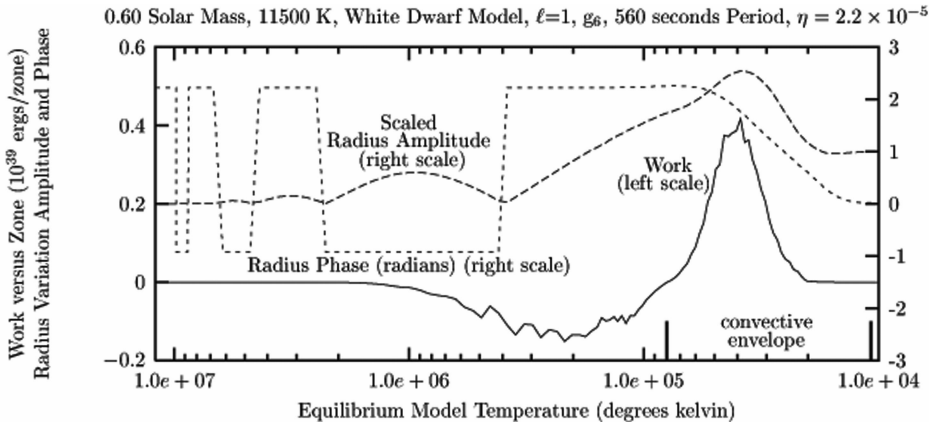


Fig. 1. Linear-theory normalized-radius amplitude and its phase as a function of internal temperature. The displacement phase lag occurs near the surface as the deep standing waves convert to running waves.

Chaos in Laser–Matter Interactions

by Peter W. Milonni

Ackerhalt
Boyd
Campbell
Farmer
Gaeta
Galbraith
Milonni
Shih

Chaos is unpredictable, random behavior in a system described by equations that do not include randomness. Theoretical Division researchers have studied the chaotic behavior of turbulent flows, lasers, nonlinear optics, the dynamics of atoms and molecules in laser fields, and laser radiation. They also investigated whether chaotic behavior can be controlled or suppressed and how orderly behavior becomes chaotic. Their analyses have compared well to experiments.

"Chaos" refers to unpredictable, random behavior in a system described nevertheless by deterministic equations, i.e., equations that do not include any randomness. Chaos is associated with an extreme sensitivity to initial conditions.

A common metaphor is the butterfly effect, which may be described as follows. Suppose one knows exactly the equations describing the evolution in time of the weather everywhere on Earth, and solves these equations using the most powerful computer imaginable. If the initial conditions are not known precisely—for instance, if the fluttering of a

butterfly somewhere is not taken into account—then detailed long-term weather prediction becomes fundamentally impossible if the system is chaotic. Such extreme sensitivity to initial conditions, or chaos, manifests itself as an erratic-time evolution that is indistinguishable from true randomness.

Chaotic behavior is of interest to the Los Alamos National Laboratory Theoretical Division for several reasons. It is important to know, for instance, whether the turbulent flow of a fluid is in fact an example of chaos and whether the chaotic behavior of a system can be controlled or suppressed. Chaos is a typical mode of behavior of nonlinear systems, and nonlinearity is the rule rather than the exception in most "real-world" systems. A simple example of nonlinearity occurs in a laser: If the electric current powering the laser is doubled, the laser power is not always simply doubled, as would be the case if it behaved linearly.

Chaotic behavior was studied intensively in T-Division in the 1980s; and the nonlinear character of laser-matter interactions strongly suggested that chaotic behavior should play a significant role in lasers, nonlinear optics, and the dynamics of atoms and molecules in laser fields.

Early Examples

One of the first examples considered was the interaction of a collection of two-level atoms with their own radiation field. [1] It was found that, in the so-called semiclassical approximation in which the photon character of the radiation field is ignored, the system evolves chaotically in time if the density of atoms is large enough. This example has remained of interest in connection with "quantum chaos," i.e., the study of the quantum mechanics of systems that exhibit chaos when treated semiclassically. An optical analog of the model has been realized experimentally and found to exhibit the predicted chaotic behavior.

A more practically oriented example studied was the multiple-photon excitation and dissociation of polyatomic molecules by laser radiation, a process studied extensively at Los Alamos and elsewhere during the 1970s in connection with laser-controlled isotope separation.

Experiments indicated that molecules could be dissociated in laser fields of relatively modest power, which was somewhat of a surprise.

Experiments also revealed that, over a wide range of conditions, the multiple-photon excitation process depends mainly on the fluence of the laser field, i.e., the total energy of the pulse, but not on the detailed intensity variations within the pulse. It was found that models of multiple-photon excitation exhibited chaos associated with the nonlinear nature of the molecular vibrations or with the coupling of molecular vibrations and rotations. As a consequence of the erratic evolution resulting from chaos, the excitation and dissociation were found to be

fluence-dependent for laser powers and laser pulse durations relevant to the experiments of interest. [2] In other words, the chaotic evolution of the molecular vibrations and rotations effectively prevented the molecule from coherently tracking the intensity variations in time, leaving the dissociation probability dependent only on the integrated intensity or fluence.

Good agreement with experiment was also obtained for the scaling of the dissociation probability with laser fluence. Specifically, the chaotic nature of the excitation and dissociation predicted by our models offered a plausible explanation of the fluence dependence observed in the experiments on laser isotope separation.

Important Developments

One of the important developments in chaos theory during the past quarter-century is the discovery of a few prevalent ways by which systems may make the transition from orderly behavior to chaos. These same "routes to chaos" have been observed in systems as diverse as lasers and fluid flows, to name but two. We studied various types of lasers that were known to behave "erratically." By demonstrating that this behavior followed from one of the known routes to chaos, we concluded that the observed behavior was in fact chaotic, i.e., it could be explained by a deterministic set of equations without any random inputs to the equations.

In one example, a so-called "bad-cavity" helium-xenon laser, hundreds of simultaneous differential equations were solved on the Cray machines available at Los Alamos and, in spite of the complexity of the system, good agreement was obtained with the experimental observations.

In a collaboration with Alexander Gaeta and Robert Boyd at the University of Rochester, we showed theoretically that the polarizations of two counterpropagating laser beams could vary chaotically if the laser intensities are sufficiently large. This behavior was later observed experimentally. [3]

The principal contributors in this work were Jay R. Ackerhalt, Harold W. Galbraith, Mei-Li Shih, and I. The first two were staff members, the third was a graduate student from the University of Arkansas, and the fourth was a visitor on sabbatical leave from Arkansas. The work benefited greatly from discussions with various staff members in T-Division, especially J. Doyne Farmer and David K. Campbell. *Chaos in Laser-Matter Interactions* (Peter W. Milonni, Mei-Li Shih, and Jay R. Ackerhalt [World Scientific Publishing Company, Singapore 1987]) resulted from an invitation from World Scientific and was based in part on lectures I gave at the Universidad de Puerto Rico and on the Ph.D. thesis of Mei-Li Shih.

The Los Alamos Suite of Atomic Physics and Kinetics Codes

Abdallah
R. Clark
R. Cowan
Mann

by Joseph Abdallah

Researchers in the Los Alamos Theoretical and Applied Physics (X-Division) divisions collaborated to produce a suite of atomic physics and kinetics codes that contributed to advances in atomic physics and plasma spectroscopy. The codes also have aided work in other areas, such as opacity, x-ray lasers, radiative power loss for magnetic fusion, diagnostics for laser removal of solid surfaces, astrophysics, electronic lithography, lighting systems, and calculation of wave functions.

The Los Alamos Theoretical Division–Applied Physics Division suite of atomic physics and kinetics codes has been widely used over the last decade or so to address issues in atomic physics and plasma spectroscopy. Over 70 papers and reports, mostly in refereed journals, have been produced based on the results from these codes. In addition, nearly 100 presentations at professional meetings, including various invited talks, have been given. Robert E. H. Clark of X-Division and I developed these codes.

One of the first and more memorable accomplishments was the use of the codes to confirm the earliest transmission measurements using laser-produced plasmas in the early nineties. The codes have also been used for a wide variety of other problems including opacity, atomic cross sections, x-ray lasers, radiative power loss for magnetic fusion, diagnostics and modeling for inertial confinement fusion, diagnostics for laser ablation of solid surfaces, electron impact coherence parameters for plasma polarization spectroscopy, astrophysics, analysis of plasmas created by pulsed-power devices such as the X-pinch and plasma focus, plasma spectroscopy of hollow atoms, hot electron effects in various types of plasmas, analysis of laboratory space plasmas, nonequilibrium modeling of high-Z plasmas, high-intensity laser irradiation of clusters, lithography, plasma display panels, and lighting systems.

The original motivation in the late 1980s for the Los Alamos suite of codes was to provide a nonequilibrium plasma-modeling and opacity capability. Since large amounts of atomic physics data are required for such calculations, it was decided to use the codes of Cowan and Mann, whose work is described elsewhere in this volume, to generate this data. This decision was made because published experimental and theoretical data was usually incomplete and coupling schemes were inconsistent. Also, the codes of Cowan and Mann provided accurate data for our purposes and were relatively fast compared to other codes available at the time. In addition, we had the great fortune of having both men available for consultation on a daily basis.

Cowan, Mann, and Beyond

The codes of Cowan and Mann, were modified to run more robustly, provide more user-friendly interfaces, and output random access files capable of accommodating large amounts of atomic physics data that could be read directly in a plasma modeling code, while maintaining the integrity of the physics in the original codes. RCN, RCN2, and RCG were combined to form the CATS code, while Mann's excitation and ionization codes were the starting points of the ACE and GIPPER codes, respectively. CATS is used to calculate wave functions, energy levels, oscillator strengths, and plane-wave Born collision strengths. ACE is used to calculate electron-impact excitation cross sections, and GIPPER is used to calculate electron impact ionization, photon impact ionization, and auto-ionization cross sections.

The FINE code was developed from scratch to calculate the populations of atomic states, number per volume, by solving the rate equations under various plasma conditions. Input to FINE consisted of all the atomic physics data discussed here. Rate coefficients for the various processes are calculated by integrating the cross sections discussed here and their reverse process over the electron energy distribution and radiation fields dictated by the imposed plasma conditions. These coefficients are used in setting up and solving the rate equations; the populations are extracted and used to compute the desired plasma radiation property. The properties usually of interest are the plasma emission (light directly emitted) or the opacity, which is the probably of light absorption.

The LINES code was written to take advantage of simplifications that occur for equilibrium plasmas where no collisional cross sections are required. However, all the capabilities of LINES have since been included in FINE. (See Fig. 1.)

More recent additions to the suite of codes have been the development of the RATS and GRATS codes, which are more suitable for generating atomic data for high-Z elements because of the inclusion of relativistic effects. Usually, RATS is used for high-Z elements instead of the CATS code. ACE and GIPPER have also been modified to accept data from RATS and, hence, yield a full complement of atomic cross sections.

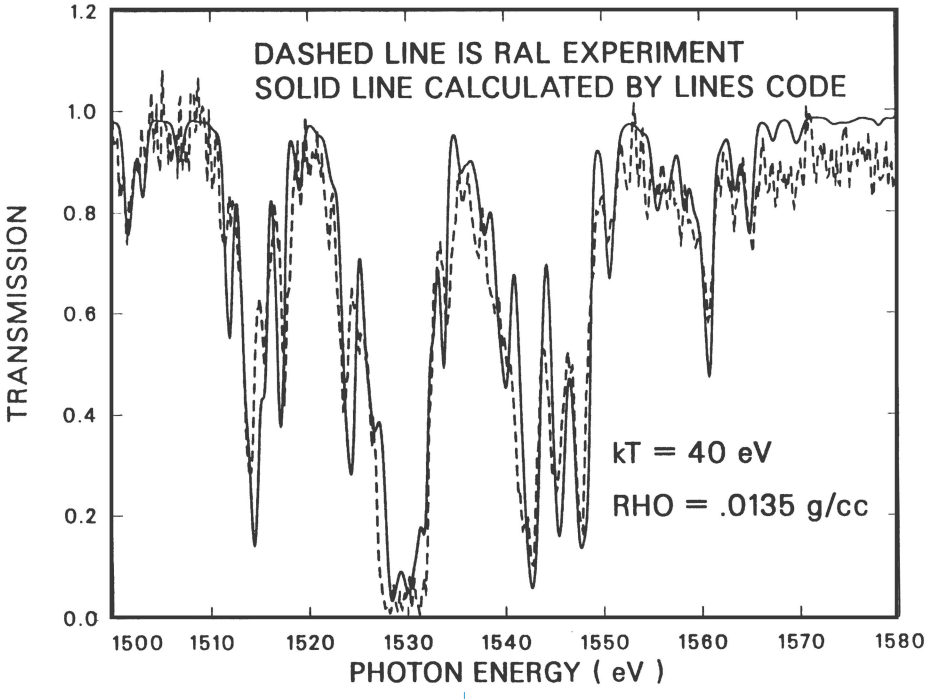


Fig. 1. A comparison of the experimental and calculated transmission fraction. Note the complexity of the structure due to x-ray line absorption and the excellent agreement.

In addition, FINE has been modified to process the relativistic models to provide a unique high-Z kinetics capability. Another enhancement to FINE is the inclusion of unresolved transition array theory that provides a more realistic treatment of spectra lines when using configuration average physics.

Contributions to Other Areas

The codes provided the first successful simulations of plasma transmission experiments. [1–3] The calculations gave confidence to the experimentalist to proceed with additional opacity measurements. Here, two laser beams are used, one to create a tamped aluminum plasma and the other to provide an x-ray source. The x-rays are used to irradiate the aluminum plasma; and the intensity of the x-rays are diminished by plasma absorption, which is measured after passing through the sample. The experimental result provides the fraction of radiation transmitted as a function of photon energy. The CATS code was used to calculate atomic data for several hundred configurations each for the sodium-like through hydrogen-like ionization stages of aluminum. These resulted in thousands of fine-structure atomic-energy levels and millions of transition probabilities in some of the more complicated ion stages. GIPPER was used to calculate the photo-ionization cross sections. The idea was to calculate all the important processes capable of absorbing an x-ray. The LINES code used a local thermodynamic equilibrium model to calculate the populations of atomic energy levels at a given temperature and density. The calculated populations predict the resulting transmission spectra. This was a valid approximation because of the relatively high electron density that was encountered.

As a second example, we consider the large plasma device (LAPD) at the University of California at Los Angeles a linear plasma research facility primarily designed to study space and fusion plasma processes. A 2.95-kiloamperes (kA) discharge current with a voltage near 50 volts (V) is used to create a 10-m-long plasma column with a diameter of 40 cm. A magnetic field of 1,200 gauss (G) is applied to confine the plasma. The plasma which is formed from the interaction of the electrical discharge with the chamber gas, usually helium or argon, has densities in the 10^{12}–10^{13} cm^{-3} range and electron temperatures in the 5–10 eV range. The geometry is well suited for making localized measurements of the plasma conditions. At early times in the discharge, the electron density builds up quickly due to the interaction of energetic primary discharge electrons, with the target gas producing secondary electrons from electron impact ionization. In this phase, the electron energy distribution is highly non-Maxwellian, and the ionization rate of the plasma is unusually rapid. The bulk of the density is quickly formed by secondary electrons, which are Maxwellian in nature. The subsequent ionization is then determined by the action of these thermal electrons. Ionization in the plasma continues until the discharge current is switched off.

Agreement with Experiments

The electron density and electron temperature have been measured in an LAPD helium plasma at various times and locations during the discharge. These measurements have been compared to the results of detailed time-dependent atomic kinetics calculations [4] using the Los Alamos codes. The codes are used to simulate a local volume element of the forming plasma. The kinetics calculations must include all the important atomic processes that can alter the populations of atomic levels existing in the plasma. Previous calculations for helium have considered only the steady state, time invariant case. Here, 300 coupled-rate equations are solved simultaneously as a function of time to predict level populations and the free-electron density using the measured temperature profile as input. The initial conditions for the integration were determined from the measured gas density of 2.6×10^{12} cm^{-3} the electron density at the earliest time point. Thus, the calculation involved no adjustable parameters. The calculated electron densities are compared to the measured values for radial positions of 0, 9, and 18 cm from the center of the plasma column in Fig. 2. The results are in excellent agreement with experiments. The results of the calculations provide a confirmation of the plasma conditions, an understanding of the important mechanisms involved in LAPD plasma formation, and a validation of the atomic physics models used in the codes.

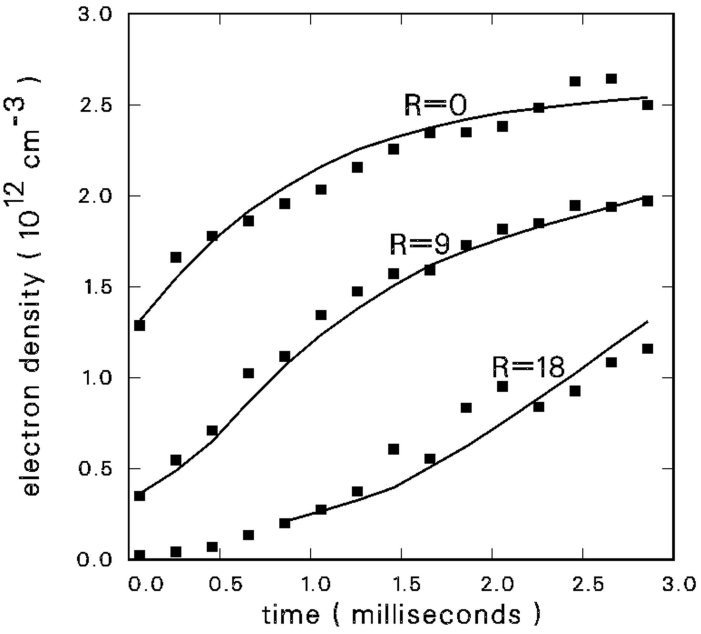


Fig. 2. Comparison of calculated electron densities to measured values for radial positions of 0, 9, and 18 cm from the center of an LAPD helium plasma.

Finally, the codes have been used to analyze various x-ray spectra resulting from the high-power, short-pulse laser irradiation of argon clusters targets. [5–10] (See Fig. 3) The interaction of such laser pulses with atomic gas clusters has been a subject of active investigation for a number of years for fusion and continues to be studied for a variety of applications to gain insight into the behavior of matter under extreme conditions. For example, laser irradiated clusters are relatively inexpensive, compact sources of extreme ultraviolet radiation (EUV) and x-rays. In addition, plasma conditions comparable to those obtained in large-scale experiments, such as inertial confinement fusion and Z-pinch, can be achieved by using relatively small experimental setups. In fact, high-plasma

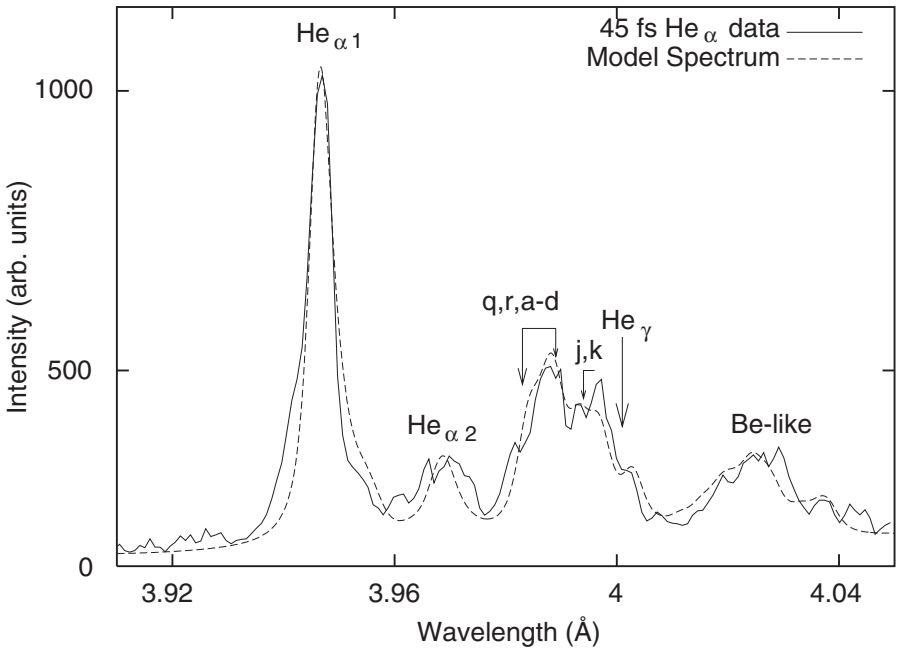


Fig. 3. Comparison of model calculations with a spectrum observed in an experiment involving the short pulse laser irradiation of argon clusters.

temperatures and densities distinguish laser interactions with clusters from interactions with gas targets that absorb laser light inefficiently. Our code calculations have shown for this case and other short-pulse laser experiments involving solid targets that the inclusion of hot nonthermal electrons is necessary to reproduce the spectral features. Also, the calculations show that the best fits to experiment combine a high-density component corresponding to a cluster core and a low-density temperature component roughly corresponding to its surroundings.

Electron–Ion Collision Studies

by George Csanak and Walter F. Huebner

Abdallah
Burgess
R. Cowan
Csanak
Huebner
Keady
Lyons
Mann
Merts

The need for electron-ion excitation cross sections came from two sources: (1) the need for nonequilibrium atomic physics data and (2) the need for energy-loss data for magnetic-fusion experiments. This latter occurred in the mid-1970s when tokamaks (plasma-containing devices) began losing radiative power. In answer to these needs, Theoretical Division researchers developed models that had and continue to have substantial influence on the United States magnetic-fusion program and the collision and fusion physics community.

The atomic physicists in Theoretical Division in the 1960s were mainly involved in performing Rosseland and Planck mean opacity and equation-of-state calculations for matter in local thermodynamic equilibrium (LTE). [1,2] However, some atomic physics processes occur on time scales that are shorter than the time it takes to establish LTE. For example, for a plasma at temperature much below the energy required to produce electron-positron pairs (510 keV), the photons in the high-energy tail of the Planck distribution can produce electron-hole pairs at a slow rate. These electron-positron pairs act as scattering centers for the more abundant lower-energy photons and can dominate the opacity at LTE. As a second example, radiative de-excitation in medium-to-heavy elements can be faster than LTE-determining collisional excitation/de-excitation. Thus, under some conditions LTE is never achieved.

Realizing that understanding the atomic physics for some very fast processes can be important, in the mid-1970s, T-Division expanded its research capabilities into the field of non-LTE. Joe Abdallah developed a non-LTE code using rate coefficients calculated from photon and electron collision cross sections. The work on rate coefficients was

expanded to include molecular processes. One of the much-cited works in this area is the paper by Huebner, Keady, and Lyons, [3] which was also published as a book. Subsequently, R. D. Cowan initiated a substantial effort in the area of atomic spectroscopy, [4] which also aided the opacity effort.

Serious Needs Draw in Theoretical Division

The need for electron-ion excitation and ionization cross sections came from two sources: one was the need for nonequilibrium atomic physics data, and the other was the need for stopping power (energy-loss) data for magnetic-fusion experiments. The nonequilibrium atomic data was mainly needed to investigate coronal equilibrium. There are also situations of great Laboratory interest where heavy-element atomic data (including electron-impact cross sections) are needed in temperature regions where LTE cannot be attained in the plasma. The real incentive for electron-ion cross-section calculations came from the magnetic-fusion program. In the mid-1970s T-Division was asked by high-level representatives of the United States Government to look into the radiative power-loss problem of the Princeton University tokamak and the Oak Ridge National Laboratory tokamak.

In magnetic-fusion plasmas, heavy-element atoms from the walls, the injector tubes, and other parts of the container are sputtered into the plasma. These atoms become totally or partially ionized due to their collisions with plasma-electrons, which also excite them. The power loss due to line radiation from the partially ionized and excited atoms becomes a serious problem when the concentration of heavy elements is as large as 1%. In two important reports, [5,6] the T-Division researchers have shown that the presence of tungsten, gold, iron, and other heavy elements in the magnetic-fusion device can be detrimental to the outcome of the experiment. These two reports exerted a substantial influence on the U.S. magnetic-fusion program. One of these reports [5] also contained a detailed discussion of the dielectronic recombination process in plasmas. The process culminated with the modification of a formula introduced by A. Burgess, which subsequently has been referred to in the literature as the Burgess-Merts formula and was used extensively in the plasma-fusion community.

Need for Codes

The two studies also pointed out the need for the development of reliable and practical computer codes for the calculation of electron-ion excitation and ionization cross sections. This led to the development of the distorted wave approximation (DWA) excitation and ionization codes by J. B. Mann. These codes also took advantage of the already developed spectroscopic codes of R. D. Cowan to generate the wave functions of the initial and final target states. The DWA is a simple approximation scheme that still might be expected to give reliable cross sections for most excitations and ionizations.

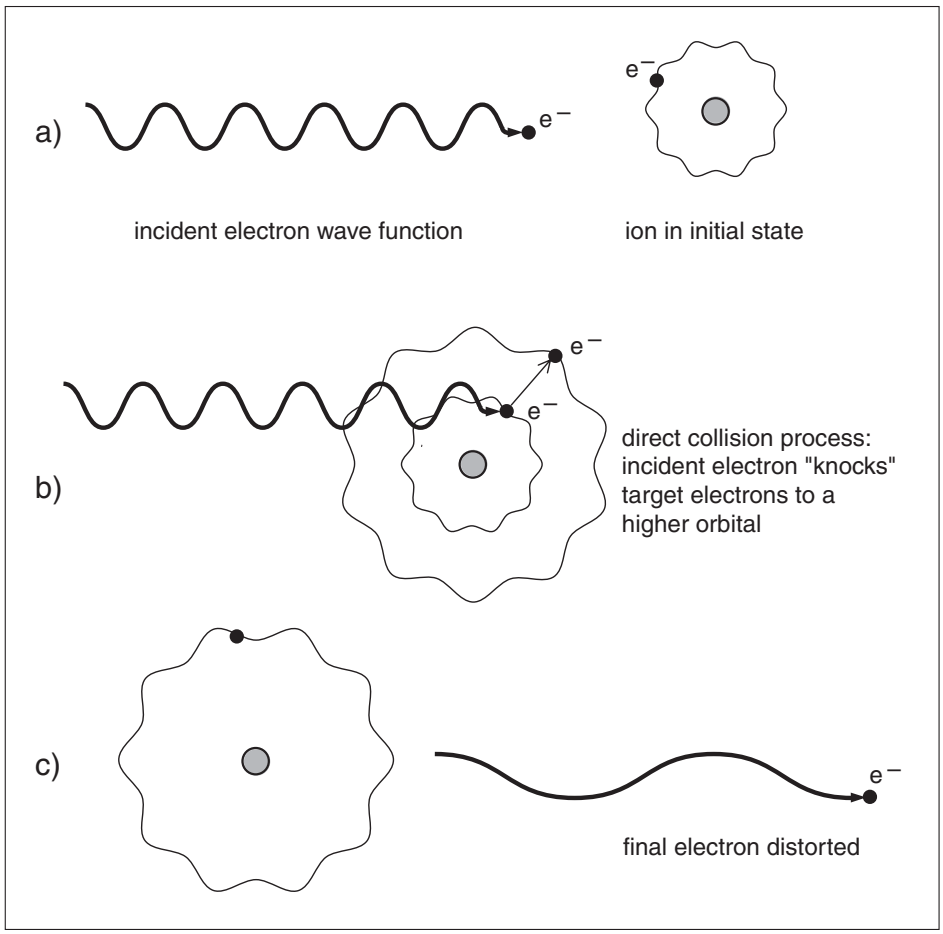


Fig. 1. For the excitation process, DWA is a one-step approximation scheme. In this one-step scheme we envision that the wave function of the incident electron is distorted by the initial target state (Fig. 1a). Subsequently, the electron in this distorted state approaches the ion and, in a single-step process, excites it to its final state (Fig. 1b). Subsequently the electron "leaves," moving in the field of the final state and being distorted by it (Fig. 1c). These intuitive elements are incorporated into the DWA formalism by the appropriate application of quantum mechanics.

For the excitation process, DWA is a one-step approximation scheme. In this one-step scheme we envision that the wave function of the incident electron is distorted by the initial target state (Fig. 1a). Subsequently, the electron in this distorted state approaches the ion and, in a single-step process, excites it to its final state (Fig. 1b). Subsequently, the electron "leaves," moving in the field of the final state and being distorted by it (Fig. 1c). These intuitive elements are incorporated into the DWA formalism by the appropriate application of quantum mechanics.

The Los Alamos DWA code produced many valuable cross sections for electron-ion collisions that were used for magnetic fusion and opacity applications. Subsequently, R. D. Cowan and J. B. Mann realized that the DWA may not give always reliable cross sections for electron-impact ionization. These cases include those when the incident electron is able to excite auto-ionizing levels of the target ion. In those cases the DWA cross sections have to be supplemented by excitation-auto-ionization cross sections. For high temperatures and for highly charged ions, the contribution from excitation-ionization could be several times larger than the DWA cross sections.

Significant Model

In the case of electron-impact excitation of ions, Cowan observed that the DWA cross section should be supplanted by the calculation of resonance cross sections, when the incident electron would be trapped by the ion and one of the target electrons is simultaneously excited to a higher level, thus forming a temporary doubly excited state called the autoionizing state. Subsequently, this autoionizing state can decay by emitting one or two electrons. Such a process can be included in the excitation process as an independent process. With these models, Cowan and Mann were able to improve results for ionization and excitation processes in many cases. Their model is still being used by the collision and fusion physics community.

The Theory of Atomic Structure and Spectra

by Robert D. Cowan

Andrew
R. Cowan
Edlén
Elander
Fawcett
Griffin
Hansen
Herman
Källne
Keady
Magee
Mann
Mark
Merts
Peacock
Pindzola
Racah
Robb
Skillman
M. Wilson

This article recounts my work in the above field during my last 25 yr (1960–85) as a Laboratory employee in Theoretical Division and during subsequent retirement years. Much of the work was done jointly with a large number of colleagues and collaborators both at Los Alamos National Laboratory and worldwide. The light emitted or absorbed by atoms (or ions) in a hot, vaporized source consists of certain discrete wavelengths known as* spectrum lines. *The wavelengths are characteristic of the element involved; therefore, the spectrum provides information on the internal structure of the atom, and spectral information can be used to deduce properties of (vaporized) bulk material such as chemical composition and temperature of both laboratory and astrophysical light sources. I have developed codes that have been used for moderately accurate, large-volume, survey-type work. Various versions have been distributed to 107 groups in 27 states and to 164 groups in 34 foreign countries.

Experimental Background

The light emitted or absorbed by atoms (or ions) in a hot, vaporized source consists of certain discrete wavelengths known as *spectrum lines*. The wavelengths are characteristic of the element involved; therefore, the spectrum provides information on the internal structure of the atom, and spectral information can be used to deduce properties of (vaporized) bulk material such as chemical composition and temperature of both laboratory and astrophysical light sources. For example, by means of comparison with laboratory-observed spectra, lines seen in the solar corona were shown in 1945 by Bengt Edlén (Lunds Universitet, Sweden) to arise from 9- to 14-fold ionized iron and from similar ions of calcium and nickel, thus showing that those elements existed in the corona and that effective temperatures there were far higher than in the photosphere, where only neutral or singly ionized atoms are present.

The reason that only discrete wavelengths are present is associated with the fact that the electron cloud around the atomic nucleus can exist only in states having certain discrete energy values ("energy levels"). The relationship between spectrum lines and the energy-level structure of an atom is illustrated in Fig. 1. When an electron in the atom changes its state so that there is a transition from one level to another, the resulting change in energy of the atom appears as emission or absorption of light having a correspondingly discrete frequency or wavelength.

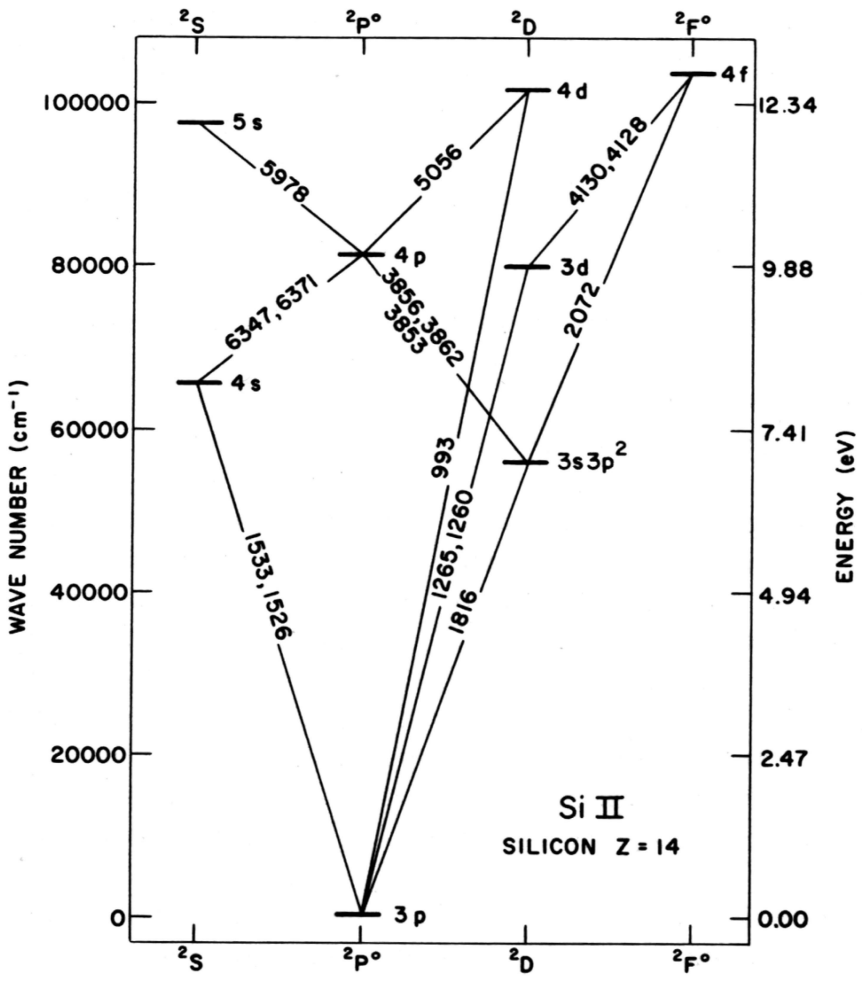


Fig. 1. The relationship between spectrum lines and the energy-level structure of singly ionized silicon atoms. The numbers on the lines representing radiative transitions are wavelengths in Ångstroms.

Stated in reverse, the problem of experimental spectrum analysis is the following. Starting from a list of observed spectrum wavelengths (hopefully, all produced by a single ionization stage of the atom under study), one attempts to fit the corresponding transition energies into a consistent and reasonably complete array of energy levels. Along with this, one tries to infer for each energy level the associated electron distribution ("electron configuration") and values of certain electron angular momenta. This process is guided by the so-called vector model of the atom and other theoretical considerations, taking into account also certain selection rules that determine whether any specific pair of levels can be connected by a radiative transition.

Initial Impetus for Los Alamos Work on Atomic Structure

In about 1957, then Theoretical Division leader Carson Mark asked me to see if I could aid members of the Controlled Thermonuclear Reaction Division in identifying (with regard to element and ion stage) spectrum lines observed in their magnetic confinement device, lines that were not included in tables of known spectrum lines. At that time, I knew little about spectral theory and was busy on other unfinished projects, so I was of no help.

However, in early 1960, while looking for a new project on which to work, Ken Andrew—a close friend from undergraduate and graduate days and by then professor and head of the spectroscopy laboratory at Purdue University—came to me for help in the analysis of spectra of germanium, a project on which he was working. In those days, the financial situation and academic atmosphere of the Laboratory were such that I was free to choose my own research topics, and I plunged 100% into the task of learning the quantum-mechanical theory of atomic structure from the ground up. This project was undertaken solely because it was of interest to me and Ken, with no particular thought of possible laboratory applications other than the one mentioned in the preceding paragraph.

Basic Quantum-Mechanical Theory

The main contributions to the energy of a stationary state of an atom are

(1) the electrostatic interactions between each electron and the nucleus and between each pair of electrons, and

(2) the magnetic interaction between the orbital motion of each electron and its own electron spin (the spin-orbit interaction).

Briefly, the theory is built upon the concept of an "electron configuration," which is defined by specifying the nl value for each electron—n and l being integers called the principal and angular-momentum quantum numbers, respectively; the angular distribution of the electron cloud about the nucleus is determined by the values of l, and the radial distribution (the variation with distance from the nucleus) is determined by both the n and l values. The theoretical expressions for the total energy and other properties of the atom (such as radiative transition rates) correspondingly involve two kinds of quantities: those arising from the angular distribution and orbital motion of the electrons and those involving the radial distribution of the electrons.

Progress of Theoretical Work

Values of the angular quantities for configurations that Ken and I initially needed were available in the journal literature. It had long been customary to treat the radial quantities as adjustable parameters to be evaluated by least-squares fitting of experimental energy levels that had been determined by classical spectrum analysis as described previously.

Ken provided me with a copy of notes on least-squares fitting procedures that had been left by the Italian theorist Giulio Racah during a recent visit to Argonne National Laboratory. I coded an appropriate FORTRAN computer program, the use of which proved to provide good fits to Ken's experimental levels. Thus I simultaneously established the suitability and accuracy of the basic theory, verified

(or corrected) Ken's interpretation of his levels in terms of the electron configuration and angular momenta of the various quantum states and predicted values for levels that he had not yet found experimentally, thereby providing valuable guidance in searching for unidentified levels.

It soon became clear that we would need angular factors for other electron configurations for which values were not available in the literature. Thus, it was necessary to develop an appropriate way of computing such factors for perfectly general configurations. This project involved extending methods that Racah had expounded in three mid-1940s *Physical Review* articles—methods now commonly referred to as Racah algebra. After a year's digression working in the weapons test group on spectroscopic collisional-ionization and eyeburn problems connected with the Redwing series of high-altitude nuclear tests, I returned in 1963 to coding a single-configuration version of a program for calculating the necessary input angular quantities to the least-squares-fitting program.

The following year, I obtained a copy of a program developed by Herman and Skillman of Lockheed Research Laboratories and RCA Laboratories for the calculation of approximate one-electron atomic radial distribution functions ("wave functions"). I elaborated on this and adapted it to my needs for theoretical calculation of the radial quantities mentioned previously, thereby producing preliminary versions of a set of programs for the *ab initio* calculation of atomic energy levels, as opposed to simply the fitting of experimental levels.

I had barely finished this stage of research when I was contacted by Nick Peacock, who was spending a sabbatical year in the Controlled Thermonuclear Research (CTR) Division, on leave from the Culham Laboratory in England. It appeared that people at the U.S. Naval Research Laboratory in Washington, D.C., had observed spectra of the sun in the far ultraviolet region by using rockets to send spectrographs above the Earth's absorbing atmosphere. They had noticed, among many weaker lines, a striking set of four equally strong, almost uniformly spaced spectrum lines with wavelengths in the vicinity of 170 angstroms (1.70 millionths of a centimeter, or 0.67 millionths of an inch—about one thirtieth the wavelength of green light). These lines had never been interpreted by laboratory work. Peacock's coworkers at Culham had seen these lines in the spectra of their CTR device (which used a stainless steel discharge tube) and in intense-spark iron spectra. The lines had also been observed in Los Alamos CTR devices when iron penta-carbonyl was inserted into the plasma. The lines were thus clearly due to iron, but guesses as to the ion stage responsible ranged from about 5- to 15-fold ionized. The task of trying to identify the lines by means of *ab initio* calculations with my codes was slow and frustrating for a number of reasons, including the fact that the four strong lines, though close together in wavelength, proved to arise not from one ion stage but from 8- to 10-fold ionized iron.

Nick's and my excitement, when we finally succeeded, was short lived. Before we could even write up our results, we learned that the Culham people had finally been able to arrive at the correct identifications by purely empirical means. Nonetheless, our experience established that predictions of atomic energy levels and spectra could be made with reasonable accuracy by means of *ab initio* calculations. I believe that this was the first such accomplishment for moderately complex atoms by means of computer calculations. [1,2]
A comparison of calculation with experiment for emission spectra of 30- to 32-fold ionized molybdenum is shown in Fig. 2—the collaboration again being with a foreign visitor to the CTR Division, Elizabeth Källne from Sweden. [3]

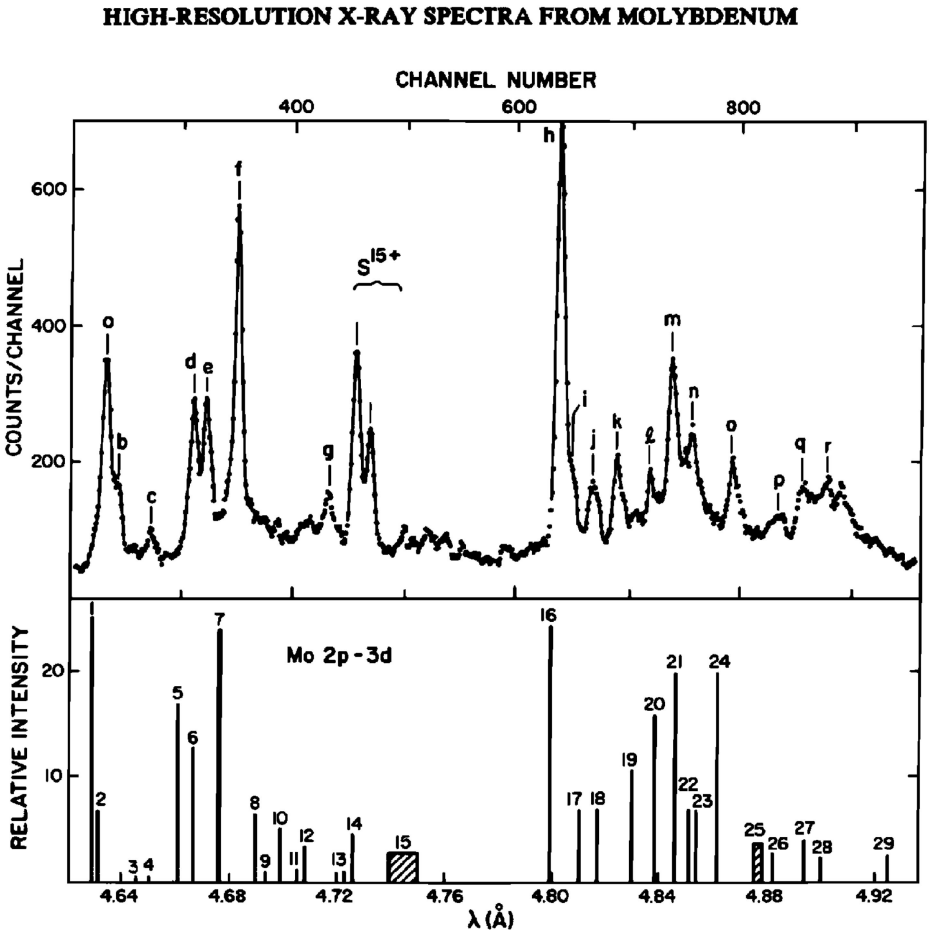


Fig. 2. Comparison of experimental and theoretical x-ray spectra in the region of Mo 2*p*-3*d* emission. Peaks in the experimental spectrum have been keyed with letters and the calculated peaks with numbers. The experimental wavelength scale has been normalized to the peak (*h*) at 4.804 Å. The experimental spectrum is an addition of 12 shots at $N_e = 3.0 \times 10^{14}$ cm^{-3} and $T_e \approx 1.4$ keV.

Figure 3 shows a comparison for inner-shell (2*p*; i.e., $n = 2$, $l = 1$) photoabsorption in neutral chlorine, the work having been done by collaboration with groups from the Universities of Central Florida and Wisconsin and the Max-Planck Institute in Göttingen, Germany; in this case, a suitable line shape has been convolved with each computed spectrum line to facilitate comparison of calculation with experiment. [4]

Expansion of Work

News of early successes spread by word of mouth and published journal articles. The result was a long series of collaborations over a space of many years with groups not only at Los Alamos, but also at the following: Argonne National Laboatory, the National Bureau of Standards (now National Institute of Standards and Technology [NIST]), with Brian Fawcett at the Culham and Rutherford/Appleton Laboratories in England, at Lunds Universitet in Sweden, and at several places in Germany (all dealing with basic atomic spectrum analysis); at the U.S. Naval Research Laboratory in Washington (primarily on astrophysical applications, such as opacities and radiation transfer, required for calculation of stellar structures, measurement of densities and heavy element abundances in stars, study of radiation from solar flares); at Stanford University (calculations on quasi-metastable excited states

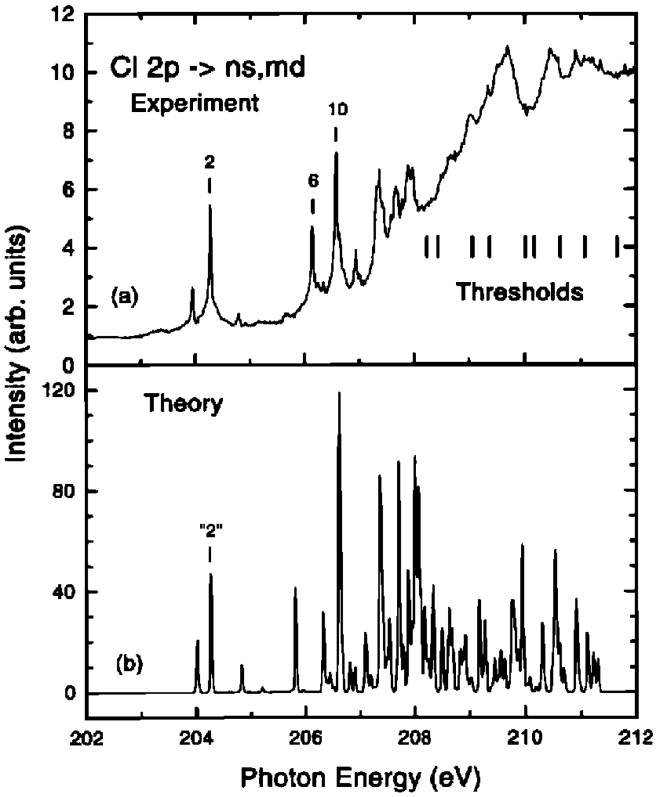


Fig. 3. Comparison of the experimental and the theoretical Cl 2*p* spectrum. The calculated spectrum comprises all transitions to 4*s*–9*s* and 3*d*–8*d* levels; it has been convoluted with a Gaussian of 40-meV width and has been shifted by –0.812 eV to match the experimental spectrum at line 2. The computed ionization thresholds are given in the upper panel, except for the highest threshold at 213.11 eV; these thresholds were not shifted.

with application to laser development); with Mike Wilson at University of London Royal Holloway Laboratory and Bedford New College in England and with Jorgen Hansen at the Van der Waals–Zeeman Institute in Amsterdam (assorted atomic processes, including stability and structure of negative ions); and with many others.

Numerous possible applications and requests by interested scientists provided the incentive to add new features to my programs, including conversion of the radial-wave-function program to more-accurate methods, addition of approximate relativistic corrections to the radial wave functions (done largely by Don Griffin, a student of Andrew's spending a summer with me in Los Alamos), addition of a means of computing radial wave functions for unbound electrons so as to be able to calculate autoionization transition and dielectronic recombination rates (needed for calculation of relative abundances of different ionization stages in plasmas), extension of the angular-momentum code to handle multiple interacting configurations and to compute magnetic-dipole and electric-quadrupole as well as electric-dipole transitions, and (with Derek Robb of T-Division) addition of options to compute electron-collisional excitation cross sections and rates in the plane-wave-Born approximation.

Required input to the programs is a bare minimum: the atomic number of the element and the ion stage of interest (thereby specifying the number of electrons in the system), and the *nl* values for each electron in each of the configurations to be included in the calculation. Output includes the radial wave function and binding energy for each electron in each configuration; the total energy of each configuration; the angular and radial quantities needed for calculating energy levels; the levels themselves, along with their angular-momentum properties; a list of the wavelengths and energies of the possible spectrum lines, together with radiative transition rates and level lifetimes. If desired, low-probability magnetic-dipole and electric-quadrupole transitions can be computed in addition to the normal high-probability electric-dipole transitions. Calculation of plane-wave-Born electron collisional rates is an option, as is calculation of autoionization rates if unbound electrons are included in the problem.

The programs can profitably be run for quite complicated systems on even very modest desktop and laptop computers; for example, a 10-configuration run for neutral germanium (atomic number 32) producing 84 levels and 1,097 spectrum lines can be run in 7 or 8 min, including setup time and 2 min of computer time on a medium-speed desktop machine. Crucial to this success has been the use of Racah algebra in order to make the energy-level/spectrum code small, fast, and capable of handling completely general multiconfiguration cases.

More accurate methods of treating detailed atomic-structure problems exist, but they require a much greater expenditure of personal setup effort and computer time. My codes are very good for moderately accurate, large-volume, survey-type work. They have become sufficiently versatile, user-friendly, and popular that from 1966 to date I know copies of various versions to have been distributed to 107 groups in 27 states and to 164 groups in 34 foreign countries—primarily Great Britain, France, Germany, Sweden, China, and Japan, but also Canada, Belgium, The Netherlands, Spain, Italy, Poland, Russia, Israel, India, and New Zealand.

In early days, I made the distribution personally by means of copies on magnetic tape, and later by means of floppy diskettes or e-mail. In 1992 the FORTRAN source programs were made freely accessible by means of downloading from the T-Division public Web site (ftp://aphysics.lanl.gov/pub/cowan). It is unknown how many additional copies have been obtained by this last means or by person-to-person transfer, but I still learn of half-a-dozen new users per year, and spend an appreciable fraction of my time answering e-mail requests for help and advice from both old and new users.

Examples of Applications

A notable application to a problem of national interest concerned radiation loss in tokamak CTR machines, resulting from high-Z (high-atomic-number) atoms knocked out of the containing walls by the electrical discharge. In 1975, several people from the Princeton Plasma Physics Laboratory visited T-Division looking for help on the theoretical calculation of such losses. Their use of a crude hydrogenic approximation for atomic structure omitted any radiation loss by means of delta-n = 0 transitions (ones with no change in principal quantum number) because this approximation gives zero energy separation for states with given principal quantum number n but different angular momenta l. We in T-Division, including Al Merts and Norm Magee, were in a position to immediately make calculations that included important delta-n = 0 losses. We found losses could be severe if even iron impurity atoms were present in the tokamak discharge [5]—let alone the impurities introduced by the much-higher-Z molybdenum and tungsten components that had been in use. Design modifications of tokamaks were thus called for to eliminate the use of tungsten and reduce the concentration of molybdenum impurity atoms in the plasma. Also, renewed emphasis was placed on the laser-implosion CTR approach.

Related work for CTR applications is being done by Don Griffin (Rollins College, Florida) and Mitch Pindzola (Auburn University), who use my programs to produce wave-function and energy-level input to their collisional-excitation codes, which use a more accurate distorted-wave method. They work closely with members of the Atomic Data and Analysis Structure program at the University of Strathclyde, Glasgow, which produces databases for plasma modeling calculations for Culham's Joint European Torus (JET), a large CTR device, and for stellar-structure calculations. My codes are also used in this project to produce large numbers of energy levels, radiation rates, and approximate (plane-wave-Born) collision strengths where more accurate results are not available.

The codes have also been applied at Los Alamos to T-Division molecular structure calculations and to Chemistry Division observations on the spectroscopy of crystalline solids containing actinide elements.

A more basic research application was calculations concerning the electronic structure of hypothetical super-heavy elements. My calculations with Joe Mann of T-Division showed with reasonable confidence [6] that 5-g electrons ($n = 5$, $l = 4$) should be present in the lowest-energy configuration of atoms with atomic numbers greater than 124, as a consequence of collapse (sudden radial contraction) of g-electron wavefunctions with increasing atomic numbers, analogous to the collapse of 3*d*, 4*d*, and 5*d* electrons ($l = 2$) at the start of the transition-element series and the collapse of 4*f* and 5*f* electrons ($l = 3$) at the beginnings of the lanthanide and actinide series of elements.

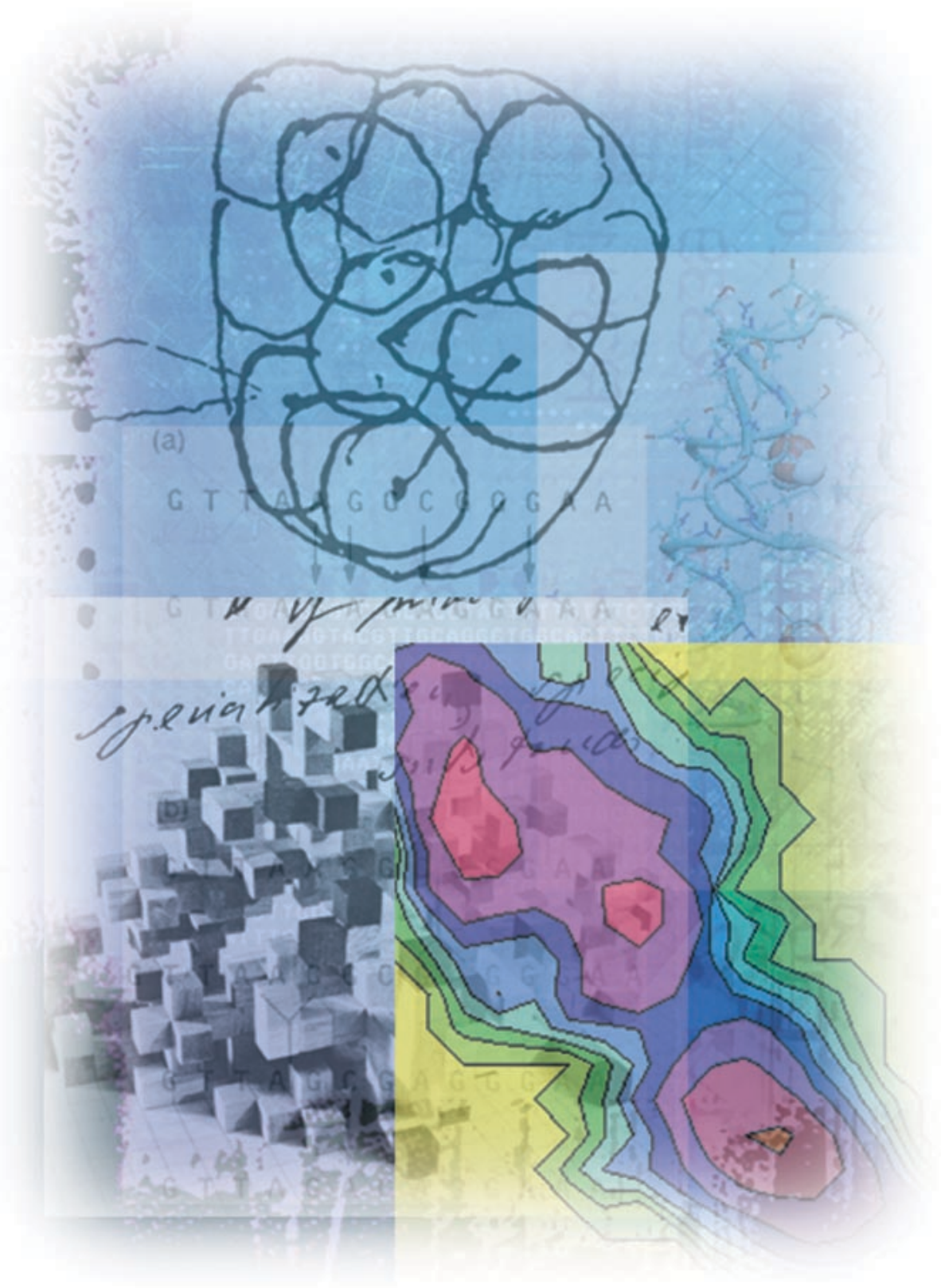
Bound-bound and bound-free transitions in high-temperature plasmas are important contributors to the radiation opacities involved in calculations on astrophysical plasmas and stellar structures and in the design of nuclear weapons. Application of my structure programs to opacity calculations is discussed in Application of my structure programs to opacity calculations is discussed in John Keady's article, "Opacity Efforts," in this "Atomic Physics" chapter.

Summary

During a 1971 sabbatical as a visiting professor with Andrew's group at Purdue, I began writing up theory and computational methods, eventually resulting in the 1981 publication by the University of California Press of *The Theory of Atomic Structure and Spectra* as the third book in the Los Alamos series in basic and applied sciences. The book has been well received (now, after 21 yr, in a fourth printing), and its appearance led directly to my 1982 receipt of an honorary Ph.D. degree from Lunds Universitet in Sweden, home of one of the world's two preeminent groups in classical spectrum analysis. The book also has been in part responsible for a number of invitations to spend visits of 3 weeks to 5 months in England, The Netherlands, Sweden, Russia, and China over the period from 1977 to 1994. Some of the Swedish visits were at Lund, others were to work with Nils Elander at the Manne Siegbahn Laboratory of Stockholm University, primarily on adaption of my codes to provide input to his close-coupling code on electron-collisional excitation.

I was very fortunate to have embarked on this field of work just at the time when large digital computers with FORTRAN compilers had come into existence, together with the availability of excellent computer facilities at Los Alamos and the presence at the Laboratory of visiting foreign scientists who brought data, ideas, and inspiration with them. The whole experience has been a classic example of basic research leading to numerous applications. During the entire period, it seemed that one person after another came to me with questions or problems just as my work had reached the point at which I had the capability of helping them.

"The whole experience has been a classic example of basic research leading to numerous applications. During the entire period, it seemed that one person after another came to me with questions or problems just as my work had reached the point at which I had the capability of helping them."


(a)

Protein-Folding Dynamics

by Angel E. García

Anfinsen
Frauenfelder
García
Leopold
Levinthal
Montal
Nymeyer
Onuchic

Proteins are the molecular machines that perform most of the functions of living organisms. Protein structure and dynamics determine the protein's functions. A protein can adopt a number of formations that grow exponentially with the number of amino acids in the chain. The problem of understanding how and why proteins act the way they do and how they bond with other organisms is a complex one. If we know how a protein folds, we can learn how other chemicals, such as drugs, bind with it. Theoretical Division researchers have developed the simulation tools to explore the interplay among various protein types, side-chain packing, and water dewetting. These tools make possible numerical investigations that could not have been performed previously and may lead to more efficient and faster-produced drugs that in turn can lead to enhanced treatment of diseases. If we know how it folds, we can learn how other chemicals such as drugs, bind with it. This research can lead to more effective and faster-produced treatments for disease.

The protein-folding problem was defined by Anfinsen in 1961, when he concluded that the tertiary structure of a protein (ribonuclease), under a given set of conditions, is determined by the protein sequence. [1]

Proteins are the molecular machines that perform most of the functions of living organisms. The functions of a protein are determined by the protein structure and dynamics. The structure of a protein is hierarchically characterized as primary (sequence), secondary (alpha helices, beta hairpins, β sheets, and others), tertiary (e.g., globular), and quaternary (the assembly of multimolecular complexes, such as hemoglobin).

Computationally, the protein-folding problem is considered complex because a protein can adopt a number of formations that grow exponentially with the number of amino acids in the chain. A typical protein has a sequence of between 60 to 150 amino acids (although there are much larger proteins). Based on this exponentially large number of possible configurations, Levinthal's paradox suggested that a protein would take the age of the universe to fold to a unique structure. [2] However, proteins do fold and do so quickly. The fastest folding protein folds in a little over a microsecond, and most proteins fold in milliseconds.

A common model accepted for many years was that the protein sequence also was encoded for a pathway that the protein chain followed. [3] Until 1974 the view about protein dynamics was dominated by the idea of sampling a single structure and assigning structures to every state detected by spectroscopy.

This view dramatically changed with the work of Hans Frauenfelder and coworkers who showed that protein dynamics were also relevant to function. Using infrared (IR) spectroscopy, he showed that the folded state of a protein consists of a large ensemble of states. Frauenfelder's work led to the concept of the protein-free-energy landscape, which gave a new framework in which to think about protein folding. [4]

The overall picture of protein folding changed in 1992 with the work of Leopold, Montal, and Onuchic. [5] They suggested that the protein folded through a funnel where a large reduction in configurational entropy was continuously exchanged by enthalpy as the protein reached a folded state. [6]

The New View

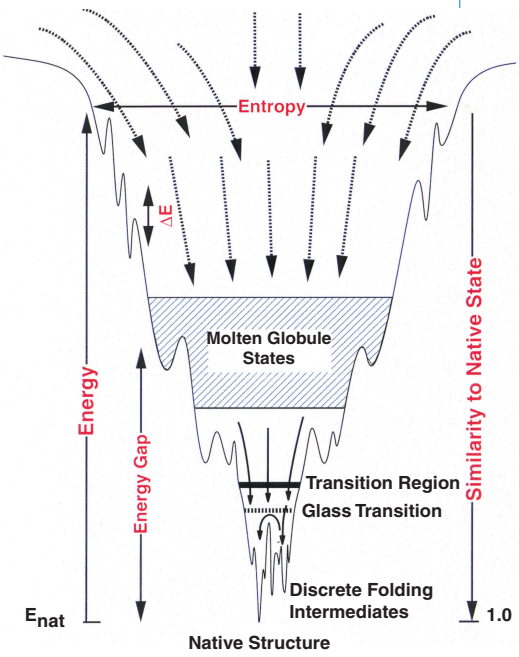


Fig. 1. A schematic representation of the energy landscape for a good folding sequence. A clearly favored native structure can be observed at the bottom of this funnel. The slope of the funnel walls imposes a bias toward the folded state. [6]

The energy-landscape theory and the funnel concept, supported by a new generation of experiments, created a "new view" for understanding protein folding. In this "new view," folding dynamics are better described as an ensemble of converging pathways toward the native structure. For good folding sequences, these paths are energetically very similar with barriers among them on k_BT scales. (See Fig. 1.) Minor fluctuations caused by environmental or mutational changes may vary these path probabilities, but the overall global landscape flow will be only weakly altered. The new view of protein folding led to one of the key results in the energy-landscape theory—that protein-folding dynamics can be properly understood as the diffusion of an ensemble of protein configurations over a low-dimensional free-energy surface, which may be constructed using many different order parameters.

A full quantitative understanding of the protein-folding problem is now becoming possible with the help of the energy-landscape theory and the protein-folding funnel concept. Good folding sequences have a landscape that resembles a rough funnel where the energy bias toward the native state is larger than its ruggedness. Such a landscape leads not only to fast folding and stable native formations but, more important, to sequences that are robust to variations in the protein environment and to sequence mutations. In a paper by Nymeyer, Onuchic, [7] and me, an off-lattice model of sequences that fold into a β-barrel native structure is used to describe a framework that can quantitatively distinguish good and bad folders. The two sequences analyzed have the same native structure; but one of them is minimally frustrated, whereas the other one exhibits a high degree of frustration.

Advances in Protein Modeling

Until recently, the modeling of proteins with full atomic detail was still limited by the inability of performing sufficient configurational sampling. This picture changed with the development algorithms that accelerated the dynamics of folding, the parallel dynamics invented by Voter of T-Division, [8] and enhancement of configurational sampling by means of multiple temperature simulations and parallel replica exchange. [9] Combining the energy-landscape view with new sampling techniques, such as replica exchange, we are now able to probe the folding mechanism at atomic resolution, including effects of explicit solvation. [10] Although it is impossible to perform sufficient kinetic folding runs at this atomic description, it is now possible, with this new approach, to perform sufficient sampling to obtain free-energy surfaces as a function of reasonable reaction coordinates.

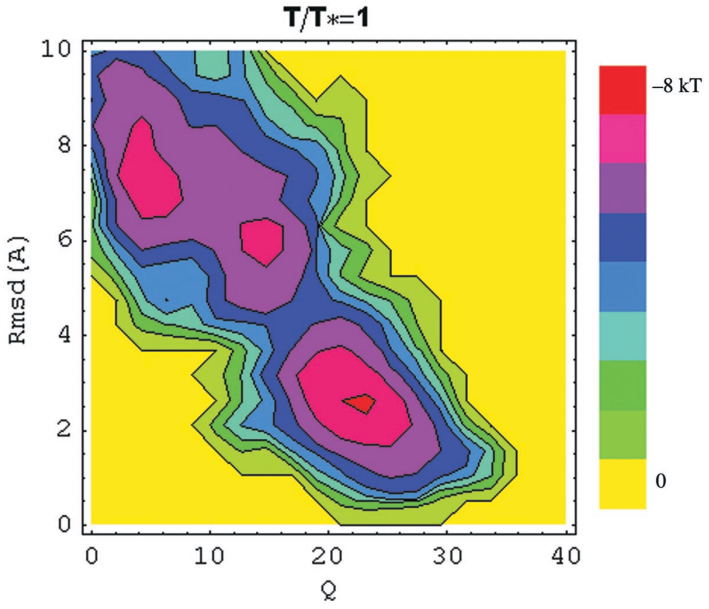


Fig. 2. Two-dimensional projection of the free-energy diagram as a function of root mean square deviation (rmsd) of the protein from the folded state and a measure of the similarity of atomic contacts to the folded state (Q) at the folding transition temperature, T*. Two free-energy basins, corresponding to the folded and unfolded states, are equally stable at this temperature. The folded basin is defined by rmsd < 4 A and Q > 10. The unfolded basin is defined by rmsd > 4.0 A and Q < 20.

Recently, Onuchic and I completed the first atomic simulation of the thermodynamics of protein folding/unfolding of a 46-amino-acid protein, protein A. The model for the system includes an all-atom representation of the 46-amino-acid protein (738 atoms) immersed in a solvent represented by over 5,000 water molecules. The replica exchange molecular dynamics method was used to simulate the thermodynamics over a temperature range of 277 K–550 K. A total of 82 copies (replicas) of the system were simulated for a 12.3 ns/replica and a total simulation time of over 1 microsecond. This simulation was performed using 82 Pentium III processors for a period of over 22 weeks (i.e., over 4,000 cpu h). The work by Onuchic and me has shown that it is now possible to explore the interplay among tertiary and secondary structure formation, side-chain packing (Fig. 2) and water dewetting (Fig. 3) during the folding event. These new simulation tools allow us to make numerical investigations that were practically impossible about 1 yr ago.

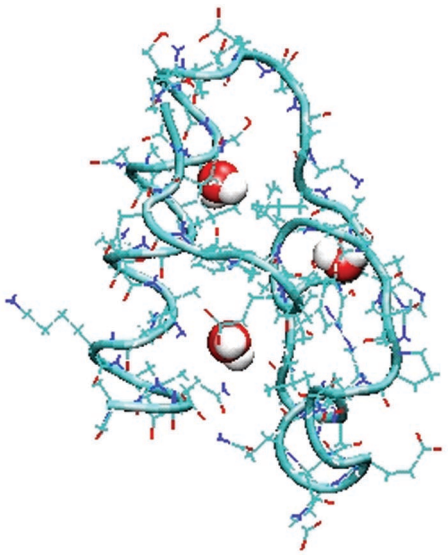


Fig. 3. Illustration of interior water molecules involved in the folding transition state of protein A.

T* = transition temperature, as distinguished from temperature (T).

The Antigenic Evolution of Influenza

Lapedes
Perelson

by Alan S. Lapedes

Influenza is a virus that causes hundreds of thousands of deaths annually worldwide. The virus continually mutates to avoid being defeated by a person's immune system. Theoretical Division scientists have developed new algorithms to analyze viral antigenic evolution. This approach, which gives a better understanding of how the influenza virus evolves, will ultimately lead to more effective vaccines and saved lives.

Influenza is a virus that causes tens of thousands of deaths in the U.S. each year and has the potential to mutate, by means of an antigenic shift, into a form that has, in the past, killed millions of people worldwide in a single year. In short, influenza is far from the popular perception of viral stomach flu or 24-h flu; instead, influenza is a specific and deadly respiratory virus that is efficient at mutating to escape being controlled by the immune system. Although antigenic shifts can be deadly, they are rare; and it is influenza's ability to mutate by means of accumulated point mutations (changes of just one or a few amino acids) that has received much public health attention. This yearly drift by means of point mutations is why many people receive influenza vaccine shots annually, in an attempt to provide protection through updated vaccines.

Our focus has been on analyzing the antigenic evolution of influenza by means of drift. Genetic evolution is a more familiar topic that refers to characterizing evolution by noting the changes in the genetic sequence of the virus. Many tools, such as reconstruction of phylogenetic trees, are available to quantitatively analyze the genetic evolution of viruses. Antigenic evolution refers to the virus's ability to escape detection by the immune system by making changes so that the antibodies of the immune system are not able to bind strongly to the mutated virus. Although much data are available on antigenic evolution (because national influenza centers worldwide collect antibody-binding data in addition to genetic sequence data), few quantitative tools exist to analyze this important class of data.

Building on a conceptual framework called "shape space," proposed over 20 yr ago by our Los Alamos colleague, Alan Perelson, to think about antibody-antigen binding, we have shown that mathematical algorithms called multidimensional scaling algorithms (first developed in the mathematical psychology literature) can be extended to construct the detailed geometry of shape space, given antigen-antibody binding data. [1] Using our approach, and in collaboration with the Dutch National Influenza Center, we have shown [2,3] that the influenza virus over the past 30 yr can be described as evolving in a

five-dimensional space and that one can even project this high-dimensional evolution down to two dimensions so that it is easily visualized and analyzed by humans accustomed to living in just three dimensions. This approach to analyzing viral antigenic evolution complements sequence-based approaches to analyzing genetic evolution and can provide new ways to analyze how viruses evolve—an important goal in this era of emerging and reemerging deadly viruses.

GenBank—From Idea to Worldwide Indispensability

by Christian Burks and Linda K. Wood

GenBank has more than 22 billion bases and 18 million sequences. This DNA sequence database is now used worldwide and is an essential tool for the following:

(1) scientists doing basic research into the molecular basis for life;

(2) biotechnology and pharmaceutical companies generating new medicines (and the associated intellectual property claims);

(3) agricultural companies creating the next generation of plants, pesticides, and nutriceuticals;

(4) educators and government agencies overseeing research grants; and

(5) workers in forensic and environmental sciences and civil law.

Daily, approximately 40,000 users search or download GenBank's base pairs. A Theoretical Division team of elite theoretical physicists and molecular biologists created this fundamental aid to biological research. Scientists hope that sequencing will lead to new and effective treatments for diseases, including cancer and Alzheimer's.

DNA SEQUENCES

Deoxyribonucleic Acid (DNA) is the genetic blueprint for life, coding for the proteins that make up the machinery, regulators, and structure of a cell. DNA is composed of two polymeric strands that are in turn composed of a chemical alphabet of four chemical monomers, or "bases." These "**bases**" are represented by A, C, G, and T; and the genetic information is coded by the specific ordering of these letters. The strands are bound to one another in an antiparallel, zipper-like fashion, with the binding mediated by the fact that A always binds to T, and C to G. These pairings are referred to as "**base pairs**." The strings of these letters are referred to as "**sequences**." J. D. Watson and F. H. Crick won the Nobel prize in 1962 for determining the underlying structure of DNA, and Walter Gilbert and Frederick Sanger won a Nobel prize in 1980 for developing biochemical methods that allow scientists to determine the base-ordering in DNA sequences. Aided by these discoveries, scientists began to determine and publish DNA sequences in the 1970s, and they continue to do so at an exponential rate until the present day.

Bell
Biolofsky
Burks
Cameron
Carrico
Crick
Doolittle
Fickett
Gilbert
Goad
Grantham
Hamm
Jordan
Kabat
Kanehisa
Kirschstein
Lewitter
Lipman
Marr
Murray
Rindone
Sanger
Simon
T. Smith
Watson
L. Wood

Monday, May 12, at 10:30 Steve Simon invites you for cake and coffee to celebrate 100,000 bases now in the DNA sequence library.

—*Los Alamos National Laboratory memorandum,* May 9, 1980 [1]

As of September 2002, GenBank had more than 22 billion deoxyribonucleic acid (DNA) bases and 18 million sequences. What began as an attempt to store information unsuitable for keeping in notebooks is now an essential tool of molecular biologists all over the world. Daily, approximately 40,000 users search or download GenBank's base pairs. This database, which has been available free since the early 1980s on magnetic tape, floppy disks, online, or in hard-copy form, was the brainchild of Theoretical Division's Walter B. Goad and a team of elite theoretical physicists and molecular biologists. [2]

The explosion of DNA sequencing engendered the birth of a crucial new field of "bioinformatics," a combination of biology and information technology. Because of advances in deciphering the genetic code, scientists are producing enormous amounts of biological data for storing and analyzing by computer. [3] A bioinformatics scientist focuses on the development and application of software tools and database resources that enable the efficient extraction of information from biological data sets that are far too large to be usefully examined by the eye. Expertise in bioinformatics bridges biology, chemistry, computer science, and software engineering. The tools and resources these scientists create are tremendously enabling for scientists who need efficient, informative ways to sift through all the known DNA sequences to find facts or patterns in their areas of interest.

GenBank is the public repository that stores DNA sequences and the descriptions of the functionality of the genes they code. This repository is, among other things, the final destination of the sequencing for the human genome.

Sequences often provide the fundamental framework whereby scientists focus their insights into the origins and causes of diseases, including cancer and Alzheimer's. [3] In fact, this DNA sequence database is now used worldwide and is an essential tool for the following:

(1) scientists engaged in basic research into the molecular basis for life;

(2) biotechnology and pharmaceutical companies generating new medicines (and the intellectual property claim associated with them);

(3) agricultural companies creating the next generation of plants, pesticides, and nutriceuticals;

(4) educators and government agencies overseeing research grants; and

(5) practitioners of forensic science, environmental science, and civil law.

GenBank is an annotated collection of all publicly available DNA sequences. Most scientific journals require that sequence information be submitted to GenBank before they will publish an article. In fact, GenBank staff played a crucial role in recognizing the need for a major paradigm shift in the way that biologists publish data, such as DNA sequences, [4] and worked closely with the editors of a number of key journals to engineer the direct deposition of sequence data with GenBank preceding publication of an article describing the importance of a new sequence. Because the database is designed to provide and encourage access within the scientific community to the most up-to-date and comprehensive DNA sequence information, no restriction is placed on the use or distribution of GenBank data. [5]

In 1998, The University of Texas BioTech Science Resources cited GenBank as the "best bet" for most sequence searches. "It offers daily exchange of information with other major sequence databases, has a variety of user interfaces, fairly detailed online help (with e-mail addresses for more information if what is already available is not sufficient), and a speedy interface." BioTech also found that, at that time, GenBank was so popular that its output was somewhat slow in forthcoming. They suggested using the database after work hours in order to get fast results. [6]

Beginnings

"There's a great story in how this all got started," says biologist Christian Burks, who joined Goad in Los Alamos in February of 1982 and who is now chief science officer at Affinum Pharmaceuticals in Toronto. "George Bell and Walter Goad really championed the idea of doing a database. They were both nuclear physicists who had gone to Los Alamos as part of the weapons effort, but during the 1970s they got interested in molecular biology. The fact that DNA sequencing was taking off was what interested them, and they had the foresight to understand how crucial a role computational approaches would play in going forward."

Burks continues, "They had spent a lot of years collecting very large data sets—nuclear cross sections crucial to calculations around the weapons program—behind closed doors…thinking about what to do with data that you can't store in notebooks." [1]

In 1970–71 Goad spent a year with Nobel Laureate Francis Crick at the Medical Research Council Laboratory of Molecular Biology, Cambridge, United Kingdom. When he returned to Los Alamos, he went into biology full time and provided theoretical support to several experimental biology programs at the Laboratory. In 1974, he joined Bell's T-Division group. Bell had also spent a swords-into-plowshares sabbatical with Nobel Laureate Walter Gilbert at Harvard University, one of the labs where DNA sequencing was first created as a methodology.

In 1979, two other Los Alamos scientists, Mike Waterman of the Analysis and Assessment Division and Temple Smith of T-Division, attended a meeting at Rockefeller University to discuss the management of sequencing data. With their encouragement, Goad became active in discovering how to analyze sequence data with computers. He began collecting sequences on a computer and writing software for the analysis of his pilot sequence database. He headed the Los Alamos effort to create a national data bank and analysis center for nucleic acid sequences. [7]

In addition to the catalytic role GenBank played in establishing electronic data publishing as a correlate to publishing affiliated findings in journal articles (discussed previously), T-Division staff working on the project played a very central role in formulating the national discussions on the feasibility of the Human Genome Project. It was clear that the output of the Human Genome Project, the complete DNA sequence for the human genome, and the complete sequences for a number of other, related organisms, would end up in GenBank and that the processes, mechanisms, and protocols by which these data got there would be largely informed by the experience with GenBank. Theoretical, Computing, and Life Sciences Divisions, along with others at Los Alamos, went on to provide the core of the Center for Human Genome Studies (CHGS) that was established by the U.S. Department of Energy at Los Alamos as part of the national Human Genome Project. George Bell served as initial head of the CHGS.

"It's a wonderful success story that the database was started in an unusual place and with an unusual cast of characters—in terms of mainstream molecular biology at the time—and now is simply essential and ubiquitous, something biologists use every day," Burks comments.

Goad began constructing the Los Alamos Sequence Library in 1979 with help from Minoru Kanehisa, Jim Fickett, and several graduate research associates. Burks joined this team in early 1982, at the point they were writing the proposal that obtained a $2 million, 5-yr contract from the National Institute of General Medical Sciences (NIGMS) to establish GenBank. The grant was based on their earlier work on the Los Alamos Sequence Library.

Other Contributors

Scientists found out that the database existed and began requesting data that would aid in their research. The need for a widely available database became apparent. When the U.S. National Institutes of Health competed the contract to develop and maintain such a database, Goad's team got the nod. The team would collaborate with Bolt, Beranek and Newman (BBN) in Boston, who would take responsibility for distributing the database to users.

Another person responsible for getting GenBank off the ground was Elke Jordan, then program activity director at NIGMS. Jordan and then NIGMS director Ruth Kirschstein brought eminent molecular biologists together to discuss particulars concerning such a database. Christine Carrico, now executive officer of the American Society for Pharmacology and Experimental Therapeutics, was the NIGMS project officer for the first 5-yr contract. Carrico says Goad won the contract because the Los Alamos team "wrote a more responsive proposal and had more facilities at its disposal." [1]

European scientists were also collecting sequence data. Such researchers were Richard Grantham in France and Greg Hamm, Graham Cameron, and Ken Murray at the European Molecular Biological Laboratory (EMBL) in Heidelberg, Germany. [1]

WHERE ARE THEY NOW?

T-Division researchers were among a handful of groups worldwide that spawned the first generation of bioinformatics scientists. Some have now taken their expertise to other organizations and countries.

- Walter B. Goad died November 2, 2000.
- George Bell died May 28, 2000.
- Christian Burks has been responsible for leading the integration of informatics and bench science at two biotechnology start-up companies, Exelixis Pharmaceuticals, Inc., South San Francisco, and Affinium Pharmaceuticals in Toronto, Ontario, Canada.
- Minoru Kanehisa is director of Kyoto University's Bioinformatics Center Institute for Chemical Research.
- Tom Marr, who joined the project several years later, went on to found a bioinformatics start-up company, Genomica, in Boulder, Colorado, and is now President's Research Professor in Bioinformatics at the University of Alaska, Fairbanks.
- Jim Fickett is global director for bioinformatics at AstraZeneca, a pharmaceutical company in Boston.
- Temple Smith is now director of the Biomedical Engineering Research Center at Boston University.

In 1981, Goad offered a 1-yr visiting staff appointment at Los Alamos to outstanding bioinformatics pioneer, Temple Smith. Smith assisted users of the Los Alamos database, computers, and software in defining and solving research problems. [1]

Gerry Myers, on the faculty at St. John's College in Santa Fe and a molecular biologist by training, also spent a sabbatical year in T-Division working on GenBank. He later joined T-Division and created a spin-off database that contains human immunodeficiency virus (HIV) sequences. This effort also was eventually funded by the U.S. National Institutes of Health (NIH). Meyer's database rapidly established itself as the premiere global resource for genetic information for HIV, both as a database and as a focus for analytical approaches to determining the evolutionary patterns of HIV. This effort blossomed into several viral databases that focused on key disease threats to society and continues to this day in T-Division under the leadership of Bette Korber and her colleagues.

So, T-Division and BBN partnered on the 5-yr contract. Bilofsky tells how it worked. "Los Alamos would enter the data and periodically we would get a distribution from them that we would massage and [apply] tools for doing quality control." In other words, Goad's group collected,

managed, organized, and oversaw the data and then sent it by file transport protocol to BBN for publishing. [1]

Overwhelming Response

Early response to the database was overwhelming. Goad's five-person team had a difficult time keeping up. Burks describes the process.

> We would go to the literature [and get] the sequence [from] the paper—a line of As, Ts, Cs, and Gs. We would tear it out or photocopy it and someone would sit down and type it into the database. As soon as there was this public database, people expected the data to be there, but it would take a year to show up. [1]

The recognition that these mechanisms would not scale with the growth of the published data drove the GenBank team to shift to reliance on direct submission, in electronic form, of sequence data from the labs that were creating the data sets.

Fran Lewitter, now head of the Biocomputing Group at the Whitehead Institute, worked on GenBank at BBN in those days. She says, "It's amazing how things have changed in the last 20 yr in terms of accessibility and the kinds of information we collect." Lewitter, who teaches a short history of GenBank, says students are amazed that GenBank was on printed matter early on. They have a hard time imagining how the tool worked before the Internet. [1]

By 1984, 2.8 million databases had been entered into GenBank, and 120 universities and individuals received the database on magnetic tape with approximately five users a day obtaining the information on line. Carrico credits two technological innovations with speeding up the process: the use of microcomputers at Los Alamos and the use of the Advanced Research Projects Agency Network (ARPAnet) to transfer data between Los Alamos and BBN. ARPAnet belonged to the U.S. Department of Defense. [1]

The use of microcomputers was an ironical innovation, given that much of the perceived strength of T-Division's proposal was based on the preeminent concentration of supercomputer facilities at Los Alamos (then CDCs, later Crays). It turned out that the text-editing and management components of the project were much better managed with small, locally controlled microcomputers and the editing and management of text files that these microcomputers supported well. T-Division's need for and use of these computers as a complement to the supercomputers played a significant role in Los Alamos expansion into a larger and next-generation picture of scientific computing, for example, with a focus on scientific workstations. GenBank's first microcomputer, a Gifford CompuPro, was assembled by Goad on a rack that sat in his office. His colleagues in T-Division were often met, when entering his office, with the sight of him poring over an octal representation of machine code as he tried to debug the computer and its controlling software.

GenBank played a central role in another high-impact development in computing, the extension of relational database management systems (RDBMS), software for managing and querying data sets into large, free-text applications. Goad and Kanehisa had explored the use of early RDBMSs for sequence data but found them awkward and inconvenient, leading to the text-file underlying format for GenBank. (This format is still in use today in distributed versions of the database.)

However, by the late 1980s, several later-generation database companies were recreating the tools to accommodate applications (such as GenBank with its long sequence fields) that had not been envisioned earlier. T-Division ended up being the first group at Los Alamos to license the newer-generation tools from a company called Sybase in California. This transaction created quite a stir at Sybase, given that Los Alamos was at the time the largest holder of licenses from their primary competitor, Ingres. The partnership catalyzed a substantial role for Sybase in biological database projects in the following decade, where the Los Alamos product stood for some time as the tool of choice.

International Collaborations

From the beginning, Goad worked on data-sharing with Europe. By 1982, Goad's team and BBN's Hamm and Wayne Rindone (now a senior data programmer/analyst in the Harvard Medical School Department of Genetics) had devised a way to transfer data between Los Alamos and EMBL by means of Telenet access to a Virtual Address Extension (VAX). Goad said, "This will permit very close coordination of our collection efforts, even if the data itself is exchanged on magnetic tape." GenBank became a global enterprise. [1]

It took a while to get the data interchange running smoothly; but, by 1984, EMBL and GenBank were working together to publish a hard-cover compendium supplement to *Nucleic Acids Research*, a journal published by Oxford University Press. This journal and all other major journals began sending Goad sequences in papers that they accepted for publication. Lewitter says, "That was a big thing. If you published a sequence, you had to get an accession number to GenBank." [1]

The next development was university access to GenBank through the ARPAnet. Progress came slowly, but as the Internet developed, things moved more quickly. Difficulties arose because GenBank didn't incorporate instructions on how to use it, but with time these problems too were ironed out.

International collaboration and cooperation were a key component of NIH's mandate in funding GenBank, and T-Division scientists worked closely with their colleagues at the sister database efforts that were established in Europe and Japan: the EMBL Data Library (then in Heidelberg at the European Molecular Biology Laboratory, now outside Cambridge, United Kingdom, at the European Bioinformatics Institute); and the DNA Data Bank of Japan (at the National Institute of Genetics in Mishima, Japan).

In 1987, NIGMS conducted a competition for a second 5-yr contract. Goad had turned over the reins of the project to Fickett and Burks by this point. They and their colleagues put together the winning proposal, under Burks' leadership, that included a five-fold increase in funding and a new commercial partner, IntelliGenetics, to take the place of BBN.

NIH had initiated GenBank as a contract because they didn't have the internal resources or framework to take on such a project. However, as the central role of database resources and analysis tools developed around GenBank and the Human Genome Project, they saw a significant advantage to developing an internal intitiative, which resulted in their establishing the National Center for Biotechnology Information (NCBI) as part of the National Library of Medicine in the early 1990s. Thus, rather than continuing to manage GenBank under the contract mechanism, in 1992 at the end of the second 5-yr contract with Los Alamos, they took over maintenance of GenBank, along with several other databases. They have continued the international collaborations established by the original GenBank effort.

At the same time, T-Division spun off a number of its staff into the National Center for Genome Resources in Santa Fe, which in turn has spun off at least one bioinformatics start-up company and established several database initiatives in its own right.

Indispensable Tool

Russell Doolittle, an evolutionary biologist at the University of California, San Francisco, who served on the GenBank advisory committee, calls the project an enormous success that now rivals the U.S. Geological Survey in terms of scientific usefulness. [1]

David Lipman, NCBI director, has plans for expanding GenBank's use for research in protein families, model organisms, and gene functions. Of the early days, he says, "I'm impressed by the view that they had of the future." [1]

ComputerWorld Smithsonian awards went to GenBank in 1992 and 1998. These awards honor innovative use of information technology and achievement of outstanding progress for society through use of computers. A news release about the award to GenBank says, "GenBank speeds medical research and helps provide insights into the causes of such diseases as AIDS and cancer by allowing scientists worldwide to electronically exchange and compare newly determined DNA sequences to others in the database." [8] Francis Ouellette, GenBank coordinator at NCBI in 1998, said, "Every major advancement in medical science today owes a debt to GenBank." [9]

And Lewitter says, "People can't understand what it would be like if they couldn't sit down at their computer and search GenBank." [1]

And it all began with Walter Goad in T-Division.

HIV Evolution and Its Implications for Understanding Viral Origins

Bhattacharya
Korber
Kuiken
Muldoon
Myers
Theiler

by Bette T. Korber and Tanmoy Bhattacharya

Theoretical Division's Human Immunodeficiency Virus (HIV) Database, begun in 1986, contains 80,000 HIV sequences that are available to the international community on the Worldwide Web. The database also contains thousands of HIV "epitopes," small portions of viral proteins that can trigger an immune response from the human body. This database serves thousands of researchers each week and provides a foundation for research projects at Los Alamos and elsewhere. These projects are directed toward producing effective vaccines to combat HIV.

The human immunodeficiency virus (HIV) Database and analysis group (part of the Theoretical Biology and Biophysics Group in Theoretical Division at Los Alamos) was founded in 1986 by Gerald Myers, when he realized that HIV was an incredibly diverse virus and that this diversity would create particular problems for vaccine design, drug development, and HIV research in general. The original purpose was to collect and curate the rapidly growing set of sequences that were generated by Acquired Immune Deficiency Syndrome (AIDS) researchers worldwide and to make it available to them in the form of a yearly publication, *The HIV Sequence Compendium*. In 1995, Bette Korber created a companion volume—*The HIV Immunology Compendium*—which provided a comprehensive, annotated overview of immunologically relevant regions in HIV proteins. In 1996, Gerald Myers moved on to found several other pathogen genetic databases and left the HIV database project to Korber, who was soon joined by Carla Kuiken.

As the World Wide Web became a fundamental resource for scientific exchange in the mid-1990s, the work at the database broadened appreciably. We now maintain a database of approximately 80,000 HIV sequences that are accessible to the world by means of a web-search interface and a collection of thousands of immunologically recognized HIV "epitopes." Our web-accessible databases have become a resource that thousands of researchers use each week. These databases also serve as a foundation for our own research projects, which are aimed at understanding the global evolution of HIV, HIV molecular epidemiology, and the interplay between the human immune response and viral evolution at a population level. All of this work is primarily driven by the hope of developing educated intervention strategies and well-designed vaccines.

The Origins of the Virus

In the late 1990s, Korber first became involved with studies to help elucidate the origin of AIDS. This work grew out of her initial analysis of viral sequences from the oldest known HIV-positive sample obtained from a blood sample taken in 1959. [1] The virus from this sample DNA was found to be "ancient" relative to contemporary samples, but it seemed premature to draw conclusions about the time of the ancestor of the pandemic strains of the virus because of limitations of the models and the methods. [2] In particular, the methods available then did not allow for statistical testing of the reliability of the results, nor did they provide a quantitative estimate of the effects of the random nature of evolution on the estimate. It was known that both of these deficiencies could be mitigated by the use of the statistically rigorous and computationally intensive methods following from the maximal likelihood principle. This approach could reveal shared lineages and mutational patterns and could be used to reconstruct elements of the evolutionary history of viruses. Also, this approach is better able to deal with biases in the data. With enough samples, the approach was known to produce more reliably both the branching patterns, i.e., genealogical relations between viruses and branch lengths that indicate the amount of mutations that arose in the period that they evolved separately. But because of the extreme computational demands, it was not feasible to study more than a few dozen sequences, whereas reliable estimates of the deep structure in the tree (i.e., estimates of ancient evolutionary history) needed several hundred large ones.

Using such a large number of sequences was possible only with the use of the modern supercomputers, but no maximum likelihood code had ever been ported to such platforms. Tanmoy Bhattacharya, a theoretical physicist in T-Division, became interested in the problem. He parallelized the maximum-likelihood tree-building code and improved the evolutionary models to more realistically model HIV evolution under the assumption of a molecular clock, which holds that genetic mutations accumulate as a linear function of time. James Theiler, a physicist in the Los Alamos Nonproliferation and International Security (NIS) Division, contributed statistical expertise to estimate the confidence limits, using Monte Carlo resampling strategies. Mark Muldoon, a mathematician from England who was visiting T-Division, contributed an alternative strategy that incorporated sampling times into the generation of the phylogenetic tree. Korber, as a biologist, contributed her knowledge of phylogenetic methods and HIV evolution and then applied the methods developed by the team. Thus the theoretical foundation of this project was a multidisciplinary effort. In the culmination of this work, global HIV-1 sequences with known time of sampling were used as a basis to estimate the timing of the ancestral sequence of the main (M) group of HIV-1—the viruses that are responsible for the AIDS pandemic. We estimated that the most recent common ancestor of the HIV-1 M group occurred in 1931 (95% confidence interval [CI] , 1914–1941) (Fig. 1). [3,4]

Additional Studies

We used coalescent theory and evolutionay analysis to do further studies and to model the history of the epidemic in the Democratic Republic of the Congo (DRC). These methods can be used to inform models of the early spread of the virus; and, for a variety of reasons, the DRC is thought to be in or near the region where the virus began its expansion in the human population. The coalescent analysis of the DRC data was most compatible with a model of epidemic growth involving increasing exponential growth rate of the infected population through time. This compatibility suggests that the early years of the HIV epidemic in this region of Africa were characterized by an extended period of slow spread. This model helps to explain how HIV-1 may have remained present but undetected for the period between the timing of the ancestral sequence of the M group (now determined to have been circulating in 1930) and the time that AIDS was finally recognized as a disease (1981). [5]

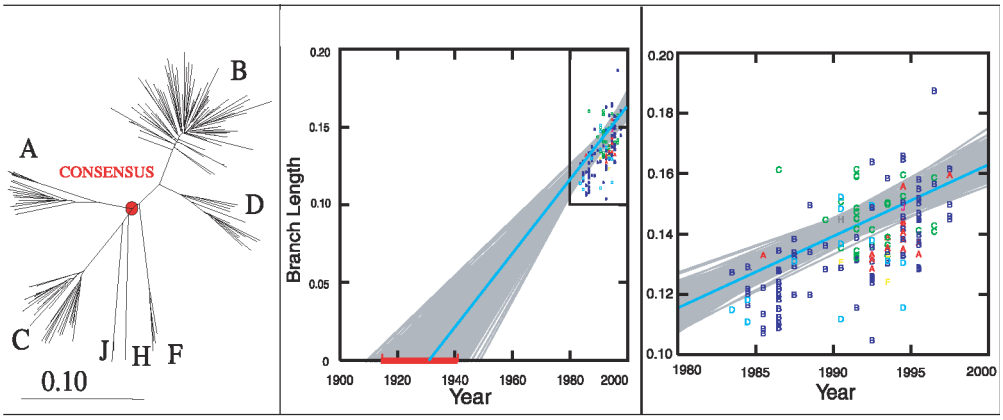


Fig. 1. Estimating the last common ancestor of the HIV-1 M group based on data collected over the two decades 1981 to 2000. (a) The gp160 phylogenetic tree used for this calculation. (b) The sum of the branch lengths from each leaf to the root of the tree is plotted against time. The subtype of the HIV sequence is indicated by colored letters. A maximum-likelihood linear fit to the data was generated as described in the text (bold blue line in figure). These lines prove to be steeper and are offset to the left (into the past) relative to lines that would be drawn using a "standard" linear least-squares fit to a line. We used 480 bootstrap fits to data points to calculate 95% CI, shown in the figure as a red line along the horizontal axis. (c) A magnified view of the boxed region in (b), showing that the points derived from different subtypes tend to be reasonably well distributed about the line, a consequence of the approximate equality of the intra clade evolutionary rates.

Finally, these efforts lead us to a practical concept for HIV vaccine design. Maximum likelihood trees enable us to model ancestral sequences within the framework of a phylogenetic tree. Such ancestral sequences in HIV phylogenetic trees are roughly twice as close to contemporary circulating strains as such strains are to each other. Our concept was to use an artificial ancestral sequence for creating a vaccine strain that might be more cross-reactive. Such approaches are currently moving from a concept stage [6] to an initial experimental testing phase being led by our colleagues at Duke University, the University of Alabama at Birmingham, Harvard University, the U.S. National Institutes of Health (NIH) and Cornell University. [7] Although much work is yet to be done to determine if such ancestral strains ultimately may be useful in a vaccine, these strains show initial promise in terms of antibody cross-reactivity [7] and so will be further tested and incorporated into vaccine design concepts.

Modeling HIV Infection and Introducting Combination Therapy

Ho
Perelson

by Alan S. Perelson

During the last 7 yr, revolutionary advances have been made in our understanding of HIV-1 that have been translated into new therapies. Much of this increased understanding and the rationale for using combination therapy, i.e., a "drug cocktail," has arisen because of mathematical modeling done in the Theoretical Division.

HIV on average takes about 10 yr to advance from initial infection to full-blown AIDS. If a patient with chronic HIV infection has multiple blood tests and the amount of virus in the blood, i.e., the viral load, is measured, one generally finds that the viral load remains nearly unchanged for periods of months and years. This suggested to many that HIV infection was very slow, because the amount of measurable virus was not increasing, and that treatment might not be needed. Modeling, however, revealed that this was not the case.

In a collaboration with David Ho, head of the Aaron Diamond AIDS Research Center, The Rockefeller University, we examined viral-load data obtained during clinical trials of one of the first HIV-1 protease inhibitors, ritonavir. [1] This trial was placebo controlled, so some patients did not receive the drug. In these patients the viral load did not change over the 1-month period of the trial. In the group that was given the drug, the viral load was observed to drop by one-to-two orders of magnitude during the first 2 weeks of therapy. Mathematical analysis of these data showed that the virus was being cleared from infected patients under treatment at a rapid rate.

One way to quantify the clearance rate is to compute the half-life of the virus in blood, i.e., the time it takes for the body to clear half of the virus in the blood. This half-life was estimated as being 2 days or less. [1] Given this half-life, it was easy to show that to maintain the constant (or steady-state) level of virus in the blood before therapy required that HIV be produced at a rapid rate. In fact, the amount of the virus in blood needed to double every 2 days to keep up with the clearance. Calculations showed that this amounted to at least a half-billion new viruses being produced each day. [1] If fewer viruses were produced, then the body would clear the virus, and one would not have observed a constant viral load in untreated patients.

A Radical Change in Viewpoint

These observations caused a radical change in viewpoint. HIV was now seen as a rapidly reproducing virus and one that could respond to therapy, as opposed to an indolent infection that was barely spreading. If virus replication and clearance were finely balanced and occurring at rapid rates, then tipping the balance in favor of clearance might have a large impact on viral load. This meant that drug therapy might succeed in wiping out this virus.

However, mathematics also pointed out a problem. At the predicted rate of rapid HIV production, one could calculate that large numbers of mutations (or changes) in HIV's genes would occur. Because the currently available drugs all worked by inhibiting an HIV enzyme, this rapid mutation process could cause HIV to evolve so as to be resistant to the drug. To bring home the point, it was known that for some drugs a change of a single amino acid in the HIV enzyme would cause the drug to lose efficacy. Our calculations showed that every single possible point mutation of the HIV genome would be made hundreds or thousands of times each day. [2] Thus, modeling suggested that the virus would quickly become resistant to any single drug, particularly those that required only one mutational change in the virus to generate resistance.

More refined models, which included the fact that HIV-1 was produced from infected CD4 + T cells, showed that the vast majority of infected cells lived on the order of one day while producing virus. This was an important finding, as HIV infection causes a depletion of CD4 + T cells. These cells play a central role in the immune system and their loss causes the immune system to malfunction. When the CD4 + T cell count, which in a healthy individual is about 1,000 cells/microliter of blood, falls below 200, the person is said to have full-blown AIDS. Why the CD4 + T cell count falls is unknown; but our work showed that cells, which when uninfected have lifespans of weeks or months, upon infection only live about a day. [3,4] Thus, blocking T-cell infection with potent antiviral drugs would be expected to have a profound effect on the course of AIDS. This in fact has been found to be true.

A similar theoretical analysis has been applied to other chronic viral infections, namely hepatitis C virus (HCV) and hepatitis B virus (HBV) infection. [5,6] A few percent of the world's population are infected with these viruses.

In the U.S. it has been estimated that 3–4 million people are infected with HCV. This infection causes liver damage, may lead to the development of liver cancer, and is the leading cause of liver transplantation in the U.S. The standard therapy for this infection involves getting injections of a compound called interferon three times a week plus taking an antiviral medication, ribavarin, once a day. Interferon monotherapy is very ineffective, clearing virus in only about 10% of treated patients.

Interferon-ribavarin therapy is more effective at clearing the virus in perhaps 30% of treated patients. A current challenge is to develop more potent therapies or to develop a vaccine. Studying the changes in the amount of virus assayed in the blood of patients treated with interferon showed that, like HIV, hepatitis C virus was being produced rapidly, about 10^{12} virions/day, cleared rapidly (half-life of 3 h) and was mutating rapidly. [5] Thus, the same obstacles as we discovered for treating HIV are present in HCV infection.

Further, the modeling of the kinetics of viral decay during therapy allowed us to estimate for the first time the efficacy of different dosing regimes of interferon. In particular, we showed that the approved schedule of giving interferon three times a week at the U.S. Food and Drug Administration (FDA)-approved dose was not very effective in blocking HCV production. We estimated that only 60% of virus production was blocked with the standard drug dose, whereas with higher experimental doses, as much as 95% of the production could be blocked.

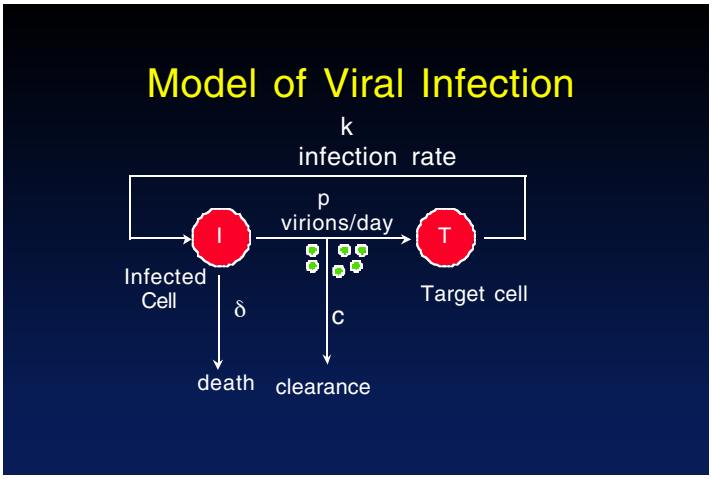


Fig. 1. Basic model of viral infection. Uninfected (target) cells, *T*, are infected by virus, *V*, with rate constant *k*. Infected cells, *I*, produce virus at rate *p* per cell and die at rate δ per cell. Free virus is cleared at rate *c* per virus particle. Not shown in the figure is the production of new target cells at rate λ and their death at rate *d* per cell.

Basic Model of Virus Infection

The model that has been used to study HIV, HCV, and HBV infection is shown in Fig. 1. The model considers a set of cells susceptible to infection, i.e., target cells, T, which through interactions with virus, V, become infected. Infected cells, I, are each assumed to produce new virus particles at a constant average rate p and to die at rate δ per cell. The average lifespan of a productively infected cell is $1/\delta$ and thus if an infected cell produces a total of N virions during its lifetime, the average rate of virus production is $p = N\delta$. In the case of HIV infection, death may be either by viral cytopathic effects or immune-mediated cellular destruction. Newly produced virus particles, V, can either infect new cells or be cleared from the body at rate c per virion. Box 1 on the next page shows the equations corresponding to the model in Fig. 1 and illustrates their use in analyzing data obtained from HIV-infected individuals undergoing antiretroviral therapy to obtain minimal estimates of the parameters c and δ. From these estimates the half-life of virions in plasma and the half-life of productively infected cells can be computed.

Box 1

The equations that describe the basic model of viral dynamics shown in Fig. 1 are

$$dT/dt = \lambda - dT - kVT\,, \quad (1)$$

$$dI/dt = kVT - \delta I\,, \text{ and} \quad (2)$$

$$dV/dt = pI - cV\,. \quad (3)$$

If one assumes that initially a person is uninfected, i.e., $V = 0$, $I = 0$, and $T = \lambda/d$, and then introduces a small amount of virus, the solution of Eqs. (1)–(3) mimic the kinetics of primary HIV infection.

To analyze the effects of giving an antiviral drug, Eqs. (1)–(3) are modified.

Reverse transcriptase (RT) inhibitors block the ability of HIV to successfully infect a cell. Protease inhibitors (PIs) cause the production of noninfectious viral particles, V_{NI}. Thus, in the presence of these drugs, the model equations become

$$dT/dt = \lambda - dT - (1 - \varepsilon_{RT})kV_IT\,, \quad (4)$$

$$dI/dt = (1 - \varepsilon_{RT})kV_IT - \delta I\,, \quad (5)$$

$$dV_I/dt = (1 - \varepsilon_{PI})pI - cV_I\,, \text{ and} \quad (6)$$

$$dV_{NI}/dt = \varepsilon_{PI}\,pI - cV_{NI}\,, \quad (7)$$

where ε_{RT} and ε_{PI} are the efficacies of the RT and protease inhibitors ($\varepsilon = 1$ being a perfect drug), V_I and V_{NI} are the concentration of "infectious" and "noninfectious" viruses, respectively, and $V = V_I + V_{NI}$ is the total amount of virus.

If the concentration of target cells is assumed to stay constant at the measured baseline level, T_0, then Eqs. (5)–(7) can be solved in closed form. If T changes, then numerical methods can be used to solve the equations. If a 100% effective PI is given to an individual at a steady state with a viral load, V_0, and one assumes that over some period of time that $T = T_0$, the viral-load decay will obey the equation

$$V(t) = V_0\,(-ct) + \frac{cV_0}{c-\delta}\left[\frac{c}{c-\delta}\left\{\exp(-\delta t) - \exp(-ct1)\right\} - \delta t\exp(-ct)\right]. \quad (8)$$

Using nonlinear least-squares regression, this equation was fit to patient data and used to estimate the parameters c and d. Two examples are shown in Fig. 2. The half-life of free virions and productively infected cells can then be computed by the formulas $ln\ 2/c$ and $ln\ 2/d$, respectively.

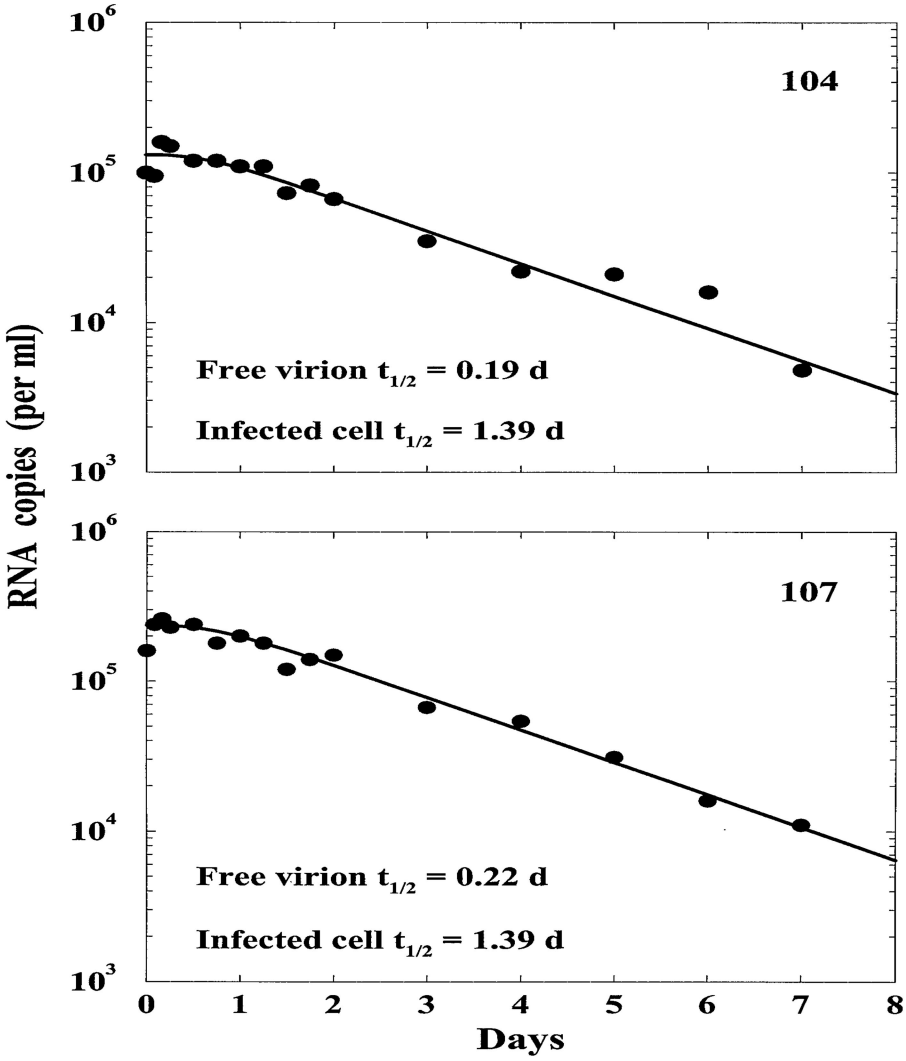


Fig. 2. Data obtained from two HIV-infected patients who initiated antiretroviral therapy at time 0. The solid line is the best-fit theoretical curve from which the parameters c and δ were estimated.

Universal Scaling Laws in Biology: From Molecules and Cells to Whales and Ecosystems

by Geoffrey B. West

Brown
Enquist
Hemmingsen
Kleiber
West
Woodruff

Life is the most complex physical system in the universe: its complexity is matched only by its diversity. Theoretical Division researchers have developed a theory that offers a comprehensive, quantitative, integrated explanation for many characteristics across the spectrum of animals, plants, and microbes. The theory explains why body size has such a powerful influence on biological structure and functions. The work has application to the noninvasive study of cardiovascular abnormalities, aging, toxicology, pharmacology, forestry and to studies of genetic code and intracellular network structures. T-Division scientists have received considerable honor and recognition for developing this comprehensive theory.

An Overview

Life is the most complex physical system in the universe: its extraordinary complexity is matched only by its extraordinary diversity. Living organisms span a mass range of more than 21 orders of magnitude from the smallest mycoplasma (10^{-13} g) to the largest mammals and plants (>10^{8} g). Overall, the process of life covers almost 30 orders of magnitude from the microscopic terminal enzymatic molecules of the respiratory complex up to macroscopic ecosystems. This range exceeds that of the mass of the Earth relative to that of the galaxy (which is "only" 18 orders of magnitude) and is comparable to that of an electron relative to a cat. Similarly, the metabolic power required to support life over this enormous range varies by over 21 orders of magnitude, a range significantly larger than the Planck (or string) unification energy scale (~10^{19} GeV) relative to that of the mass energy of a proton (~1 GeV).

Over this immense spectrum of biological size, life uses basically the same chemical constituents and reactions to create an amazing variety of forms, functions, and dynamical behaviors. All life functions by transforming energy from physical or chemical sources into organic molecules that are metabolized to do the work of building, maintaining, and reproducing complex, highly organized structures and processes. Understanding the origins, structures, and dynamics of living systems from the molecular level up to ecosystems is one of the grand challenges of modern science. Finding the universal principles that govern this enormous diversity is critical to understanding the nature of living structures and to managing biological systems in such diverse contexts as medicine, agriculture, and natural resource management. Precious few universal principles are known beyond the genetic code and the Darwinian evolutionary process that determine the uniquely biological features observed at all levels of organization.

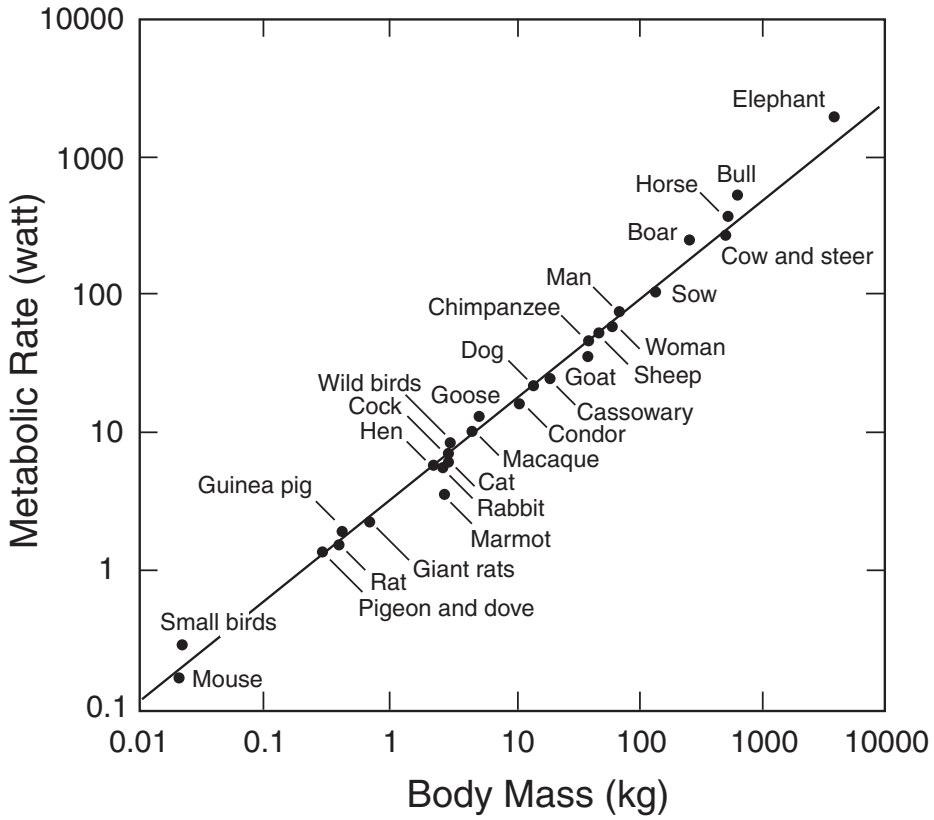


Fig. 1. Basal metabolic rate (in watts) compared to mass (*M* in kg) for mammals and birds on a log-log scale. The slope of the best straight-line fit is very close to 0.75, illustrating Kleiber's law that sets metabolic rate scales as $M^{3/4}$.

In stark contrast to their amazing diversity and complexity, biological systems obey a host of remarkably simple and systematic empirical scaling laws that reflect how organismal features change with size over many orders of magnitude. Among the large number of biological variables that satisfy these allometric scaling relationships (a term coined by J. S. Huxley) are fundamental quantities such as metabolic rate (the power required to sustain life), time scales (such as lifespan, growth rate and heart rate), and dimensions of structures (such as lengths of aortas, heights of trees, densities of mitochondria, and numbers of respiratory complexes). [1]

It is a remarkable fact of biology that these scaling relationships are invariably power laws with exponents that closely approximate simple multiples of 1/4 (e.g., 1/4, 3/4, 3/8). Perhaps the best-known of these phenomenological laws is for basal metabolic rate, which was first shown by Kleiber in 1932 to scale as $M^{3/4}$ for mammals and birds (Fig. 1). This was later extended by Hemmingsen to include ectotherms and unicellular organisms (Fig. 2). A collaboration centered in the Theoretical Division involving James Brown (Regent's Professor of Biology at the University of New Mexico), William Woodruff (Los Alamos), and me, has recently shown how this law can be further extended to over 27 orders of magnitude down to the terminal oxidase molecules within mitochondria (the cellular respiratory machinery). [2] (See Fig. 3.)

When almost any physiological variable is plotted in this way, it reveals a similar power-law scaling behavior with an exponent that is typically a simple multiple of 1/4. For example, heart rate decreases with the 1/4 power of mass, as illustrated in Fig. 4.

These quarter-power scaling laws appear to be valid at all scales and for almost all forms of life, whether mammalian, avian, reptilian, unicellular or plant-like. [1] Their universality clearly indicates something fundamental about how life is organized and the constraints under which it has evolved. For nearly a century, understanding the origin of these scaling relationships and, in particular, why the exponents are almost invariably simple multiples of 1/4, has been a major challenge in biology.

New Principles Postulated

As a result of an unusual cross-disciplinary collaboration between an elementary particle theoretical physicist in T-Division and two biologists at UNM, James Brown and his student Brian Enquist (now on the faculty at the University of Arizona), we believe that the problem of the origin of such scaling laws has been solved. [3] Using the analytic power of mathematics and physics, we proposed a set of universal underlying principles that provide a basis for constructing explicit quantitative physical models for biological systems at all scales. These principles can be viewed as both complementary and supplementary to the genetic code and the principle of natural selection. The basic idea is that, in order to support the huge number of localized microscopic units of an organism, life at all scales is sustained by hierarchical fractal-like branching networks. Functionally, biological systems are ultimately constrained and limited by the rates at which energy, metabolites, and information can be supplied through these networks. The model accounts, in a well-defined, testable fashion, for 1/4-power scaling in diverse biological phenomena and, as such, provides a quantitative framework for addressing many important problems, both basic and applied.

Three generic principles are postulated:

(1) networks must be space-filling in order to service all local biologically active regions in an organism;

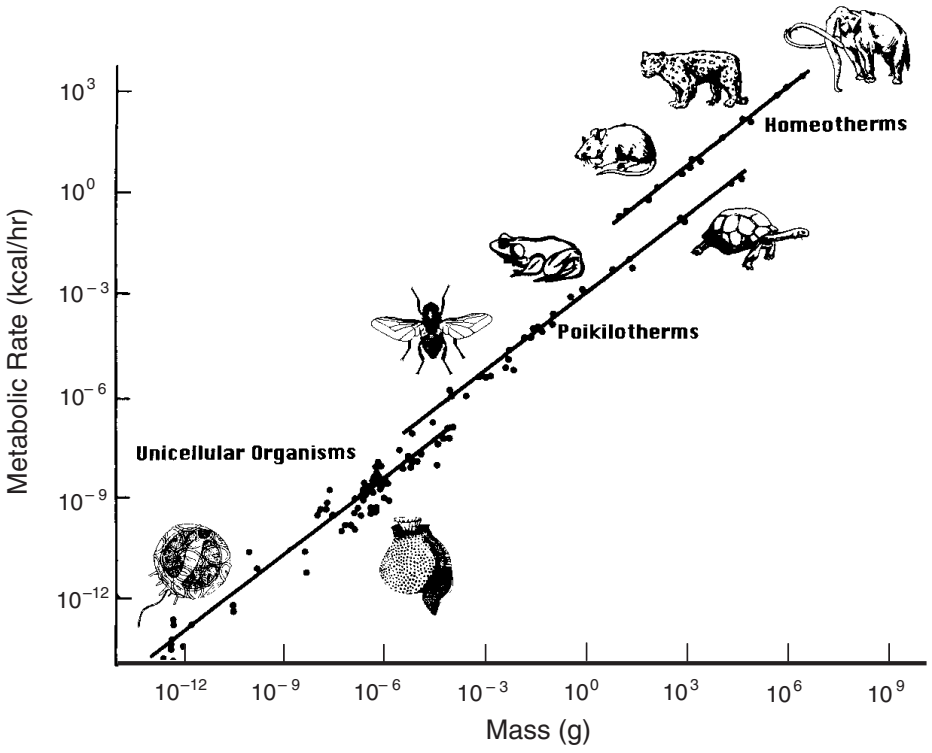


Fig. 2. A similar plot to Fig. 1, showing basal metabolic rate (in kcal/h) compared to mass (in g) for endotherms (homeotherms), ectotherms (poikilotherms), and unicellular organisms. All of the straight lines have a slope of 3/4.

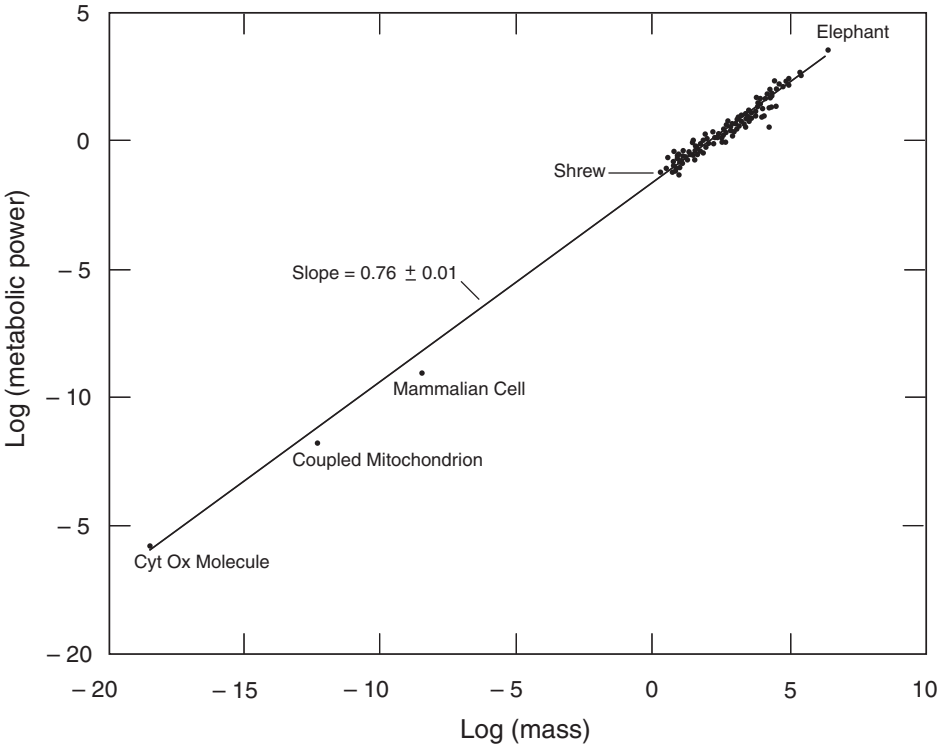


Fig. 3. Extension of Kleiber's scale to more than 27 orders of magnitude down to the terminal oxidase molecules within mitochondria.

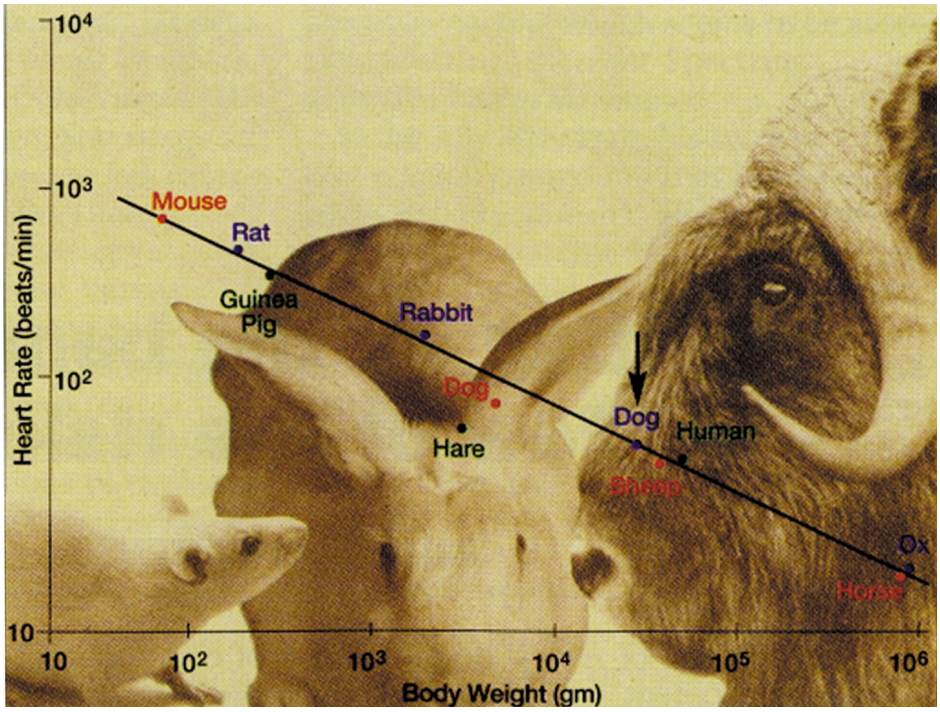


Fig. 4. Logarithmic plot of heart rate (beats/min) compared to body weight (in g) for a sequence of mammals. The straight-line best fit has a slope very close to –1/4.

(2) their terminal units, which interface with the resource environment (e.g., capillaries, petioles, mitochondria, cytochrome oxidase molecules, etc.), are invariant within a class or taxon; and

(3) organisms have evolved so that the energy (and possibly other appropriate quantities) required to distribute resources and sustain them is minimized.

Starting from these principles, biological systems can effectively be described by field theoretic equations of motion with well-defined natural physical cut-offs. Detailed integrated analytic models of the mammalian circulatory and respiratory systems, [3] and of the plant vascular system, [4] have been successfully constructed. Many scaling laws have been derived between organisms (e.g., quarter-power allometry between a mouse and an elephant), within an individual organism (e.g., from the aorta to the capillaries), and during ontogeny (e.g., from a seedling to a giant sequoia). [5,6] The model puts severe physical constraints on the structural design and functional characteristics of organisms leading to a quantitative integrated description of the entire system. Where data exist, excellent agreement is generally found; where data are not available, the model provides testable predictions.

Although the theory was originally motivated with circulatory systems of macroscopic organisms in mind, its applicability is much more general because it is based on generic principles of resource distribution in complex hierarchical network systems. Indeed, we have shown that, if power-law behavior is assumed (reflecting a space-filling fractal-like hierarchy) and scaling of the area of interface of the organism with its resource environment (or, equivalently, the total number of terminal units) is maximal (driven by evolutionary competition), the ubiquitous 1/4 arises because organisms effectively function in four spatial dimensions, even though they physically exist in three. [7]

This principle can be understood as follows. The length of a line, such as a ball of string, that tightly folds upon itself to fill a volume clearly scales linearly with the volume, i.e., as $(\text{length})^3$, rather than as a length. In much the same way, a volume-filling exchange surface, such as the total area of interface of all network terminal units (e.g., capillaries) with the resource environment (e.g., cells), scales with the volume,

i.e., as (length)^3, rather than as a simple Euclidean area, i.e., as (length)^2, where the lengths are network lengths. The volume of the network (e.g., the total volume of blood) and of the tissue serviced (the whole body) both scale as total body mass, i.e., as (body length)^3. However, in terms of network lengths, these scale as (exchange surface area × network length) $\propto$ (network length)^3 × network length $\propto$ (network length)^4. Thus, the ratio of scaling exponents of exchange surface area (which determines metabolic rate) relative to body volume is 3/4. This provides a "geometric" argument, independent of dynamical details, for why 1/4-power scaling dominates network transport systems with very different physical structures and pumps (e.g., pulsatile as compared to nonpulsatile or diffusion as compared to hydrodynamic flow). Motivated by this finding, we recently showed that such scaling extends down to the intracellular and organelle levels, even though details of their transport networks are not understood. [2] Indeed, this finding opens up new and exciting ways of viewing such structures, and one of the future challenges is to begin constructing realistic models of intracellular transport networks.

Theory Extension to Other Systems

The original dynamical model was derived for the pulsatile circulatory system of mammals, which is a hierarchy of branching tubes. Using the same set of basic principles, we derived an analogous model for plant vascular systems, which are effectively fiber bundles of long tubes (xylem and phloem) driven by a nonpulsatile pump. This model for plants and trees, which incorporates biomechanical stability constraints, closely predicts many empirical scaling relationships and answers some long-standing questions, such as how plants deal with resistance in their microcapillary vessels and what determines the maximum height of trees. [4] The model also predicts that vascular tubes must have a small taper as they run from trunk to leaf so that their hydrodynamic resistance is independent of the total path length. This leads to the surprising prediction that the productivity of ecosystems and the flux of power per unit area are independent of plant size, a prediction borne out by observation. In a related paper, [4] we show how we used these ideas to model the structure of plant populations and communities and derived a resource-based thinning law that predicts average population-density scales as $M^{-3/4}$, in agreement with data. This paper received the Mercer Award from the American Ecological Society.

More recently, we began investigating how our principles impact both the temporal and temperature dependence of various biological phenomena. We applied our paradigm to derive a general growth equation from first principles that should apply to all organisms. [5,6] The basic idea is that metabolic energy is transported through the network to cells where it is used either for maintenance (including replacement of dead cells) or to grow new ones. The parameters of the resulting growth equation are determined solely by cellular properties, such as their in vivo metabolic rate and the energy required to create a cell. The model gives a natural explanation for why we stop growing

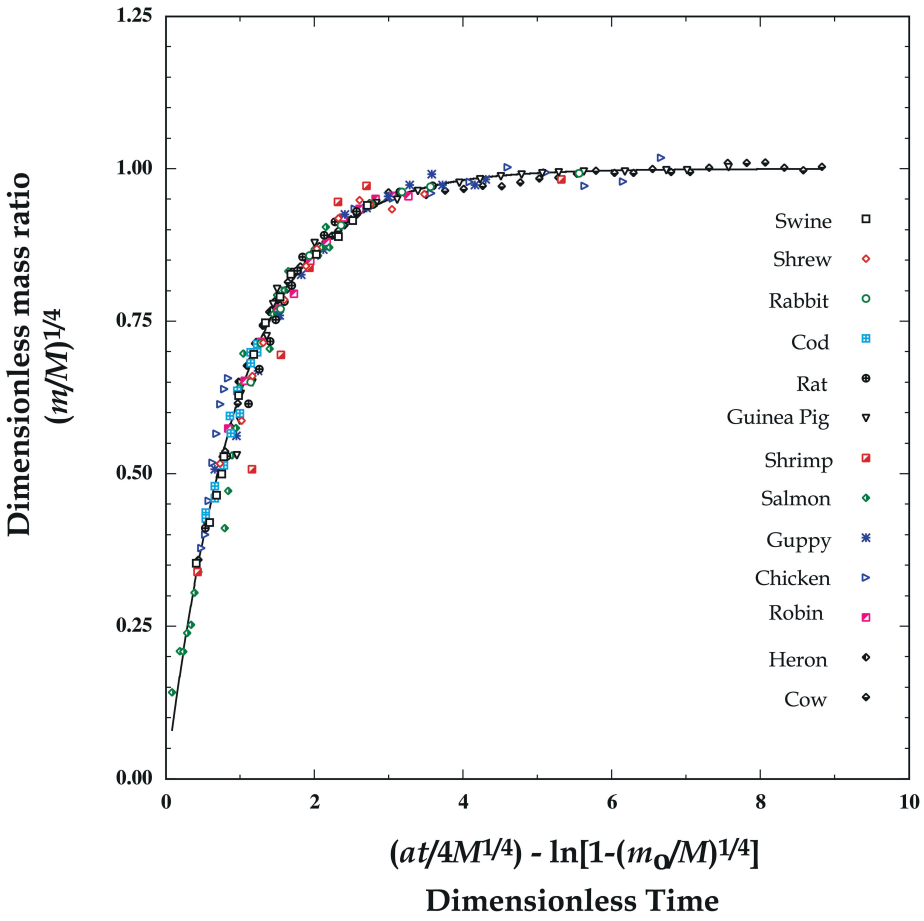


Fig. 5. The dimensionless mass ratio $(m/M)^{1/4}$ compared to a dimensionless time variable showing the universality of growth. Shown are data for mammals, birds, fish, and Crustacea that, when plotted this way, are predicted to lie on a single universal curve; m is the mass of the organism at age t, m_0 is its birth mass, M is its mature mass, and a is a parameter determined by theory in terms of basic cellular properties.

(The number of cells supplied scales faster than the number of supply units, namely capillaries.) and leads to a formula for the asymptotic mass of the organism. From the ensuing equations, we derived a universal parameterless scaling curve for growth on which data for all organisms was shown to fit (Fig. 5). Ontogenetic development can, therefore, be viewed as a universal phenomenon determined by basic cellular properties. In addition, we derived many scaling laws related to growth, such as relative energy devoted to maintenance, doubling times, and so on.

Temperature (T in degrees kelvin, affects metabolism, and therefore growth, through biochemical reaction rates that are governed by a classic Boltzmann factor, $e^{-E/kT}$, where E is the activation energy and k is Boltzman's universal constant. Using our model, we showed that temperature effects in growth, lifespan, and metabolic rates for ectotherms, endotherms, and unicellular organisms can, in large part, be accounted for by using an average chemical activation energy (E ~ 0.6 eV) for the rate-limiting processes responsible for adenosine triphosphate (ATP) production in the respiratory complexes of mitochondria. [8,9] The model fits embryonic development times, spanning a wide range of egg sizes and incubation temperatures for birds and aquatic ectotherms (fish, amphibians, aquatic insects, and zooplankton) and nearly 75% of the variation in post-embryonic development among zooplankton. The remaining variation can be mostly explained by stoichiometry, specifically the whole-body carbon-to-phosphorus ratio. These results suggest a general definition of biological time that is approximately invariant and common to all organisms: to a good approximation, all organisms run by the same universal clock with similar metabolic and growth rates when adjusted for size and temperature.

Since publication of the original paper, our work has received considerable attention in both the scientific and popular press (Examples include feature articles in *Science*, *Nature*, *The Scientist*, *Science News*, *Physics World*, *BioScience*, *The New York Times*, *the Washington Post*, *U.S. News & World Report*, *New Scientist*, and *Financial Times*.) I have given several public lectures and have several times been a lecturer in

"distinguished scientist series" (including Stanford University Medical School, Institute for Advanced Study at Princeton University, The Scripps Research Institute, Bell Labs, Hewlett-Packard Company, and a centenary speaker for the American Physical Society). I have also been a colloquium speaker at many universities and laboratories (such as Princeton, Stanford, the University of Washington, California Institute of Technology, and the U.S. National Aeronautics and Space Administration) and a plenary speaker at many international conferences.

Summary and Conclusions

The paradigm and principles that have been developed suggest novel ways of attacking many important and fundamental problems across the broad spectrum of biology in which quantitative analytic thinking familiar in physics can play a central role. This work has enormous potential at all scales and in a variety of different contexts ranging from the highly practical (such as noninvasive study of cardiovascular abnormalities, aging, toxicology, pharmacology, and forestry) to the conceptual (such as formulating quantitative physical principles that are complementary to the algorithmic principles of the genetic code, or a theory of intracellular network structures). It provides some fascinating and challenging physics questions, such as determining the universality class of biological networks or the dynamics of the evolution, emergence, and formation of such networks. Furthermore, there are clearly interesting possibilities for extending the paradigm into analogous complex hierarchical network structures such as rivers, urban development, and corporations.

The success of the model should be viewed as a beginning rather than an end. By its very nature, the model should be thought of as a "zeroth order" description of "average idealized organisms," which embodies the essential features of biological systems. As such, it can serve as a point of departure for more detailed analyses which would take into account, for example, specific environmental variables that lead to deviations from the predicted idealized behavior. Indeed, one of the conceptual challenges presented by our work is, "Why does it work so well?" Is there some fixed point, or deep basin of attraction, operating within the general dynamical structure that ensures that, in spite of their enormous complexity, the general features of biological systems are robust against significant perturbations?

Up until now the work has predominantly been analytic in nature, using many of the standard, powerful techniques of theoretical physics, including the language of field theory and the renormalization group. In the future, as less tractable problems surface, we will employ computer simulations. Whatever techniques are used, however, it is important to appreciate that systems must be treated as integrated entities. In this sense, biological systems exhibit a long-range order. A subsystem treated in isolation behaves and scales quite differently than the system as a whole. Thus, for example, the extraordinary role of distribution networks and exchange surfaces, in providing powerful macroscopic

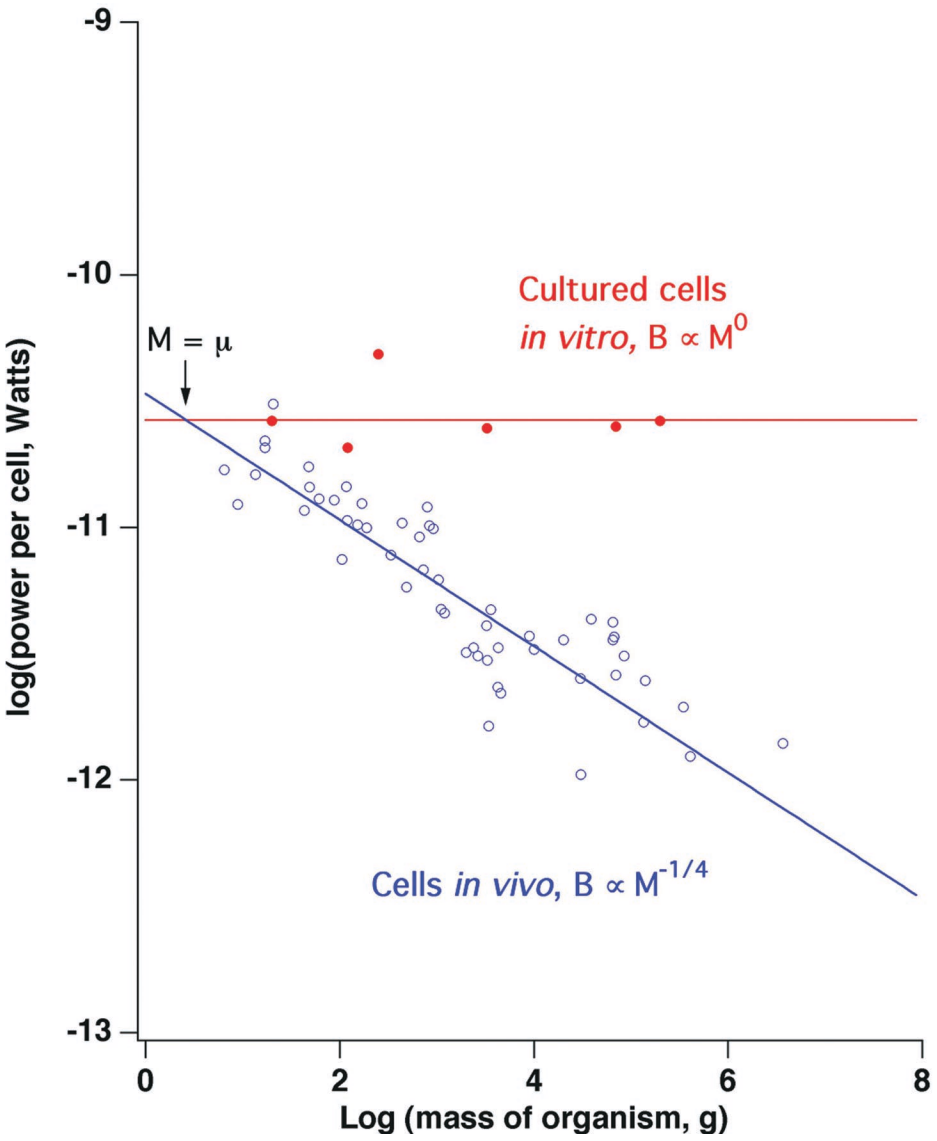


Fig. 6. Cellular metabolic rates (in watts) plotted against organismal mass (*M* in g) on a log-log scale for in vitro (red line) and in vivo (blue line) cells. The latter is predicted to scale as $M^{-1/4}$. When cells are removed from the body, in vitro cellular metabolic rates converge to a constant value predicted by the theory.

constraints on microscopic processes, predicts that, when these constraints are lifted by removing cells from the body, in vitro cellular metabolic rates should converge to a constant value. However, in vivo they should decrease with body mass as $M^{-1/4}$, [2] in agreement with observation (Fig. 6).

It is also important to recognize that we carried out this work in a close, intense collaboration between biologists, biochemists, and physicists; and the research represents an excellent example of successful interdisciplinary science. It is the kind of collaborative interaction that has been the hallmark of T-Division and that can contribute to understanding the structure and dynamics of complex biological systems, potentially leading to significant progress in some of the most fundamental and challenging questions in biology.

In conclusion, our theory offers a comprehensive, quantitative, integrated explanation for allometric scaling of many characteristics in a wide variety of animals, plants, and microbes. It shows how the geometric, physical, and biological constraints on distribution networks and exchange surfaces with fractal-like designs give rise to the ubiquitous quarter-power scaling laws that have so fascinated biologists for the last century. Our theory explains why body size has such a powerful influence on biological structure and function at all levels of organization, from the oxygen affinity of hemoglobin molecules and the density of mitochondria, to the rate of fluid flow in the vessels of plants and animals, to ontogenetic and population growth. It potentially offers the possibility of constructing a fundamental quantitative theory based on universal principles that govern the structure, function, and organization of life at all scales, including the constraints under which it has evolved.

Sequence Analysis
Contributions by Ulam to Molecular Genetics

by Walter B. Goad

This article is reprinted with permission from *Los Alamos Science, Special Issue: Stanislaw Ulam 1909–1984* (Los Alamos National Laboratory, 1987), pp. 288–291.

Results of this Theoretical Division research include GenBank, the national genetic-sequence database.

Lord Rayleigh once introduced a key idea with "It is tolerably obvious once remarked...." Yes, I think now, that is just how it was—Stan Ulam providing us with a steady stream of ideas and observations "tolerably obvious" only in retrospect, and then striking in the way they became integral to one's tangible world of evolved and evolving forms and actions. Here I would like to sketch ideas developed during the 60s and 70s as an avalanche of detail, still growing, gathered about the way sequences of nucleotide bases in DNA encode instructions for development and propagation of living organisms. Stan showed us a very general way of thinking precisely about relationships among sequences, in particular, how to devise quantitative measures of relationship that, together with the computer, are of immense help in ferreting out meaning in the very great quantities of data now pouring forth.

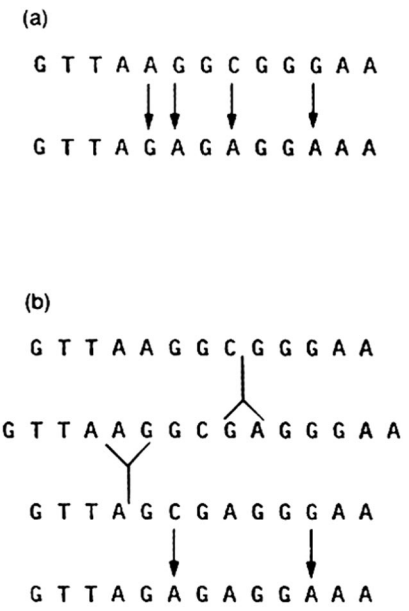
I met Stan soon after arriving in Los Alamos at the end of 1950. I came ostensibly to finish a thesis begun at Duke under Lothar Nordheim, who had arrived several months earlier while I stayed in Durham awaiting security clearance. At last a telegram came from Carson Mark that read, "Your clearance not available." An anxious telephone call established that the "not" had been garbled in transit from "now." I was immediately swept up in the thermonuclear program, kept busy with the rest dissecting schemes and designs, and sometimes new phenomena, usually standing around a blackboard. Introducing the right factors, right at least in order of magnitude, was both vital and enjoyably competitive, laced with humor—esoteric, malicious, or plain—and an occasional flash of ego. The key, of course, was to discern the dominant phenomenon and to estimate its role in the matter at hand. One always had a feeling, almost visceral, as to how deeply an argument was rooted in the web of our knowledge of physics and mathematics. Stan habitually turned things to view from a variety of directions, much as he would see an algebraic structure topologically, and vice versa, and often supplied the connection that dispelled a gathering fog.

Around 1960, Jim Tuck invited Leonard Lerman, who was in the thick of the gathering revolution in biology and then at the University of Colorado, to visit Los Alamos. The "phage group" gathered loosely around Max Delbruck had established a mode of analysis that is still driving the biological revolution: Changes in a single DNA molecule are amplified by biological reproduction, usually in a microorganism, to the macroscopic level; there the consequences of those changes, however ramified, can be studied with the resources of physics and chemistry. The amplification is made possible by an immensely powerful, and growing, armory of molecular tools based on enzymes that carry out specific operations on specific DNAs. As we grasped those ideas from Leonard and began to see the clarity and concreteness with which the mechanisms of life would emerge from such analysis, many of us were galvanized. We soon responded in a way typical of the culture, organizing a seminar, hungrily seeking out the many aspects of the subject. As I recall, the seminar continued through the 60s and early 70s with a varying membership but with Stan, Jim Tuck, George Bell, and me as regulars. We were frequently visited, and enormously encouraged, by Ted Puck, who has built a distinguished school of molecular and cell biology at the University of Colorado and who was, and is, exceedingly optimistic about the contribution systematic theory can make to biology.

A quick tour of systematic theory inevitably would start with Darwin's grand synthesis. For physicists, a key way point would be the publication in 1944 of Erwin Schrödinger's short book *What is Life?*, which equates that grand question with one congenial to physicists: What generates "negentropy," the high degree of order that living systems are continually creating from the environment? Ever since, theorists of all kinds have looked to the formulation of some powerful physical theory of life. Short of that, what we do know is that living systems escape from the determinism of ordinary chemistry by interposing molecular adaptors to control molecular interactions. An example is provision by the complex protein structure of hemoglobin of an effective interaction between O_2 molecules that is completely unrelated to their interactions as free molecules: Within a hemoglobin molecule, up to four O_2s bind at distinct sites and thus effectively stick together. Furthermore, three or four stick more tightly than one or two. So, where there is much oxygen, four are tightly bound: where there is little, departure of one causes the others to more easily depart. Invoking the adaptor principle, Francis Crick predicted the existence of what are now called transfer RNAs—small RNA molecules, a particular species of which adapts each three-base codon to molecules of a particular amino acid. A Zen-like consciousness of physical necessity—for the way in which electrons and nuclei, and thus atoms and molecules, do what they must—leads first to puzzlement at living systems and then to resolution: Molecular adaptors free the logic of higher levels of organization to adopt and express a logic of their own, exploiting, not circumventing, physical necessity.

Proteins and RNAs provide an array of complex and highly specific adaptors, and their structures are encoded in sequences of nucleotide bases in DNA. To a large extent the double-helical structure of DNA wraps the information-conveying part of the DNA into a protected interior and so in the main removes chemical constraints on the propagation and selection of sequences.

Working on DNA as a substrate, evolution has produced the marvelously complex web of living systems we see today. The working hypothesis, to which no exception is yet known, is that all of the information for propagation and development of individual organisms is encoded somehow in the sequence of four bases adenine (A), thymine (T), guanine (G), and cytosine (C) along the DNA molecules (or, in some cases, RNA molecules) that compose its genome. The "somehow" includes the great triumphs of the past two decades, the present frontiers of molecular biology, and, undoubtedly, a great deal that we do not now even glimpse. Less than a decade after Watson and Crick determined the structure of DNA, researchers at the laboratories of Nirenberg, Khorana, and Ochoa fully worked out the "genetic code" by which the base sequences of particular segments of DNA—genes—are translated into sequences of amino acids that fold up as particular proteins. For a few years, many people felt that, in principle, DNA function was now completely understood. But in the mid-70s, methods were worked out for determining sequences of bases in DNA, and it almost immediately emerged that not even the sequences that are translated into proteins are simple, continuous coding sequences. The last few years have seen the discovery of a great many distinct "signals" that control the replication of DNA and the expression of genes. However, it is not yet known how the action of those signals is coordinated, as it must be, to yield the patterns seen during reproduction and development. On the other hand, an outline is emerging of the organization within DNA of repetitive sequences, which make up a substantial fraction of the genome in higher organisms. That organization may or may not have signaling capabilities, but it is almost surely important in evolution. Perhaps most striking of all is the growing knowledge of phenomena—such as the mobility and duplication of pieces of DNA and its rearrangement—that introduce into the genome a degree of dynamism far beyond what classical genetics had led us to suspect. Most of this was yet to come in the late 60s, when the amino-acid sequences of a few proteins were the only biological sequences known. However, it was already clear that the information on which a cell acts is encoded in sequences of bases, and the question of how to characterize relationships among sequences hundreds or thousands of bases long was at hand. With his almost visceral feeling for representation of natural phenomena by general mathematical structures, Stan immediately framed the question in terms of defining a distance between sequences or, more generally, of defining a usable metric space of sequences (Ulam 1972). This he did by considering certain elementary base changes by which one sequence might be transformed into a second: Replacement of one base by another and insertion or deletion of a base. (Combinations of these changes can



DISTANCE BETWEEN DNA SEQUENCES

Consider the two short DNA sequences GTTAAGGCGGGAA and GTTAGAGAGGAAA. As shown in (a), one of these can be transformed into the other by four base solutions. If the "weight" assigned to a base substitution is *x*, then the "measure" of the set of changes in (a) is 4*x*. Alternatively, as shown in (b), one sequence can be transformed into the other by two base insertions, two base deletions, and two base substitutions. Since base insertions (deletions) occur less frequently than do base substitutions, the weight *y* assigned to an insertion (deletion) is different to that assigned to a substitution; in particular *y* is assigned a value greater than that of *x*. The measure of the set of changes illustrated in (b) is 2*x* + 4*y*, which is greater than 4*x*. The distance between the two given sequences is defined as the minimum of the measures calculated for all possible sets of elementary changes that transform one sequence into another.

result from errors in DNA replication, chromosomal crossover during meiosis, insertion of viral or other DNA, or the action of mutagens.) Obviously, one sequence can be transformed into another by more than one set of elementary changes, as shown in the accompanying figure. What Stan proposed was to compute a measure a "size" for each such set and to define as the distance between the sequences the minimum value of the measure.

In simplest form the measure is a sum of weights, one for each of the elementary changes that compose a transformation set. The set corresponding to the minimum measure—the distance between the two sequences—can be interpreted as the minimal mutational path by which one sequence could have evolved from the other. In 1974 Stan, with Bill Beyer, Temple Smith, and Myron Stein, applied the idea of distance to discerning evolutionary relationships among various species from variations in the amino acid sequences of a protein they all share. Also in 1974, Peter Sellers, after hearing Stan talk at Rockefeller University, proved that such a distance can indeed satisfy the conditions of a metric, the most demanding of which is satisfaction of a triangle inequality. Without that, one's sense of what it means for some among several sequences to be close and others distant would be quite unreliable.

Finding the distance between two sequences of length N by brute force, that is, by computing the measures for all the possible sets of elementary changes, requires on the order of $N!$ computer operations. An algorithm for determining the distance in N^2 operations was discovered by the biologists Saul Needleman and Christian Wunsch in 1970 and independently by Sellers in 1974. Essentially, the algorithm proceeds by induction: The minimal set of changes needed to transform the first n bases of one sequence into the first m bases of the other is found by extending already computed minimal transformations of shorter subsequences, then n and m are increased, and so on until the ends of the sequences have been reached.

By the end of the 1970s, it was apparent that DNA sequencing would take off and that investigators from all areas of biology, biomedicine, and bioagriculture would increasingly apply it to their particular research problems. It was also obvious that computer manipulation and analysis of sequences, much of it flowing from Stan's idea for a metric, would play an increasingly large role in exploiting the information. Mike Waterman had joined Beyer and Smith in working on sequence analysis and Minoru Kanehisa, a postdoc from Japan, and I made genetic sequences and their analysis our principal preoccupation from then on. In 1982, a consortium of federal agencies funded GenBank, the national genetic-sequence data bank. Los Alamos collects and organizes the sequence data and Bolt Beranek and Newman Inc. distribute them to users. By the end of 1986, DNA sequences totaling about 15 million bases, from several hundred species, had been deposited in GenBank.

In the 80s, a series of problems in sequence comparison have been faced with varying degrees of success. One problem now solved concerns global versus local closeness (closeness, that is, in the sense of a distance between sequences). Often of interest are sequences that are close to each other although embedded in otherwise unrelated longer sequences. Peter Sellers first introduced the important distinction between local and overall closeness in 1980. A measure suited to the local problem (essentially the number of weighted changes per base, formulated so that the algorithm of Needleman, Wunsch, and Sellers can still be used) was introduced in slightly different forms by Kanehisa and me in 1982 and by Smith and Waterman in 1981. Another class of problems stems from the sheer quantity of data—examining 15 million bases, even with an N^2 algorithm, requires hundreds of hours on a Cray. That problem has been reasonably successfully dealt with by prescreening sequences for likely candidates for significant relationships. A table of pointers to the locations of short subsequences (a simple hash table) is created and searched for short matching sequences. At this writing the method is being implemented with new hardware features of the Cray XMP… .

Devising a metric appropriate to the investigation at hand is probably not a problem that can be precisely posed, much less solved. A simple metric in which each elementary change is given the same weight may well suffice when the object of study is a virus under great pressure to preserve a small genome. But such a metric may show misleading relationships when applied to segments of DNA from a more complicated organism, as Fitch and Smith found in 1983 for mammalian hemoglobins. Some relationships may depend on similarities in three-dimensional structure of DNA that are preserved through a set of sequences, as may be the case for the elements that control initiation of expression of particular genes. To discover such relationships, one needs a measure of structural similarity, expressed of course in terms of sequences. That problem is just beginning to be faced. A good sense of the problem, and of the limitations of sequence comparison, is given by analogy to another idea of Stan's. He proposed that perception, and thought itself, be considered in terms of a metric space. This frames the question: How is the distance between the visual fields corresponding to, say, two tables—which will vary greatly with circumstances—computed in our brains so that it is small compared with the distance between the visual fields corresponding to a table and a chair? Clearly the metric appropriate to a particular class of problems depends on the mechanisms one hopes to discover or illuminate.

Mathematical analysis has spread into nearly every corner of molecular genetics; its spread and development is still accelerating. In early 1986, the Department of Energy took the initiative in seriously exploring sequencing of the complete human genome, some 3 billion bases. In that project computerized management and analysis of information will play a key role.

Speaking of sequence analysis, GenBank, and all that, Stan once said, "I started all this." Yes.

Consider the two short DNA sequences GTTAAGGCGGGAA and GTTAGAGAGGAAA. As shown in (a), one of these can be transformed into the other by four base substitutions. If the "weight" assigned to a base substitution is x, then the "measure" of the set of changes in (a) is $4x$. [See figure on p. 122.] Alternatively, as shown in (b), one sequence can be transformed into the other by two base insertions, two base deletions, and two base substitutions. Since base insertions (deletions) occur less frequently than do base substitutions, the weight y assigned to an insertion (deletion) is different from that assigned to a substitution; in particular, y is assigned a value greater than that of x. The measure of the set of changes illustrated in (b) is $2x + 4y$, which is greater than $4x$. The distance between the two given sequences is defined as the minimum of the measures calculated for all possible sets of elementary changes that transform one sequence into the other.

Reflections on the Brain's Attempts to Understand Itself

by Stanislaw M. Ulam

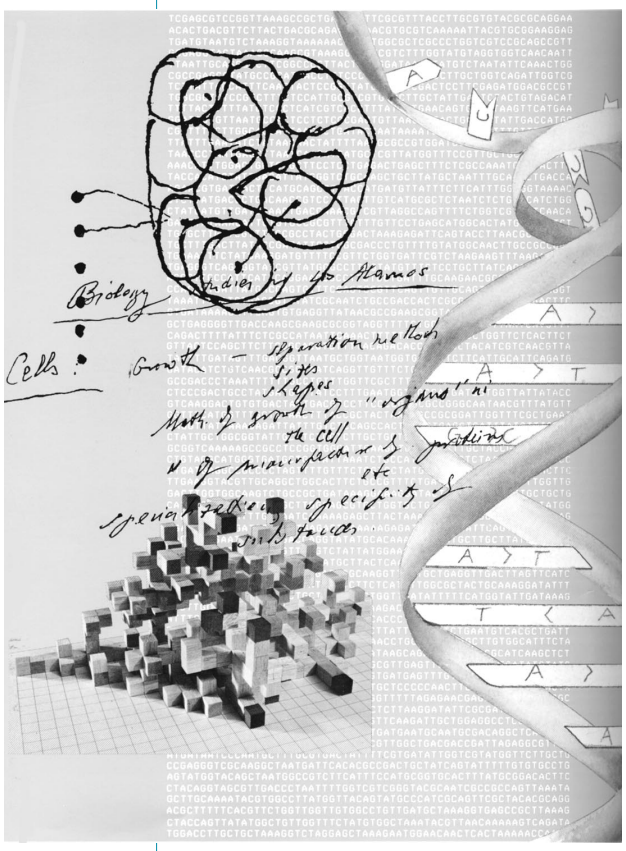
This article is excerpted with permission from *Los Alamos Science, Special Issue: Stanislaw Ulam 1909–1984* (Los Alamos National Laboratory, 1987), pp. 283–287. Ulam delivered this George Gamow Memorial Lecture at the University of Colorado, Boulder, on October 5, 1982.

Theoretical Division mathematician, Stanislaw M. Ulam, made numerous significant contributions in various fields of scientific study. He even applied mathematical equations to understanding some functions of the human brain. His work in this area deserves attention here.

Beyer
Ehrenfeucht
Gamow
Gödel
Hilbert
Margoliash
Mycielski
Schmidt
Ulam
von Neumann

My choice of subject for this talk may seem strange, since I am not a psychologist, a physiologist, or a neurologist, merely a mathematician and an amateur, a dilettante, in the workings of the brain. However, it is fitting that I give such a talk in memory of the late George Gamow, a friend of mine. Though by training a physicist, he was able to make famous contributions in other sciences, such as astronomy and biology, that interested him toward the end of his life. He was, like me, an amateur, a dilettante, in biology. Nevertheless one of the most important discoveries of recent times in that field is due to him. It was Gamow who first pointed out that ordered arrangements of four

chemical units—four "letters"—along the DNA double helix, or chain, as he called it, might be codes for many biological processes and that the codes for the manufacture of proteins might consist of three- or four-letter "words."



What I want to do today is talk about several of my own speculations, with some mathematical symbolism, concerning the operation of the brain. I believe that discoveries and breakthroughs within the next 20 yr will lead to a better understanding of the mechanisms of the brain, of the processes of thought. It will not be a complete understanding—that would be too much to hope for—but it will give us some ideas of how the nervous system operates in lower animals and in humans.

Mathematicians may help in reaching this understanding, although, for the time being I think that 99% of the progress will come from physiological and anatomical experiments. However, mathematics can be useful, for it is clear that the similarities between electronic computers and the nervous system are of great importance.

Another friend of mine, the late John von Neumann, was one of the pioneers in the planning and building of electronic computers. His book *The Computer and the Brain*, which was published posthumously in 1957, is still one of the most elegant and understandable general introductions to the subject. I remember the discussions we had on how the advent of computers would enlarge the scope of experimentation in mathematical and physical sciences, and about his specific interest in the partial analogies between computers, as they were planned in the early 40s, and the processes of deductive thinking. We saw each other frequently at the time, either in Los Alamos or in Princeton, and we would marvel at the few physiological facts then known about the brain, such as the number of neurons it contains. That number was of the order of 10 billion, and their interconnections in the human cortex were known to be still more numerous. He would say, "Not only are there 10 billion computing elements, but each is connected to many others, one hundred maybe! And maybe even to one thousand in the central part of the brain."
Well now, forty years later, the number of interconnections has been shown to be of the order of thousands, up to 100,000 in the central part of the brain. And the total number of connections, of axons and synapses, is on the order of 10^{14}. So you see, in the recent past, the purely anatomical and physiological knowledge has vastly changed.
The locations of certain centers in the brain and the differences between its right and left halves are also better known. And today, more information is being gathered through studies of the electromagnetic signals being emitted constantly by the brain.

However, I do not believe that now, or even in the near or distant future, it will be possible to gain what might be called a complete understanding of the brain's operation. My belief rests on very important and strange results in pure mathematics. These results, which date from 1930, are associated mainly with the name of Gödel, a mathematician who worked at the Institute for Advanced Study in Princeton. Gödel proved a theorem that says, roughly speaking, that in any mathematical system, any logical system, there exist statements that have sense but cannot be proved or disproved. So, in every mathematical discipline one can conceive of at present, there are undecidable propositions, finite statements that, starting from axioms, one cannot demonstrate or show to be false.

Mathematics has a store of problems, some very old, whose solutions are not known. But it was assumed that, ultimately, yes or no solutions would be found. That was the belief of Hilbert, one of the greatest mathematicians of the last hundred years. Then Gödel came along and showed that such a belief is no longer valid, that there are statements that are undecidable. This fact is of great philosophical significance. And beyond that, it could be a sort of consolation for our inability to attain a complete knowledge of various real phenomena.

So it is possible that some of the still unresolved mathematical problems are in principle undecidable on the basis of our present system of axioms. Many such problems are technically complicated, but let me give you one that is simple to state and understand.

A prime number is an integer that is not divisible by any number except itself. The numbers 2, 3, 5, 7, ..., 41, 43, 47, et cetera, are all prime. The Greeks knew that there are infinitely many prime numbers. That is one of the oldest, greatest, and most beautiful discoveries in mathematics. Now certain pairs of prime numbers, such as 5 and 7, 11 and 13, 17 and 19, are called twins because they differ by only 2. The question is: How many twin primes are there, a fixed finite number or an infinity? Nobody knows the answer to this question, and it may be undecidable. I asked Professor Schmidt, a very famous number theorist, if he knew who first proposed this very old problem and whether he thought it might he undecidable. He did not know the answer to the former and to the latter he answered, "One might not be able to decide whether it is undecidable!"

I mention Gödel's theorem to show the limitations of man's program to try to understand everything, even in a restricted domain. Perhaps the scope of the human brain is finite, or conversely, perhaps the growth of humanity, of its collection of brains, will, in terms of evolution, continue indefinitely and may reveal new points of view.

To continue the speculation on what the role of mathematics might be in the study of the brain, the time is not yet ripe to say its operation can be understood with abstract theories alone. But Gamow, who was perhaps the last great amateur in science, has shown us that it is possible to speculate fruitfully, given some luck on the great mysteries of nature. A Greek

philosopher said that many are the wonders of the universe, but the greatest of all is the human mind. And Spinoza said that it is better to begin with small and modest truths. Starting from these premises, I want to give you now a few examples of biological questions that I think mathematics has already proved somewhat useful in answering, and how similar attempts and schematizations might possibly be of some use in partially understanding the nature of human perception.

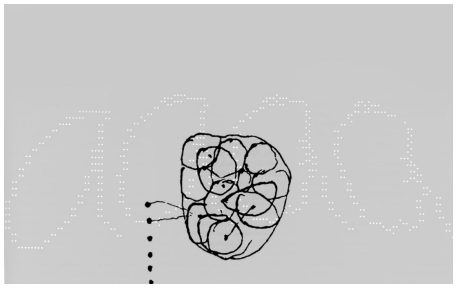
One such question concerns the mechanism of recognition of external stimuli, say sights or sounds, and ultimately of ideas. Before recognition, there is perhaps discernment, discrimination. *A priori*, it seems easier to see the difference between two objects than their similarity or analogy. We need to map the tremendous web of connections in the human brain into overlapping classes. But before we do this, here is an example of a mathematizable biological idea, one concerning the codes for the manufacture of proteins.

Gamow's suggestion about the existence of three- or four-letter codes for the constituent amino acids of proteins was almost correct. Many of the characteristics of living organisms are coded in very long sequences of four chemical units, which biologists call by the letters A, C, G, and T. Words are short strings of these letters. Finite sentences of several hundred words are codes for proteins, such as hemoglobins of various kinds. Today, tens of thousands of these codes for proteins are known, and in some cases even the spatial forms of the proteins are known. A "reader" molecule goes along the DNA "tape," reads the code, and deposits the information in other parts of the cell, in the ribosomes. This much is now understood. The functions of other parts of the long sequences, such as those called introns, are not yet understood, but they are not codes for proteins.

Some biologists are beginning to speculate on the importance of small differences that have been found to exist between the codes for a given protein in different species. For example, cytochrome c, which is important for the transmission of electrical impulses in nerves, differs slightly from one species to another but remains the same within a species. The biologist Emanuel Margoliash has tried to establish an evolutionary tree based on the quantitative differences in cytochrome c codes, on the gradations among them.

Continuing from p. 286:

The conjecture is that in the brain, in the visual system and in the memory, perhaps only a few visual perceptions are permanently stored and, when presented with another, the brain produces, for comparison, many deformations either of what is in the memory or of what is presented. If this is so, the storage capacity of the memory would be enormously enhanced.



At present one can only speculate about the mechanisms by which the brain might produce the deformations. Some are obvious, such as a tilt of the head or a change in size. One can also only speculate about what distances or how many are used in the decision. One may also speculate that a similar mechanism directs the recognition of objects within the body. Could it be that the antibodies produced by the immune system have an analogous way of recognizing antigens? Again, deformations might he used to produce a large number of examples for such discrimination and recognition.

The next higher stage in the operation of the brain might he a more complicated analysis of impressions. Instead of considering impressions of single objects, the brain might study a succession of two or three, even a "movie" of 10 or more. Combined with recognition of the passage of time, this could lead to development of primitive logic or elementary reasoning, perhaps in the form of the statement *post hoc ergo propter hoc* (after, therefore because) or its reverse *ante hoc ergo qua hoc* (before, therefore as a reason for).

Our comprehension of less elementary learning should involve the mathematical idea of measuring complexity. In recent years quite a number of mathematicians, including Jan Mycielski and André Ehrenfeucht, both professors at this university, have done some very interesting work on this subject. With proper changes, some of their results could be applied to investigating the operation of the nervous system.

It is clear that one of the most important mysteries about the brain is the organization of the memory, including the means of access. As I surmised earlier, some form of memory must exist in the visual, auditory, olfactory, and immune systems—and even in the system for differentiation itself. A mechanism for producing many examples from one would certainly seem a very efficient way of using the storage capacity of the visual and auditory memories. In the course of evolution, special devices, or tricks, must have developed to increase the scope of recognition and of the complementary process of registering perceptions as new.

Let me give an example of a trick for efficient use of a computer. Suppose we have stored in its memory a great many, say 10^6, eight-digit numbers arranged sequentially and want the computer to decide whether a given number is among those stored. The computer can do this extremely fast by comparing in succession the digits from first to last. Suppose now that we want the computer to decide whether the given number differs from any of the stored numbers by, say, 1 in any of the eight positions. We might program the computer to do this by deciding whether any of the 10^6 numbers in its memory is that close. That would be a very lengthy operation. There is a much better way to proceed, a way that requires only sixteen times the effort required for the computer to decide whether a single number is among those stored.

We first program the computer to produce from the given number the sixteen numbers that do differ by 1 in any of the eight positions and then to decide whether any of the sixteen is among those in its memory.

This example illustrates that a mechanism for producing auxiliary perceptions for comparison with perceptions stored in the memory would be an advantageous acquisition of the nervous system. So also would a mechanism for producing variations of what is stored in the memory for comparison with external stimuli. Perhaps a physiological or anatomical arrangement might serve such functions. Clearly these are merely guesses as to special characteristics the nervous system may have acquired in the course of evolution.

An Ulam Distance

by William A. Beyer

Stan had often referred, as he did in this lecture, to a distance between sets based on an encoding of the set points in terms of orthogonal functions. However, he had never explicitly defined such a distance. I do so now to honor the originator of so many seminal ideas.

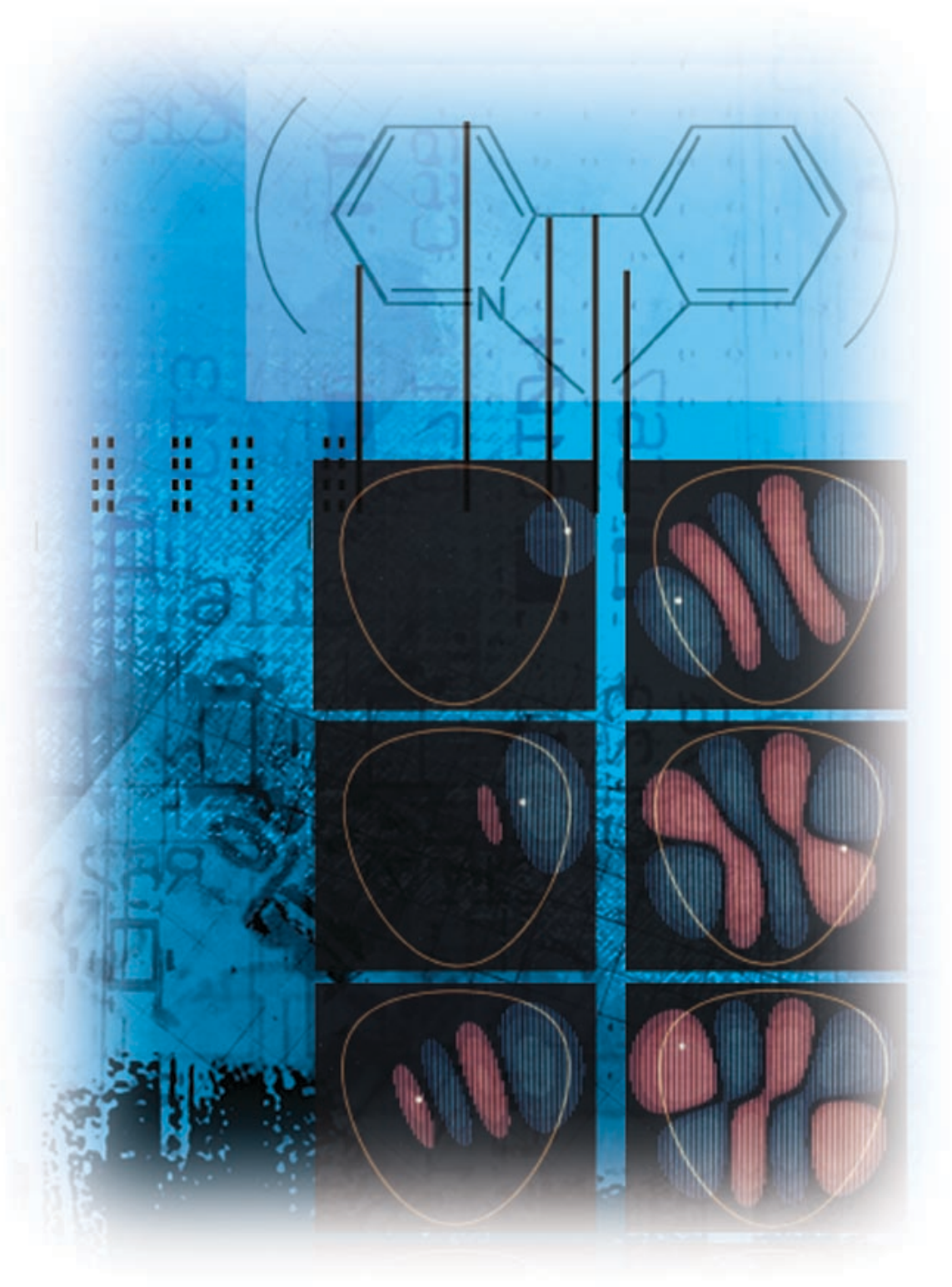
Let A and B be two-dimensional finite sets enclosed in a square. Let n_A and n_B be the number of points in A and B, respectively. Let $\{f_{i,j}\}$ be a complete set of orthogonal functions on the square, such as 2-D Fourier trigonometric functions. Define $\mu^A_{i,j}$ and $\mu^B_{i,j}$, the encodings of A and B mentioned above, as follows:

$$\mu^A_{i,j} = \frac{1}{n_A} \sum_{x \varepsilon A} f_{i,j}(x)$$

and

$$\mu^B_{i,j} = \frac{1}{n_B} \sum_{x \varepsilon B} f_{i,j}(x) \quad .$$

Then $\mu^A_{i,j}$ and $\mu^B_{i,j}$ are functions on the nonnegative lattice points of the plane. Finally, let $r(f_1, f_2)$ be some selected distance between such functions. Then $\rho(\mu^A_{i,j}, \mu^B_{i,j})$ is a distance between the sets A and B—an alternative to the Hausdorff distance defined in the lecture.


N

Extending Quantum Chemistry through the Periodic Table

by P. Jeffrey Hay and Willard R. Wadt

R. Cowan
Hay
Kahn
Martin
Wadt

The description of electronic properties of molecules forms the fundamental basis of chemistry. Concepts of chemical bonding and reactivity address the distribution of the electrons about the nuclei in the molecule. The strength of chemical bonds in turn determines the fundamental energies involved in chemical reactions. The solution of the molecular Schrödinger equation provides the molecular wave function that describes the electronic distribution about the nuclei. From this wave function, the electronic properties, bond energies, and spectroscopic quantities are predicted that can be compared to experiments. In this segment, the development of quantum chemistry during the 1970s and 1980s is discussed in the context of the Theoretical Chemistry Group which emerged from the Laser Theory Group in Theoretical Division in 1976. In particular, the challenges of carrying out calculations involving molecules comprised of heavier atoms in the periodic table are described and the impetus for such calculations received from the laser programs at Los Alamos. This work led to innovations in widely used quantum chemistry software and to the "Hay-Wadt potential" for treating elements in the lower levels of the periodic table in molecular electronic structure calculations.

Quantum Chemistry in the 1970s

During the 1970s, researchers in many of the activities in Theoretical Division were addressing processes in molecules of interest to the inertial confinement fusion (ICF) program and the molecular laser isotope separation (MLIS) program. At this time modern quantum chemistry was still in relative infancy. Solution of the molecular Schrödinger equation using "first principles" approaches was restricted both in terms of the size of molecules that could be treated and to the nature of the atoms of which the molecules are composed. Calculations were being carried out on such species as diatomic molecules, simple hydrocarbons, and other organic molecules. Some attention was given to second-row compounds containing sulfur and chlorine and only limited applications to work on first-row transition metal compounds.

The main reason for this situation is that the computational effort increases rapidly with the number of electrons in the molecule. In these calculations one of the key steps is obtaining the 1-electron wave functions, or molecular orbitals (MOs) that describe the electrons in the molecule. The total electron density in the molecule at a particular point in the molecule is obtained by adding the value of the square of each orbital $|f|^2$ at the point to obtain the total density. These MOs are represented mathematically in terms of atomic orbitals centered on the

nuclei. Overall the computational effort in the solution of the molecular Schrödinger effort scales as N^4 with the number of atomic orbitals, N, and the number of orbitals is roughly proportional to the number of electrons in the molecule.

As an aside, we should mention that one of the important principles in such calculations is the decoupling of the rapid electronic motion from the slower motion of the nuclei according to the Born-Oppenheimer Approximation, which treats the positions of the nuclei as fixed in solving for the MOs. This development was due to J. Robert Oppenheimer, the first director at Los Alamos, during his postdoctoral studies with Prof. Max Born in Germany.

One should also remember that this was the era of the early "computer mainframes," and Los Alamos was a leading center with CDC 7600 and later the first Cray supercomputers. These capabilities enabled calculations to compute the electronic states of small molecules such as fluorine, chlorine, and krypton fluoride of interest to laser programs at Los Alamos and elsewhere in the world.

Development of Relativistic Effective Core Potentials

The MLIS program was an attempt to demonstrate the feasibility of separating uranium isotopes for possible application in the commercial nuclear fuel cycle. The concept was based on exploiting laser chemistry to preferentially excite one of the isotopic species $^{235}UF_6$ and $^{238}UF_6$ in a gaseous mixture. While much theoretical and experimental progress was made on the interpretation of the fine details of the infrared spectroscopy of these molecules, there was also interest in understanding the ground-state chemistry and excited-state photochemistry of UF_6.

For a molecule such as UF_6 the sheer number of electrons for an atom as heavy as uranium presented a daunting barrier for quantum chemistry calculations. However, nearly all chemistry is determined by the outer valence electrons which are the most weakly bound to the nucleus. The tightly bound inner-core electrons are chemically inert for all practical purposes. It had occurred to many people to replace these annoying inner electrons by an "effective core potential," or alternatively a "pseudopotential." This concept has been especially prevalent in the solid-state literature since the early days of quantum mechanics, and there was also activity in the quantum chemistry community to develop reliable potentials.

A key condition for the use of effective core potentials (ECPs) is that one should get the same overall result treating only the valence electrons as if one had done the calculation explicitly including all the electrons. The only major difference is that the orbitals, or more precisely the "pseudoorbitals," in the presence of the core potential are much smoother than the original valence orbital, which was more complicated in the presence of the core electrons. The removal of the core electrons and the smooth nature of the pseudoorbitals dramatically

reduced the number of atomic orbitals needed in a calculation compared to the original all-electron calculation.

Through support of the MLIS program, a collaboration was begun with Luis Kahn of Battelle Columbus Laboratories, who had already been working on effective core potentials. Using his approach to construct core potentials for halogen atoms such as chlorine, for example, one could compare the result with all-electron calculations; and this led to improvements in the process of obtaining core potentials.

Relativistic Effects

Even with the problem of the core electrons surmounted, however, another challenge arose: relativistic effects become important for molecules comprised of the heavier atoms in the periodic table. Physically the classical "velocity" of the electrons in heavier atoms (starting about atomic number Z = 30, or roughly around copper and zinc in the periodic table) starts to become appreciable relative to c, the speed of light. If one were to ignore these effects and solve the nonrelativistic Schrödinger equation one would introduce significant errors. Another important collaboration involved Bob Cowan in the T-Division who had spent a career in theoretical atomic physics. In particular, he and a former colleague D. C. Griffin had developed a relatively simple approach for solving the Schrödinger equation for heavy atoms, which had become known as the Cowan-Griffin method. This method was much simpler than other methods for treating heavy atoms but gave good predictions of the atomic properties of heavy atoms. More important, the orbitals from Cowan-Griffin could be readily adapted to the ECP approach to incorporate both the relativistic effects and to remove the core electrons.

Using these approaches, we generated relativistic ECPs for heavy atoms, including gold, mercury, and uranium and proceeded to carry out molecular calculations as well. [1] The importance of relativistic effects is illustrated for the gold atom in Fig. 1, where inclusion of relativistic effects alters the calculated atomic transition energies several eV (increasing in one case by 56% and decreasing by 64% for the other case). This actually reverses the relative ordering of states as predicted in a nonrelativistic calculation. In addition, relativistic effects alter the size of the orbitals in the atom, including even the outermost valence electrons, which one naively would assume would be less affected by relativistic effects. The change in the energies and sizes of orbitals, in turn, has profound effects on the chemical properties, such as bond lengths and bond energies. In the case of a simple diatomic hydrid molecule, gold hydride (AuH), the calculated bond length decreases from 1.81 to 1.52 Å when relativistic effects are included.

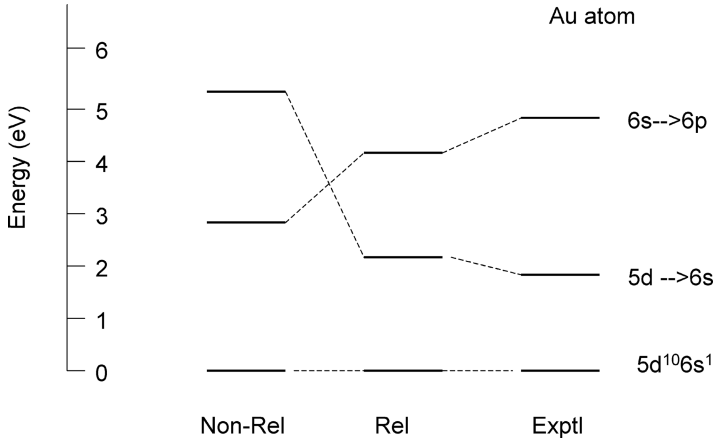


Fig. 1. Relativistic effects in gold atom.

The latter value is very close to the experimental bond length of 1.523 Å. Similar effects are seen on the bond energies. In fact, more detailed study of the influence of relativistic effects in the bulk metal explains gold's color and lack of chemical reactivity, which account for humanity's fascination with this shiny substance.

In UF_6 where the fluorine atoms are arranged in an octahedral structure about the uranium, the calculations using relativistic effective core potentials gave remarkably good predictions for the uranium-fluorine bond length and for the energies of the positive and negative ions when compared to experiments. [2] The analysis of the contributions of the atomic orbitals of uranium to the resulting MOs showed uranium involvement of both the 6d and 5f orbitals to the bonding in UF_6. As one proceeds along the actinide row of the periodic table containing uranium, one is filling the 5f atomic shell, but the low-lying 6d shell is also very close in energy. Further studies on the molecule UF_5 were more speculative because there was much less definitive experimental information on this reactive intermediate. The calculations showed a very "floppy" molecule which could change structure with relatively little barrier between structures with 4-fold and 3-fold symmetry. [3] Later more extensive calculations on the electronic-excited states of UF_6 showed that the optical spectra over the region 3–8 eV arose from charge-transfer excitations from bonding orbitals to essentially nonbonding 5f uranium orbitals. In addition, the assignments of these and the electron-impact energy-loss spectra, which shows forbidden transitions, could be made. [4]

In the early 1980s, the uranium MLIS program came to an abrupt end, but a second program came to the fore to assess the applicability of laser chemistry for isotope separation of plutonium using PuF_6. In contrast to UF_6, the excited states of PuF_6 are a combination of excitations among the 5f orbitals ("f-to-f" transitions) of plutonium and charge-transfer excitations from the F 2p orbitals to the plutonium 5f orbitals. Calculations carried out using ECPs on the f-to-f excitations proved to be remarkably accurate and provided insights into the relative ordering and nature of these transitions. [5]

Later Developments

While the main emphasis in the initial stages had been making the "leap" to actinides in the "high-Z" part of the periodic table, the technology was now in hand to perform calculations in other areas of the periodic table, such as transition metal chemistry. In addition to the studies already mentioned on gold, the first *ab initio* calculations on transition metal complexes of third-transition-series metals such as platinum and rhenium were carried out. [6]

Other key developments came from Richard Martin's arrival in the late 1970s in the Theoretical Chemistry Group. In early collaborations, the relativistic effects on transition metal atoms and their implications for chemistry were systematically studied. [7] Martin also implemented an alternative method for evaluating the matrix elements needed in calculations with core potentials that had been developed by McMurchie and Davidson of the University of Washington. These developments made the calculation of forces on atoms in a molecule much easier. This enabled the optimization of geometries when one lets the molecule relax to the equilibrium structure when there are no net forces operating on any of the atoms. These capabilities were incorporated into several electronic structure codes which are still used in the group. One of the important collaborations involved Steve Binkley of Sandia National Laboratories and Michael Frisch of Carnegie-Mellon University, who were among the developers of the electronic structure code that has evolved to GAUSSIAN98 at this writing. This has become the most widely used software by the chemical community for performing electronic structure calculations.

As the technology of core potentials was disseminated freely throughout the community, it became increasingly more important to have a standardized set of potentials for users to carry out such calculations rather than generating them on a case-by-case basis. At that time, other groups were developing alternative approaches to core potentials including efforts by Prof. Ken Pitzer and students at the University of California, Berkeley, Morris Krause and Walter Stevens of the U.S. National Bureau of Standards, now the National Institute of Standards and Technology, and Professor Stoll and colleagues at Universität of Stuttgart, Germany. With essentially no "official" funding and no hordes of graduate students, we initiated an effort in the early 1980s to develop a consistent set of ECPs. In between our principal management jobs over the period of 2 yr we generated the potentials for the transition metals scandium to gold and essentially all the main group elements starting with sodium through bismuth (Fig. 2) along with the necessary basis sets needed to perform molecular calculations. We published these in 1985 in a set of three papers [8–10] describing what have become to be known either as the "Hay-Wadt" potentials or the "LANL" potentials as they have become designated in the Gaussian programs. They permitted the first systematic studies by the quantum chemistry community of molecules comprised of elements in the lower levels of the periodic table. (Fig. 3) To date, the original references have been cited several thousand times as they have become used throughout the world.

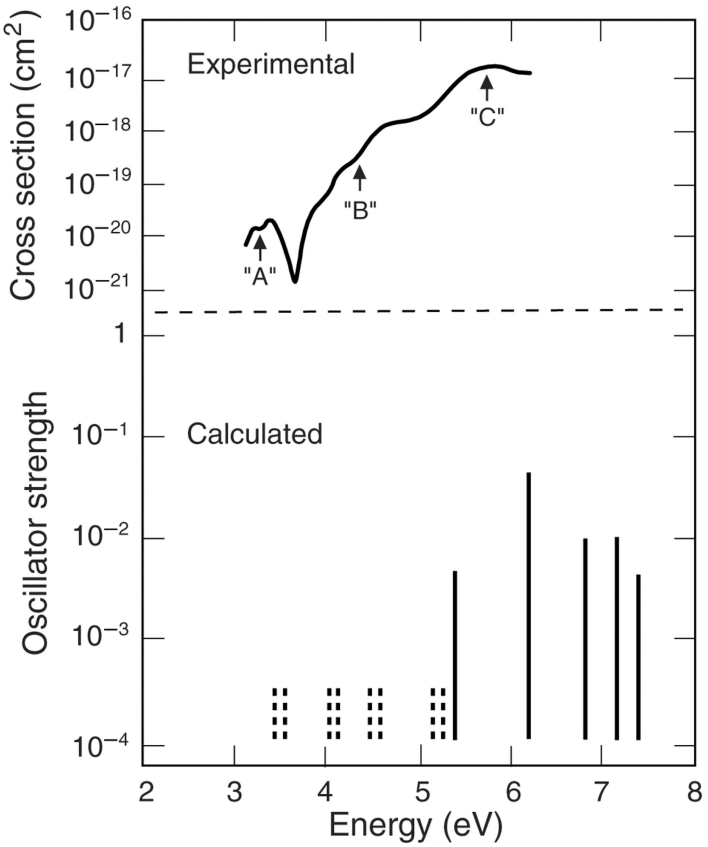


Fig. 2. Experimental spectrum of UF_6 and calculated electronic transitions. Dashed lines indicate forbidden transitions. [4]

H																	He
Li	Be											B	C	N	O	F	Ne
Na	Mg											Al	Si	P	S	Cl	Ar
K	Ca	Sc	Ti	V	Cr	Mn	Fe	Co	Ni	Cu	Zn	Ga	Ge	As	Se	Br	Kr
Rb	Sr	Y	Zr	Nb	Mo	Tc	Ru	Rh	Pd	Ag	Cd	In	Sn	Sb	I	Te	Xe
Cs	Ba	La	Hf	Ta	W	Re	Os	Ir	Pt	Au	Hg	Tl	Pb	Bi	Po	At	Rn
Fr	Ra	Ac															

Ce	Pr	Nd	Pm	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lu
Th	Pa	U	Np	Pu	Am	Cm	Bk	Cf	Es	Fm	Md	No	Lr

Fig. 3. Regions of the periodic table where effective core potentials were developed.

At Los Alamos these effective core potentials were used to study various molecules involved in inorganic chemistry (Fig. 4), including the first dihydrogen W complexes, synthesized by Greg Kubas at Los Alamos, and iridium dopant complexes for light-emitting diodes. In recent years, a resurgence of interest in actinide molecules and solids at the Laboratory and elsewhere has led back into this area. [11]

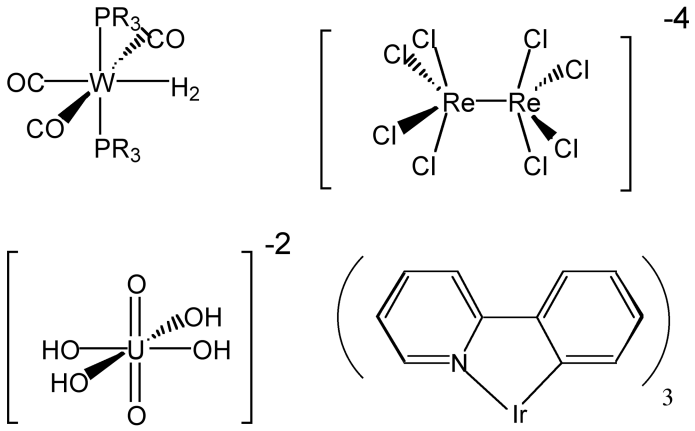


Fig. 4. Molecules studied in recent years using relativistic effective core potentials.

Retrospective

Today quantum chemical calculations are routinely performed on systems that were not possible 25 yr ago. This is a result of a combination of factors, including increase in computer capabilities, improvements in software, and the ECP developments. The environment at Los Alamos enabled this progress that would have been difficult at other institutions at the time.

Donald Stokes, [12] in his insightful treatise on research and development, grouped research according to two criteria:

(1) whether the research is inspired by a quest for fundamental understanding or strictly for applied purposes, and

(2) whether there were potential considerations of end-use of the research.

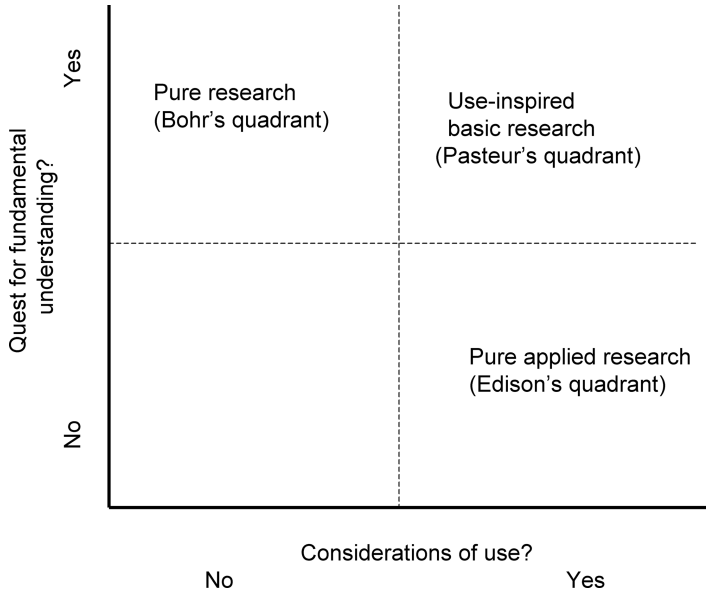


Fig. 5. Donald Stokes' taxonomy of research and development. [12]

Depending on the answers to the motivation for the research, it is categorized into four quadrants (Fig. 5). Bohr's quadrant, where research is carried out primarily for fundamental understanding with no particular consideration of use, reflects the research generally found in universities. Edison's quadrant, applied research motivated primarily for actual use, reflects the research generally found in industrial laboratories.

Pasteur's quadrant represents activities primarily driven by a desire for fundamental understanding but with potential use of the research also a key component. The development of effective core potentials, as driven to a large degree by the needs of the uranium MLIS program, represents a wonderful example of use-inspired basic research. From this standpoint, the Manhattan Project is probably the most famous and important example of Pasteur's quadrant. The strength of the research at Los Alamos National Laboratory continues to flow from the synergism between serving society and pursuing the basic intellectual curiosity that underpins science. That synergism was certainly the engine behind our success on ECPs.

Acknowledgments

In addition to the collaborators specifically mentioned here, we also wish to acknowledge our colleagues David Cartwright, Paul Robinson, Reed Jensen, and Richard Burick in the MLIS programs that supported these activities and our experimental and theoretical colleagues at Los Alamos. Finally, we thank George Bell, T-Division director during much of this period, who played a key role in maintaining the structural integrity of the Theoretical Division during the times of rapid funding changes.

Body-Frame Axes and Angular Momentum Decoupling in Molecular Collisions

Archer
Arthurs
Bacic
Bernstein
Born
Butcher
Clebsch
Colavecchia
Curtiss
Dalgarno
Gordan
Hayes
Hunter
Kendrick
Kouri
Kress
Lagana
Lester
McGuire
Oppenheimer
Pack
Parker
Schrödinger
Secrest
F. Smith
Takayanagi
Tsien
Walker
Whitten

by Russell T Pack

All chemical reactions involve molecular collisions; hence, it is important to understand those collisions and have an accurate working theory for them. Molecular collisions occur in a different regime from the collisions studied in most other branches of physics; namely, a semiclassical regime in which behavior is often almost classical, but in which quantum effects also often come back to bite those who do their calculations using classical mechanics. Hence, a quantum theory of molecular collisions is required if chemical reactions are ever to be truly understood. Theoretical Division researchers have developed such theories that have become standard tools in theoretical chemistry and molecular physics.

At the rather low (thermal) energies of molecular collisions,

(1) only a few electronic states of the molecules are usually energetically accessible;

(2) several vibrational states are often accessible; and

(3) hundreds of rotational states are accessible, populated, and strongly coupled by the collision.

It is this last fact and the fact that reactive (rearrangement, dissociation, etc.) collisions are often also possible that cause the difficulties.

Let us consider one of the simplest possible examples; namely, the collision of a ^{1}S state atom A with a $^1\Sigma$ state diatomic molecule BC with only one electronic state energetically accessible. Then, solution of the electronic Schrödinger equation in the Born-Oppenheimer [1] approximation produces a potential energy surface (PES), which is an effective potential in which the nuclei are then allowed to move. We assume that our quantum chemist friends have already solved the electronic problem for this example system. That reduces the many-body problem to a three-body problem. (I pause briefly at this point to note that the Oppenheimer referred to here is our own J. Robert, the first Director at Los Alamos, and he never was given the credit he deserved for this remarkable approximation which lies at the heart of all of theoretical chemistry.)

Now, let us consider a case in which the energy is too low to produce any exchange or dissociation reactions. Then, the number of molecular states that one must include is $N = N_v \times N_r$ where N_v is a little larger than the number of energetically accessible vibrational states of the molecule, and N_r is a little larger than the number of accessible rotational states. If the highest vibrational state included has vibrational

quantum number v_{max}, then the number of vibrational states is simply $v_{max} + 1$, and v_{max} is typically of the order of 10, so that $N_v \sim 10$ is not much of a problem. However, each rotational state labeled by quantum number j is $(2j + 1)$-fold degenerate because of the m-type quantum numbers that label the direction of the angular momentum. A little arithmetic shows that if rotational quantum numbers from 0 through j_{max} are included, the total number of rotational states involved is $(j_{max} + 1)^2$. Since j_{max} is often larger than 30, one sees that N_r is then larger than 1,000, and that causes problems.

I pause to note that if, instead of this simple atom-diatomic molecule collision problem, the molecule were a polyatomic, or this were a diatom-diatom collision, the number of coupled rotational states would be much larger yet.

When an Atom Collides with a Molecule

When one considers the system composed of the atom colliding with the molecule, the center of mass motion of the whole system can be separated off; but one must include, in addition to the angular momentum j of the molecule, the angular momentum l of the atom about the molecule. If one is not careful, this adds yet more angular momentum states. However, it was recognized long ago that total angular momentum, $\mathbf{J} = \mathbf{j} + \mathbf{l}$, is conserved, and one can couple the angular momenta together to make states labeled by J. In doing so, one trades the m quantum numbers for l (partial wave) quantum numbers without increasing the number of coupled states or channels above $(j_{max} + 1)^2$. Further, states of even and odd inversion parity do not couple to each other, so that the number of coupled channels is about half this quantity. If the molecule is a heteronuclear diatomic, BC, no further reduction occurs; but if it is a homonuclear diatomic, B_2, the symmetry reduces the number of coupled channels in half again.

If one uses the vibrational and rotational states just described as an internal basis and expands the wave function in them, the Schrödinger equation governing the scattering, $(H - E)\Psi = 0$, becomes a set of coupled second-order differential equations in the distance, R, of the atom from the molecule. These are known as the close-coupled or coupled-channel (CC) equations. There are N coupled equations, and they must be solved for each J from 0 up to such a large value that the collision is so distant that nothing happens in it. That J_{max} is typically at least 100. Hence, to calculate a cross section, one must solve about 500 coupled differential equations 100 times, and that is the problem!

When, as a postdoctural researcher, I first started studying molecular-scattering theory in 1966, over 36 yr ago, other scientists working in the field were quite discouraged. Almost all the numerical calculations that had been done to that point had used the first Born approximation or the Born Distorted Wave (DW) approximation. These are approximations in which the intermolecular potential (PES) is treated as a perturbation; and in many branches of physics, where the kinetic energy is large and the potential is a small perturbation, the approximations

work very well. However, they do **not** work well for molecular collisions. When I express this to physicists, they often ask, "Why not just do second-order Born or DW calculations then?" The best answer I have found for this is to cite an example calculation that we did some years later in which we obtained an accurate scattering matrix, **S**, and expanded it in orders of the DW approximation. It required 30 orders before it converged! For this example, the exact CC result for the cross section for the $j = 0$ to the $j = 2$ transition is 58.9 $Å^2$. The DW result is over 10 times too large. For the $j = 0$ to $j = 4$ transition, the exact CC cross section is 16.4 $Å^2$; the DW cross section is identically zero.

In molecular scattering problems, the kinetic energy is quite small, and the potential energy is large and dominant. Hence, it is better to try making approximations to the kinetic energy. My student, Tom Tsien, and I did that, [2] and discovered that, for the intermolecular potentials then in vogue, if we neglected

(1) the rotational energy differences; and

(2) the differences in the centrifugal potentials due to the different partial waves, l, that are coupled together at a given J, the equations for rotationally inelastic scattering could be completely uncoupled by a simple unitary transformation.

We first called this approximation [2] the "strong coupling" approximation and then later [3] called it the "infinite order sudden" (IOS) approximation because it is a sudden approximation which, unlike the usual sudden approximations, treats the potential exactly; i.e., to infinite order. It is extremely simple; in fact, even this first version was an order of magnitude faster than the simple DW approximation.

A Simple Solution

At this point we suspected that something so simple should have been thought of before, so we looked in the literature and found that, sure enough, in 1963 Takayanagi [4] had noted that this approximation could be made but had passed it by thinking it would not work. Indeed, this was a problem I encountered for a few years with reviewers of my proposals; they felt that something so simple could not possibly work. But work it did. At this time, methods for the exact numerical solution of the CC equations were just being developed, and with considerable effort with the computers and methods of the day, Lester and Bernstein [5] had recently succeeded in accurately solving a problem in which there were $N_v = 1$ and $N = N_r = 16$ states coupled together. Their "parameters were chosen to simulate a physically realistic case, inaccessible to evaluation by approximate methods." However, the IOS approximation [3] reproduced their entire scattering matrix almost exactly. Subsequent tests that Tom Tsien, Greg Parker, and I did on problems with $N_v = 1$ and $N = N_r$ up to 100 gave equally encouraging results. [6]

The form of the IOS approximation which existed up to that point could only be applied if the angular and radial dependence of the intermolecular potential could be written in special forms which often do not hold for real molecular collision problems. Also, the formulation of molecular scattering theory that was being used in virtually all serious calculations was that of Arthurs and Dalgarno, [7] in which the coordinate axes are always kept pointed along space-fixed (SF) directions, and the coupling to get states labeled by total angular momentum is done by means of the Clebsch-Gordan theorem. In that formulation, all coupling appears in the potential energy terms in the Schrödinger equation.

However, having done my Ph.D. work at the University of Wisconsin, I was aware of the work of Chuck Curtiss and coworkers, [8–12] in which the representations of the rotation group are used to transform the Schrödinger equation to body-frame (BF) axes which rotate with the system. The BF z axis is usually chosen to point from the diatomic molecule toward the atom. The papers of Curtiss are difficult to read, and most people had ignored them. However, when I derived the differential form of the Schrödinger equation in BF Jacobi coordinates, I discovered that, in this formulation,

(1) the coupling of the rotational j states; and

(2) the centrifugal coupling occur in separate terms in the resulting CC equations, so that the two approximations of the IOS approximation could be made separately.

I call the neglect of the first coupling the "energy sudden" (ES) approximation, and the neglect of the second coupling the "centrifugal sudden" (CS) approximation.

A Standard Tool

From the equations, it was apparent that the CS approximation could be made for any arbitrary PES, regardless of its form, and that made it much more generally applicable. Furthermore, the potential is block diagonal in the BF formulation which made the IOS even faster than before; i.e., made it faster than CC calculations by a factor of 9,000 for the test example. I wrote up a clear comparison of the SF and BF formulations and these observations. The resulting paper, [13] which appeared in 1974, became a citation classic. As of this writing (February 2003), the paper has been cited 782 times, and it was honored by the Los Alamos library a few years ago as one of the most-cited articles ever to come out of Los Alamos. This approach has allowed accurate calculations of cross sections for many molecular collisions that would otherwise have remained intractable.

At this same time, McGuire and Kouri [14] were also working on this problem and made the same discovery completely independently and from a different point of view. What I called the CS approximation they called the "j_z conserving coupled states" approximation, and their

paper also became a citation classic. Later, we agreed to both call it the CS approximation and let CS stand for "centrifugal sudden" or "coupled states" at the preference of the user.

These papers, published at almost the same time, set off a flurry of activity that resulted in refinements to the approximation, clear interpretations of the parameters involved, simplified formulas for resulting cross sections, etc. These will not be detailed here. Suffice it to say that the CS approximation is valid and accurate for a wide range of molecular collision problems, and it has become one of the standard tools of the trade. Calculations using it continue to appear regularly, and it has enabled very many problems to be solved and understood which would have been too expensive to do by exact CC methods. The whole approach has come to be known as angular momentum decoupling theory. A number of people have contributed to it, and it has contributed greatly to the advance of the quantum theory of molecular collisions. [15]

More does need to be said about the IOS approximation. In 1975, Secrest [16] and Hunter [17] independently showed that the IOS approximation is equivalent to an approximation given in 1968 by Curtiss [11] in which the potential angle, the angle *between* the axis of the molecule and the vector pointing from the molecule to the atom, is held fixed during the collision. In the process, Secrest and Hunter also showed that the IOS approximation could thus be applied to an intermolecular potential with an arbitrary angular and radial dependence and could be used with an infinite basis of rotational states.

Working from there, Greg Parker and I gave a very simple and transparent derivation of the IOS approximation and showed that cross section formulas simplify markedly. [18] Many of the total integral, differential, and transport cross sections that are needed in analyzing experiments reduce to averages over the potential angle of cross sections calculated using simple structureless spherical particle formulas. That meant that experimentalists could, with trivial modifications of their existing computer programs, immediately begin analyzing their experiments in terms of nonspherical intermolecular potentials and see just what the effects of the nonsphericity are.

The IOS approximation then became one of the standard methods of molecular scattering theory, and it continues to be heavily used by both theorists and experimentalists. [15] It has been applied to atom-polyatomic molecule and molecule-molecule collisions, and they pose no problem for it. The more rotational states that are present and the stronger the coupling, the better the approximation works.

A More Difficult Problem

The preceding discussion has been limited to inelastic but nonreactive molecular collisions. When reactive (rearrangement) scattering of the type A + BC → AB + C is possible, the quantum scattering problem becomes even more difficult. In addition to all the coupled states

involved, the initial and final states can only be accurately described if one uses different coordinate systems. As a result, progress in reactive scattering calculations was slower. The first accurate CC calculations of reactive molecular scattering in the full three-dimensional physical space appeared [19,20] in 1975, and they were a major accomplishment. However, that was for the simplest reaction ($H + H_2$), and, despite the many very bright people working on the problem, it was a full 10 yr before calculations of comparable accuracy appeared for any more complicated system.

In an attempt to contribute to the solution of this problem, I began looking at coordinate systems that would go smoothly from those of the reactants to those of the products and allow a CS approximation that would be valid everywhere. I found a very promising system of this type by making a dynamic kinematic transformation to coordinates in which one of the vectors was as long as possible. That leads to internal coordinates that are essentially the same as the symmetrized hyperspherical coordinates of Smith [21] and Whitten [22] but in which the BF z axis lies in the triatomic plane and is along the longest vector possible. I called these "adiabatically adjusting, principal axis, hyperspherical" (APH) coordinates. [23] With a lot of help from Greg Parker, Bob Walker, and Bill Archer, the theory of reactive scattering in APH coordinates was published and working computer programs were written. [24–26] These codes were used and improved over the next several years, and some of the first-ever accurate, converged CC results were obtained for such complicated and diverse rearrangement reactions as $e^+ + H$, $Li + FH$, $F + H_2$, $H + O_2$, and $t\mu + D_2$. This work will not be detailed here, as references are available on request, but I do want to note that the coworkers who made the largest contributions and without whom the work simply could not have been accomplished were, in addition to those already named, Joel Kress, Zlatko Bacic, Antonio Lagana, Eric Butcher, and Ed Hayes.

In the course of this work, it became clear that the APH coordinate approach does provide a framework in which angular momentum coupling is minimized for reactions whose transition state is linear. However, for floppy systems that can go all the way from linear configurations to triangular configurations in which the two principal moments of inertia in the triatomic plane are equal, there are singularities in the APH Hamiltonian that are difficult to treat in numerical work. Indeed, I have shown [27] that, for such floppy systems, no BF exists which is optimum or convenient everywhere.

Recent and Current Work

In more recent years, Brian Kendrick, Bob Walker, Ed Hayes, and I have shown how to treat conical intersections of electronic potential energy surfaces in reactive scattering to properly include the "Berry" or geometric phases they introduce [28] and how to handle the Eckart singularities that occur with floppy systems. [29] Using massively parallel computers, Brian Kendrick has gone on to successfully carry

out converged exact reactive scattering calculations [30–32] in systems such as H + D_2 at such high energies that diagonalization of matrices of dimensions N' as large as 41,248 to get the internal basis functions and propagation of N = 1,590 CC equations has been required. Since computational effort scales more like N^3 than N, it is clear that the field has come a long way from the time when N = 16 was a significant achievement.

Together with Flavio Colavecchia and others, I am currently working on yet more complicated collision-induced dissociation (CID) and recombination reactions of the form A + BC ⇔ A + B + C, in which the system can break up into three free bodies, but that is a story which is still in progress and will be saved until later. [33,34]

All of the coworkers listed in this article, except Tom Tsien, have spent time in T-Division here at Los Alamos. While many have gone on to other lives elsewhere, several are still here and doing very fine work. I owe them all a great debt of gratitude. Without them the work described herein would never have happened.

Semiclassical Dynamics—Insight into Physics of Complex Systems

by Dmitri Babikov and Joel D. Kress

Babikov
Hamilton
Heller
Herman
Kluk
Kress
Pack
Schrödinger
Walker

Even in this age of fast, multiprocessor computers, the calculation of the exact quantum dynamics of three heavy atoms is state of the art, especially when recombination (dissociation) or chemical reactions are involved. (See "Body-Frame Axes and Angular Momentum Decoupling in Molecular Collisions," by Russell T Pack, immediately preceding this article.) When a collision process involves four heavy atoms, an exact quantum description is impossible. Classical mechanics, on the other hand, is affordable for large systems, but has many limitations. Thus, it is always important to develop an approximate method with two important features: computational affordability and physical correctness. Theoretical Division scientists have done just that.

A standard quasi-classical (WKB) approach was derived at the beginning of quantum mechanics. WKB adopts an expansion in powers of $\hbar$ and demonstrates the classical limit of quantum mechanics as $\hbar \to 0$. ($\hbar$ is a standard symbol called Plank's constant.) Practical implementation of this method was, nevertheless, impossible due to the problem of singularity at turning points.

A new chapter started in the mid-1970s with Eric Heller at Los Alamos. Instead of making an expansion in $\hbar$, he considered the time-dependent Schrödinger equation, made an expansion of the potential near the center of the Gaussian wave packet, and derived a completely new semiclassical approximation for the "quantum dynamics of localized wave packets on smooth potential energy surfaces." [1] The evolution of a semiclassical Gaussian wave packet is defined by four time-dependent parameters:

(1) the position,

(2) momentum of the center of the wave packet,

(3) the complex width of wave packet, and

(4) the phase.

The equation of motion for the position and the momentum obey Hamilton's equations (classical mechanics), and the equation of motion, for the phase is reminiscent of the action integral. Thus, the propagation scheme is almost as simple as classical mechanics but accounts for several quantum features.

First, the semiclassical Gaussian wave packets remove the problem of singularities at the turning points and provide an analytic continuation into the classically forbidden region. Second, the quantum zero-point energy (ZPE) appears in the formalism very naturally. Absence of ZPE in classical mechanics often leads to unphysical behavior. For example, a molecule like ArHF is bound according to quantum mechanics and spectroscopic measurements, but it is not stable according to classical mechanics: a large amount of energy from the strong HF bond flows into the weak van der Waals bond and releases the argon atom. This problem doesn't occur in the semiclassical method of Heller. Here, the total energy of a molecule is a sum of ZPE (quantum part) and the energy of the trajectory (classical part). The trajectory is not allowed to consume all the energy of the system, but only a part which is above the ZPE, so the system is always guaranteed to have at least the quantum ZPE. Thus, the existence of bound states of molecules like ArHF is allowed. Other demonstrated examples of this kind were stable $(HBr)_2$ and neon-12 clusters.

Another Important Feature

The next important feature is quantum interference. Classical trajectories do not interfere. Though the semiclassical Gaussian wave packets also evolve along independent trajectories, they accumulate phases and contribute coherently into overall wave function, giving rise to quantum interference. This property is vital, for example, for description of elastic and inelastic scattering processes, where such phenomena as rainbows and diffraction do occur. For example, we applied this method to calculate cross sections for state-to-state transitions in atom-molecule scattering. We observe a correct quantum behavior of opacity

functions for all possible inelastic- and elastic-scattering channels. The last is remarkable, because the classical scattering theory is known to fail completely in its description of elastic scattering due to the lack of phase information. Other examples include $H + H_2$ and $He + H_2$ scattering and gas-surface scattering (helium on LiF).

Several other applications rely on the phase of semiclassical Gaussian wave packets. The dynamics of photofragmentation of polyatomic molecules is especially accurate here because the fragmentation of a molecule excited to the repulsive state is fast, and error due to the semiclassical approximation does not accumulate. Examples include H_3^+, CO_2 and CH_3I molecules, and Ar_nI^- ($2 < n < 6$) clusters. A modern example is coherent control of iodide/carbon/nitrogrem (ICN) photodissociation. In this process, the two dissociation pathways interfere due to the phase difference and provide a possibility for the control of the reaction outcome. The semiclassical method adds the phases of each pathway coherently, thus providing a good tool for theoretical study of coherent control.

Another application is the calculation of the bound states of polyatomic molecules. Figure 1 shows how a semiclassical Gaussian wave packet (moving in the classical phase space) builds the wave function of a quantum eigenstate. The phase of the wave packet produces an interference pattern and reproduces the nodes of the wave function. The penetration of the resultant wave function into the classically forbidden regions (outside of the phase-space boundary) is also reproduced.

A new implementation of semiclassical theory, recently proposed in T-Division, shows that the dynamics of semiclassical Gaussian wave packets preserves the quantum symmetry of the system. Symmetric molecules occur only in even rotational states, while nonsymmetric molecules may be in both even and odd states. We build the states as a superposition of Gaussian wave packets, study how the symmetry affects the outcome of a collision, and obtain the correct quantum behavior of state-to-state transition probabilities (even-odd, even-even, and odd-odd). This may help us to understand the anomalous isotope effects in recombination reactions (such as ozone formation), presumably driven by quantum symmetry.

The field of semiclassical dynamics, started almost 30 yr ago by Eric Heller at Los Alamos, is very active; and people all over the world are exploring the merits of this method. Various simplifications ("frozen" Gaussians) and refinements (Herman-Kluk propagator, cellular dynamics, and the initial value representation) have been developed. Still, the semiclassical picture remains very attractive to many chemists and physicists because it provides a simple intuitive picture related to classical mechanics; and, at the same time, it captures the major quantum mechanical effects for systems in which exact quantum treatments are impossible.

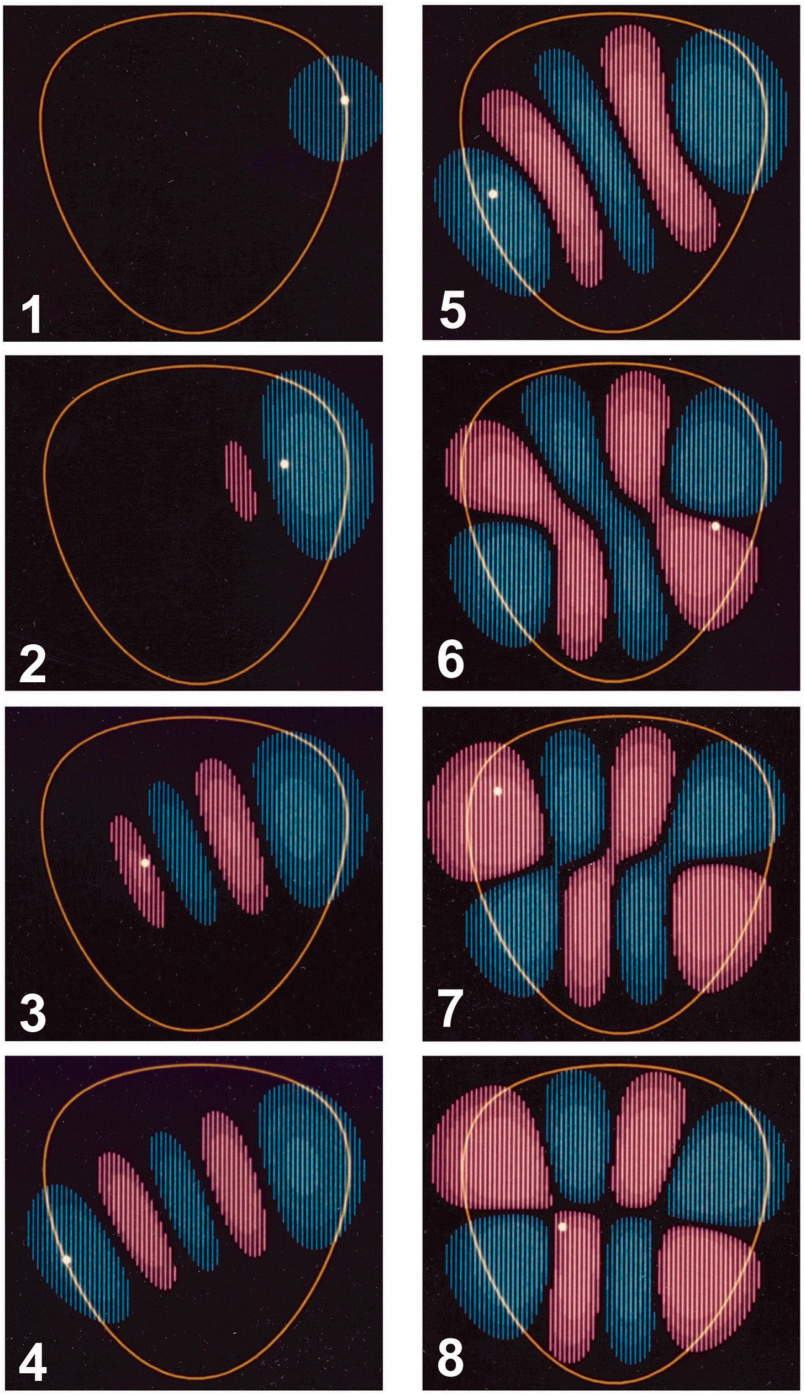


Fig. 1. A semiclassical Gaussian wave packet is used to calculate a two-dimensional wave function of a quantum eigenstate (three quanta in the horizontal direction and one in the vertical). Evolution in time is shown in frames 1 through 8. Also shown are the center of the wave packet (white dot) and the classical phase space (yellow contour). The wave packet starts from the turning point in the left upper corner (frame 1). Integration over time produces an interference pattern, revealing the nodes of the quantum eigenstate. (Figure courtesy of R. B. Walker)

Computers

by Francis H. Harlow

Los Alamos National Laboratory has played a major role in the development of high-speed computers throughout the history of their evolution. The electronic numerical integrator and calculator (ENIAC) in the middle 1940s was not a Los Alamos invention but saw important usage by Laboratory personnel. Until around 1952, however, the principal reliance in Theoretical-Division for calculations was on desk calculations, slide rules, and accounting machines. But in that year the computer age was launched, and Los Alamos, especially T-Division, was at the forefront.

Introduction

The early history of computers at the laboratory was principally written by T-Division. It consisted of two branches: computer hardware and computer usage. The first of these, as described in this chapter, consisted of building the mathematical analyzer, numerical integrator, and computer (MANIAC) and working with International Business Machines (IBM) in the development of a remarkable sequence of machines designed especially for scientific programming. The second branch of Laboratory computing lay in the use of the machines, as described in virtually every other chapter in this book. It is not an exaggeration to say that computing at Los Alamos is as crucial to the Laboratory's success as are the awesome experimental facilities that have evolved here over the full 60 years covered by this book.

The hardware branch of computing actually means much more than wires and tubes and chips. Equally necessary is the software that establishes the bridge between the machines and the solution of problems.

The main parts of this chapter were written many years ago, and we have extracted from them much of the following text. In particular, the report "Computing at LASL in the 1940s and 1950s" (LA-6945-H), written in 1978, does an excellent job in describing this exciting period. It was written by four of the principal contributors to launching the age of scientific computing: Roger B. Lazarus, Edward A. Voorhees, Mark B. Wells, and W. Jack Worlton. [1] Another very significant publication on computers appeared in *Los Alamos Science*, Number 22 (1994). The article "Windows on Computing" was written by David W. Forslund, Charles A. Slocomb, and Ira A. Agins. [2] We quote almost verbatim extensive passages from those reports. Their text refers to the Laboratory as LASL, the Los Alamos Scientific Laboratory, as it was known at the time.

Computers in 1943

In 1943, computer technology was in an extremely primitive state compared to the rapidly growing needs of nuclear research and development. The analog computers in use included the slide rule and the differential analyzer. The slide rule was the ubiquitous personal calculating device, and LASL made 18-in. slide rules available from stock. Although slide rules have now been largely replaced by electronic calculators, they once played an important role in computational physics—one which is still important and often overlooked. J. Carson Mark, head of the Los Alamos Theoretical Division at the time, pointed out that computers can be downright dangerous if their results are not checked with preliminary estimation techniques. In other words, estimation should precede computation. Using the results of computations without at least a rough estimate of what the answer should be will inevitably lead to technical errors, because flaws in the models, the codes, or the data are essentially impossible to eliminate in the very complex models used at the frontiers of scientific research and development. In those early years when computers were less trusted than they are now, "Carson's Caution" was well understood, but we now accept computers so readily that we sometimes forget this very basic lesson from the past. [2]

The other analog computing device used in the early 1940s was the mechanical "differential analyzer." This was a one-of-a-kind device that was not readily available, and thus it was not used at LASL. [2]

Digital computing in those early days was done either with electromechanical desk calculators or with accounting machines. Both of these methods were used in the early weapons calculations. [2]

Chronological Overview

Figure 1 shows the various categories of computing devices that have been used by LASL from the 1940s to the early 1970s. Note that not all of these have been used *at* LASL; in the late 1940s and early 1950s, some of the unique early computers at other sites were used in an attempt to complete some of the more critical calculations. In a sense, the "modern" era of LASL computing began in 1953 with the installation of the first IBM 701; this installation ushered in a period in which the major computing requirements at LASL (and other large computing sites)

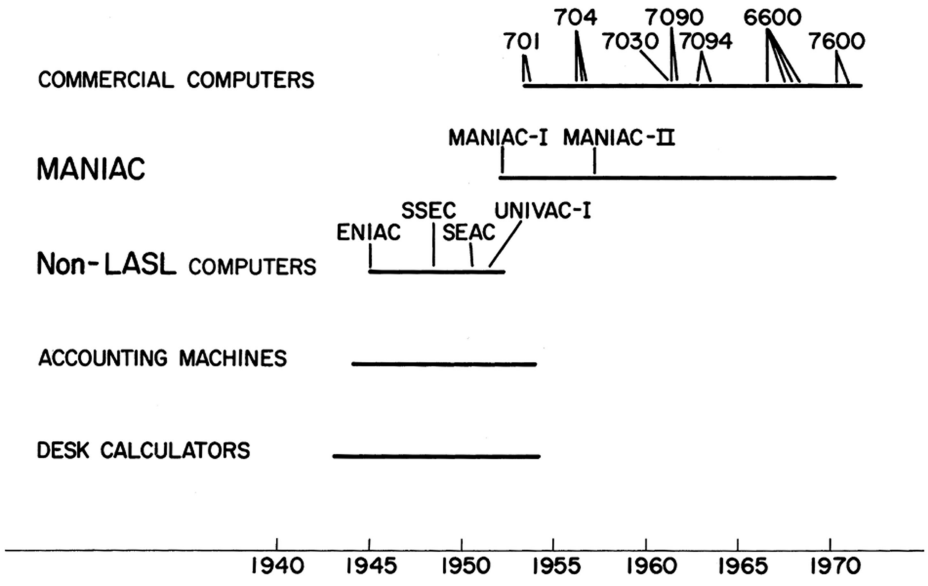


Fig. 1. This time line shows the use of various kinds of computing technology at Los Alamos, beginning with analog computers—slide rules—in the early 1940s.

would be met with equipment developed by commercial vendors rather than with computers developed as one-of-a-kind devices by government laboratories and universities. [2]

History of Computers at Los Alamos: 1943–1968 [2]

1943–45

Desktop calculators and punched-card accounting machines are used as calculating tools in the Manhattan Project.

1945

ENIAC, the world's first large-scale electronic computer, is completed at the University of Pennsylvania. Its "shakedown" calculation is the "Los Alamos problem," a calculation needed for the design of thermonuclear weapons.

1949

IBM's first Card Programmable Calculations are installed at the Laboratory.

1952

MANIAC is built at the Laboratory under the direction of Nick Metropolis. It is the first computer designed from the start according to John von Neumann's stored-program ideas.

1953

The Laboratory gets serial number 2 of the IBM 701. This "Defense Calculator" is approximately equal in power to the MANIAC.

1955

The MANIAC II project, a computer featuring floating-point arithmetic, is started. The Laboratory begins working closely with computer manufacturers to ensure that its future computing needs will be satisfied.

1956

MANIAC II is completed. The Laboratory installs serial number 1 of the IBM 704, which has about the same power as MANIAC II. From this point on, the Laboratory acquires supercomputers from industry.

Late 1950s

The Laboratory and IBM enter into a joint project to build STRETCH, a computer based on transistors rather than vacuum tubes, to meet the needs of the nuclear-weapons program.

1961

STRETCH is completed and is about 35 times as powerful as the IBM 704. IBM used much of the technology developed for STRETCH in its computers for years afterward.

1966

The first on-line mass-storage system with a capacity of over 1,012 bits, the IBM 1360 Photo Store, is installed at the Laboratory. Control Data Corporation introduces the first "pipelined" computer, the CDC 6600, designed by Seymore Cray. The Laboratory buys a few.

Desk Calculators

Shortly after scientists began arriving at Los Alamos in March 1943, a desk-calculator group was formed under the direction of Donald Flanders. This group consisted of both civilian and military personnel, including Women's Army Corp personnel (WACs), Special Engineering Detachment personnel (SEDs), and the wives of scientists. By 1944 this group was the largest group in the Theoretical Division, with 25 people. The calculators used were Marchants, Monroes, and Fridens, although Flanders soon decided that it would be best to have a standard calculator and selected the Marchant. (However, two advocates of Monroes refused to give them up.) Repair of the calculators was a continual problem at the isolated Los Alamos site, so many of the problems with a sticking calculator were solved simply by dropping the end of the offending device, in the military tradition of the "drop desk." Jo Powers also notes that when the problems became acute, they called Dick Feynman (later a Nobel laureate) who, according to Jo, could fix anything. [2]

To avoid problems with manual errors, many of the calculations were executed by more than one person, with intermediate check points to ensure that no errors had been introduced. Flanders designed special forms to aid in the setup and execution of the calculations. These calculations were typically done by a manual form of parallel processing; that is, the problem would be broken down into sections that could be executed independently. It seems that parallel processing is part of the "roots" of scientific computing, rather than just a recent innovation, as sometimes thought. [2]

Punched-Card Accounting Machines

The Punched-Card Accounting Machines (PCAMs) of the early 1940s were designed primarily for business applications, but they could also be used for scientific calculations. In early 1944, Stan Frankel (who worked with Metropolis) recognized that PCAM equipment could be used for some of the calculations at LASL, and that spring the following equipment was delivered:

- three IBM 601 multipliers,

- one IBM 405 alphabetic accounting machine,
- one IBM 031 alphabetic duplicating punch,
- one IBM 513 reproducing punch,
- one IBM 075 sorter, and
- one IBM 077 collator. [2]

The 601 multiplier was the "workhorse" of this array of equipment. Its basic function was to read two numbers from a card, multiply them together, and punch the result on the same card, although it could also add and subtract (division was done by multiplying the reciprocals). The 601 was an important advance over its predecessor, the IBM 600, because the 601 had a changeable plug board that made changing the functions being performed very rapid compared to rewiring the 600 for every such change. The 405 could add or subtract and list results. The 031 was a "keypunch" in modern terminology. [2]

Early accounts of computations with these machines indicated that a single problem took about 3 months to complete; later methods reduced the time so that nine problems could be completed in a 3-month period. [2]

The first of a series of the new IBM 602 Calculating Punch models was delivered to P. Hammer on November 6, 1947, along with a new model of the sorter, the IBM 080. Although still an electromechanical device, the 602 had over 100 decimal digits of internal storage and could perform the following operations: add, subtract, multiply, divide, compare, and negative test. Property records indicate that eight 602s were delivered to Los Alamos in 1947 and 1948, and both 601s and 602s were in use for some time. [2]

IBM's first electronic calculating punch was the 604, which was also used at LASL. It had 50 decimal digits of internal vacuum-tube register storage, although the sequencing control was through a prewired plug board. Input and output was through punched cards, read and punched at a rate of 100 cards per minute. The 604 could perform the same operations as the 602, plus zero test, positive or negative test, repeat, and shift. [2]

The IBM card programmable calculators (CPCs) were delivered to the Laboratory in 1949; eventually LASL had six of these, the last of which was removed in October 1956. The CPC employed both electronic and electromechanical technology, with 1,400 vacuum tubes and 2,000 relays. It had 290 decimal digits of internal vacuum-tube registers plus up to three IBM 941 storage units that provided sixteen 10-digit words of relay storage each. The card reader and printer operated at 150 cards per minute and 150 lines per minute, respectively. The operations performed by the CPC included add, subtract, multiply, divide, repeat, zero test, suppress, shift, plus there were wired subroutines for transcendental functions. [2]

The main functions of those accounting machines and the very early electronic computers was to simulate through numerical computation the extreme physical conditions and complex nonlinear processes that occur within a nuclear weapon. The continued role of computer simulation as the primary tool for designing and predicting the performance and safety of nuclear weapons has been a major driver behind the development of increasingly powerful computers. The larger the computer, the greater the complexity that could be simulated accurately. [1]

Fig. 2. Some of the batteries used to power the MANIAC I.

The MANIAC

A major advance in computer power at LASL was introduced in 1952, with the construction of an electronic computer called MANIAC.

Actually, there were three MANIAC computers. MANIAC was at the Laboratory from 1952 to 1957. MANIAC II was there from 1957 to 1977. MANIAC III was never at LASL; it was built in the late 1950s at the University of Chicago where Nick Metropolis, the prime instigator for all three machines, spent a few years. The word MANIAC is an acronym for mathematical analyzer, numerical integrator, and computer. [2]

MANIAC I was a vacuum tube machine powered by a room full of batteries (Fig. 2). Figure 3 shows MANIAC I. The arithmetic unit with the three registers is in the middle with the operation controls on the sides and in the back. The word length was 40 bits; if you look closely, you can see 10 bits of register in each of the 4 central panels. The memory is on top. There are 2 Williams tubes in each of the 20 boxes. The 2 monitor tubes at the ends were for viewing the contents of any of the 40 tubes. Each tube could store 32 by 32, or 1,024 bits of information; hence, MANIAC I had a memory capacity of 1,024 words. On the far left is the row of switches that served as the user's console as well as part of the engineer's console. The user's console was later moved to a table (Fig. 4). The controls used by the engineers to tune the memory and view its contents were accessible on the front of the memory boxes and below the memory. [2]

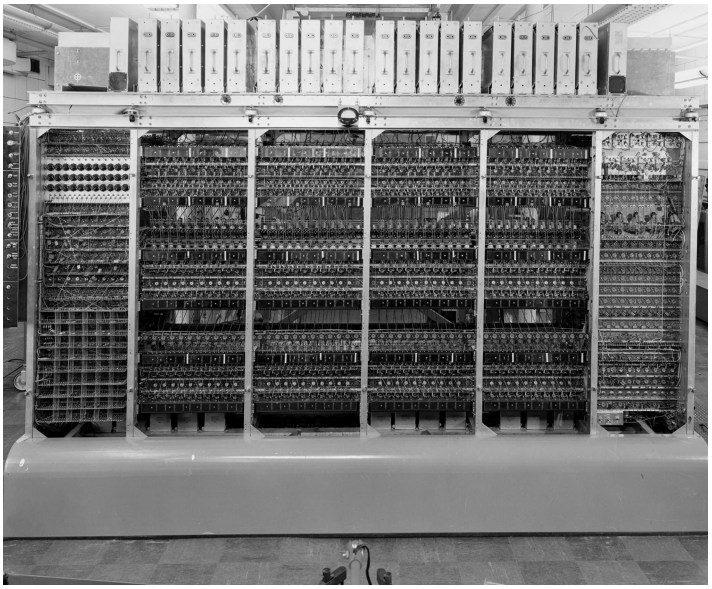


Fig. 3. MANIAC I. Controls for operating the machine are on either end; the four central panels are the arithmetic unit.

There was good rapport between the engineers and the programmers, or coders as they were called in those days. Nick and later Jack Jackson, now with IBM, were the primary design architects of

MANIAC I and its system, but suggestions for hardware modifications as well as operational procedures were proposed arbitrarily by users or engineers. There are many examples of this user-engineer interaction throughout the service of MANIAC I and MANIAC II.

Fig. 4. Two operators and the MANIAC I.

Not all of the computing on MANIAC I was numerical. We had some fun with combinatorial problems. We also had a chess-playing program on MANIAC I. However, because of the slow speed of MANIAC (about 10,000 instructions per second) we had to restrict play to a 6 by 6 board, removing the bishops and their pawns. Even then, moves averaged about 10 min for a two-move look-ahead strategy. The program played several games, both against itself and against humans; it even won one game against a beginner who had been taught how to play specifically for our experiments. We wanted to determine the level of play of the program. As I remember, we concluded that the program was equivalent to a beginner with about a half-dozen games of experience. [2]

There were, of course, numerous serious problems solved on MANIAC I. Foremost among these were the nuclear-weapons yield calculations. Another was something called Quintet, which was a five-dimensional integral for scattering corrections in experimental physics.

A more important problem was the Fermi-Pasta-Ulam nonlinear oscillator investigation that led to what is now still quite a live field, called soliton theory. There was the first Monte Carlo equation of state, where you tried to calculate the actual equation of state of the material by stochastic processes on sample molecules. There was the first stellar evolution calculation done by Art Carson and George Gamow. [2]

There was a code worked on by Verna Gardiner with Gamow for trying to discover the code for deoxyribonucleic acid (DNA) selection of amino acids, which was not successful. Other things were virial coefficients, nuclear-scattering problems, and Schrödinger equation integrations. Intranuclear cascades were another Monte Carlo thing; in a heavy nucleus, one actually tried to count the production of cascades of pions within the nucleus. [2]

There was a calculation of group characters on MANIAC I, done by Paul Stein and others. There was perhaps the beginning of what a lot of people now think of as experimental mathematics, where you try things and explore using the computer, trying to form conjectures or get some new insight. [2]

It is interesting to note that the computing power of MANIAC I, which required a room full of hardware, can now be contained in the palm of one hand.

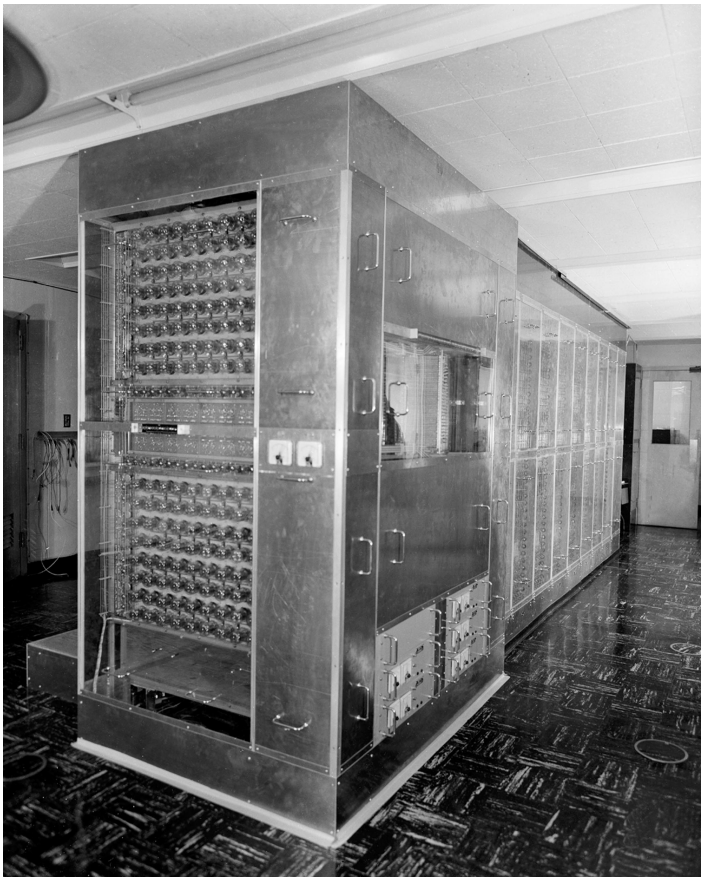


Fig. 5. MANIAC II.

MANIAC I did not actually leave service until 1965, but it was replaced at LASL in 1957 by the faster, more powerful, easier-to-use MANIAC II (Fig. 5). [2]

The chief advantage of the second-generation machines, MANIAC II and IBM 701, was floating-point arithmetic. A substantial portion of the software effort on the early machines involved producing subroutines to do our arithmetic in floating point, letting the user think more at his own level without complicated scaling. [2]

With the advent of the more powerful second-generation machines, attention could be given at a higher level, and we then saw the beginning of real programming languages: FORTRAN on the 704 (1957) and Madcap on MANIAC II (1958). However, whereas Madcap has evolved and improved over the years along with MANIAC II and with our understanding of languages and algorithms, FORTRAN has been essentially static. [2]

MANIAC I (1952) had been among the first general-purpose digital computers to realize von Neumann's concept of a stored-program computer—one that could go from step to step in a computation by using a set of instructions that was stored electronically in its own memory in the same way that data are stored. (In contrast, the ENIAC [1945], the very first general-purpose electronic computer, had to be programmed mechanically by inserting cables into a plug board.) Maniac II (1956) embodied another major advancement in computer design, the ability to perform floating-point arithmetic—the kind that automatically keeps track of the position of the decimal point. The peak speed of MANIAC II was 10,000 arithmetic operations per second, an impressive factor of 1,000 higher than that of the electromechanical accounting machines used to perform numerical calculations at Los Alamos during the Manhattan Project. [1]

Computers from External Vendors

After 1956 and MANIAC II, the Laboratory stopped building computers and instead relied on close working relationships with IBM and other vendors to ensure that the industry would be able to supply the necessary computing power. [1]

The IBM 701 originally was announced as the "Defense Calculator." LASL was the recipient of the serial number 1 machine, which arrived in April 1953 and remained until the fall of 1956. A second 701 calculator, as it was later called, came in February 1954. The machine was fixed binary. All numbers were assumed to be less than one; that is, the binary point was on the extreme left of the number. The electrostatic

storage (which was not too reliable) could hold 2,048 36-bit full words or 4,096 16-bit half-words. It was possible later to double the size of that memory, which LASL did indeed do. Instructions were all half-words; data could also be in half-word format. Instructions were a single address with no indexing. There was no parallel input/output (I/O). The system consisted of one card reader, one printer, one memory drum storage unit, two tape drives (which often did not work), and one card punch (the primary device for machine-readable output). [2]

The memory register held word transfers to and from memory. If you wanted to add or subtract, you would put one number into the accumulator; if you wanted to divide or multiply, you would use the multiplier-quotient (MQ) register. There were two overflow bits in the accumulator register in case of a spill, a condition the program could sense with an instruction. A 72-bit product was formed when multiplying two full words. A 72-bit dividend, in the accumulator and MQ, divided by a 36-bit divisor yielded a 36-bit quotient in the MQ and a 36-bit remainder in the accumulator. Six sense switches in the upper right-hand corner of the console could be interrogated by the program so as to alter the course of the program and thereby provide six manually set conditional branches. There were also four programmable sense lights, which could be used to visually indicate where in the code the execution was occurring or could monitor some other internal condition. In the extreme lower left-hand corner, two buttons, one labeled "half step" and the other "multiple step," permitted the programmer to step his way slowly through his program and observe the contents of the accumulator, MQ, and the memory register when the 701 was switched to "Manual Mode." [2]

Before the first IBM 701 had been delivered, three principal methods of programming had been developed at the LASL: longhand (which is perhaps better understood to today if called "machine language"), SHACO, and Dual. SHACO and Dual were originated and implemented at LASL. All these programming systems were developed by 8 to 10 people and were operational soon after the delivery of the first machine. [2]

Machine language came into operation in 1953 and was based on an early IBM assembly program called SO2. LASL's first version, "606," was soon followed by "607," which was then used during the remaining service of the 701s at LASL. [2]

Machine language allowed the user to get intimately close to the computer. There were no monitors or other software stored in the computer. Everything that was in the computer was there because the user loaded it from cards. When he got on the machine, he loaded a bootstrap card loader that loaded the rest of his deck. He loaded his own print program (fixed-output format) and manually put the corresponding print board in the printer. If there was trouble on a run, he then loaded an appropriate debugging program and associated printer board of his choice (generally, either a printed dump or an instruction-by-instruction

tracing of a portion of the code as it executed). Memory check-sums and frequent dumps were made as protection against the short mean time (minutes) between computer failures. [2]

With machine language, the user had to remember the following.

- When dividing, the divisor always had to be less than the dividend so that the quotient would be less than 1. If not, a Divide Check would occur, and the machine would come to a screeching halt.
- An instruction was often used also as a data constant; that would be unheard of today.
- Because the 701 was a fixed-point binary machine, the user had to think in binary octal, especially when working at the console or poring over an octal dump.
- Scaling was necessary and often difficult to do without causing undue loss of significance or result overflows. The programmer had to mentally keep track of the binary point as he programmed.
- When decimal data were loaded, the programmer specified both the decimal and binary points for the conversion. For example, the integer 13 (decimal) would have a decimal scaling factor of 2 (for two decimal digits), but the binary scaling factor would have to be at least 4 to accommodate the number of binary bits from the conversion.
- When adding two numbers, users had to have the same binary scale factor for proper alignment; otherwise, the user would have to shift one of them until they were aligned. He also had to allow room for a possible carry or to check the overflow indicator after the addition. [2]

The Program Library included various card loaders, print programs, and debugging programs. The debugging programs would look for such things as transfers or stores to certain specified locations. Memory errors frequently resulted in a “divide check” and a machine stop. Occasionally, instruction sequence control would be lost, and a jump would occur to some part of memory where it should not be. In this case, when the machine stopped, you had no idea how control got to that location. [2]

There were no machine room operators as such; there was a dispatcher who kept records of usage, downtime, and the schedule of assigned users. A user operated the computer while his programs were running. During the day, he would request and was allocated 2- to 5-min periods on the computer for debugging, checkout, etc. When his turn came, he would have his program card deck plus debugging programs in hand.

The preceding user would normally be printing or punching at the end of his run, and he would usually allow the next user to stack his cards in the card reader. When the preceding user finished, the next user would be ready to initiate loading of his deck. Sometimes the preceding user would overrun his allocation. Usually, he would be allowed to continue for another minute, but then the situation could get tense indeed. More than once, the waiting user pushed the clear button to erase memory. The habitual overrunners developed a reputation for such; but, all in all, things went rather smoothly. [2]

At night, users with long-running jobs were able to do something other than tend the computer. An AM radio was attached to the machine. Different sections of a code had a unique rhythm and sound so the user could do other work if he kept one ear half-cocked to the sound of the radio. If he heard something unusual, he took appropriate action. Night runs would range from 30 minutes to 4 or more hours. Occasionally, when computer time was oversubscribed, upper management might have to decide who would get the time available during the night. [2]

The IBM 704, initially announced as the "701A" in mid-1954, came to LASL in January 1956. It represented a significant improvement over the IBM 701. For one thing, instead of electrostatic storage, it had core memory, which was much larger (up to 32 k words) and far more reliable. It had floating-point binary, so SHACO and Dual were unnecessary. It had index registers; although there were only three, they could be used simultaneously in combination if you were a very tricky coder. It had a much larger vocabulary and some logical operations. It was two-and-a-half times faster. By September 1956, three 704s had been installed at LASL and two 701s were released. There was a lot less procurement red tape at that time. [2]

A particularly important collaborative effort was STRETCH, a project with IBM started in 1956 to design and build the fastest computer possible with existing technology. For the first time, the new semiconductor transistor would be used in computer design. Transistors have a much faster response time and are much more reliable than the traditional vacuum-tube elements of digital circuits. The STRETCH, or IBM 7030, computer was made from 150,000 transistors and was delivered to Los Alamos in 1961. It was about thirty-five times faster than the IBM 704, a commercial vacuum-tube machine similar to the MANIAC II. [1]

In 1966 the Laboratory initiated a collaboration with Control Data Corporation and acquired a few CDC 6600s.

In 1968, a new Laboratory division (Computing Division) was formed from T-Division to carry on the activities necessary to keep Los Alamos at the forefront of the rapidly evolving computer age. Thereafter, T-Division continued its involvement principally through its extensive usage of each new machine.

e-Print arXiv Project

Bhattacharya
J. Carlson
Doyle
Fendley
Ginsparg
Gough
Hartill
Jen
Schwander
Warner
West

by Paul H. Ginsparg

The e-Print archives (http://arXiv.org/), initiated within Theoretical Division at Los Alamos National Laboratory in August 1991, have effectively transformed the research communication infrastructure of multiple fields of physics and are moving towards providing a unified set of global resources for physics, mathematics, and computer sciences. Entirely automated and scientist-driven, the archives have also spawned a number of independent imitators (including the major National Institutes of Health PubMedCentral project launched in 1999). They have grown into a major international project, with dedicated mirror sites in 16 countries and collaborations with U.S. and foreign professional societies and other international organizations. They have also provided a crucial lifeline for isolated researchers in developing countries. They have served as the core around which the "Open Archives Initiative" was organized starting in late 1999, which is an effort to provide an interoperable framework for networking a wider variety of institutional electronic resources. Ideally, these archives serve as a model for the rest of scientific research communication, providing much greater efficiency and functionality at far lower cost than any pre-existing system, and suggesting how a global unified "knowledge network" might be the best way to communicate knowledge and hence to create new knowledge.

One simple measure of the enormous impact of the Los Alamos e-Print arXiv project can be found in the IEEE/AIP Computing in Engineering and Physics Year 2000 calendar and poster of "top 10 algorithms of the century," whose timeline for the entire 20th Century has an entry for 1991 that reads: "Paul Ginsparg puts e-Print archive online at Los Alamos National Laboratory."

Background

In August 1991, a year after joining Theoretical Division at Los Alamos, I wrote the first version of the e-Print archive software to assess the feasibility of an automated system for preprint distribution in theoretical high-energy physics. The working conditions and intellectual freedom in T-Division during the early 1990s were essential components in the implementation and development of the system. At the time, T-Division was one of the few high-powered research environments that afforded the flexibility to undertake such an open-ended project and also benefited from the attitude of the Laboratory's computer network support that such projects were a natural use of the Laboratory's growing networking resources. This was also a transitional period for T-Division at Los Alamos, shortly following the

fall of the Berlin Wall and presumed end of the Cold War, when diversification from a dedicated weapons agenda had to be contemplated and before the security implications of the Internet were fully appreciated.

The initial user base for the system (called "hep-th" for "high-energy physics theory") was assembled from pre-existing e-mail distribution lists in a subdiscipline of the field, namely, two-dimensional gravity and conformal field theory. Recall that 1991 was a few years before the global establishment and acceptance of the World Wide Web, and even the notion of the Internet back then was familiar only to very few outside of the academic research community. Starting from a subscriber list of 160 addresses in mid-August 1991, hep-th quickly grew within 6 months to encompass the full range of formal quantum theory and string theory. After a few years, the original hep-th grew to over 4,000 subscribers, received over 200 new preprint submissions per month, responded to thousands of requests per day, and transmitted more than 1,000 copies of papers on peak days.

In the latter half of the 1990s, the server network based at Los Alamos experienced continued dramatic growth in both its breadth of coverage and worldwide usage. The physics community rapidly moved to realize the vision for the future as expressed in, "Report of the American Physical Society (APS) Task Force on Electronic Information Systems" [1]: "The dominant mode (of dissemination) will be via a single electronic physics library, or Physics Database, which will be the heart of a worldwide Physics Information System."

After ten years of operation at Los Alamos, the entirety of the archives collectively grew to serve over 70,000 users in over 100 countries, received over 33,000 new submissions in calendar year 2001, and processed many millions of electronic transactions of all sorts per month. (See usage graphs Fig. 1 and Fig. 2.) The system remained entirely automated (including submission process and indexing of titles/authors/abstracts) and allowed access by means of e-mail, anonymous file transport protocol (ftp), and the World Wide Web. In early 1999, the system adopted a new domain name, "arXiv.org."

Initially conceived as a feasibility assessment experiment for a small research subcommunity, the e-print archives are now more properly viewed as a prototype model demonstrating the potential of electronic communication to improve and transform scientific information exchange. Some measure of the success of the e-Print archives is given by the following:

- widespread testaments from users that it is an indispensable research tool—for many eliminating their reliance on conventional journals;
- decisions by virtually all physics institutions to discontinue their preprint mailings by the mid 1990s;

- the present custom of providing as reference an article's e-print arXiv identifier rather than a local report number or a published reference; and
- commonplace arXiv usage to the extent that science articles in, for example, the *New York Times*, give direct links to the version that the authors have placed on the archive before publication.

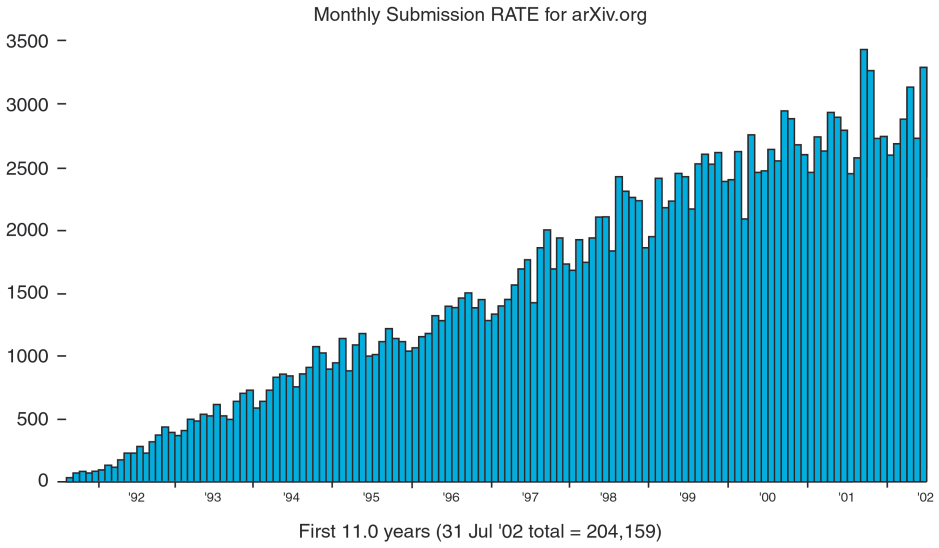


Fig. 1. The number of archive submissions per month, from August 1991 through July 2002. (See http://arXiv.org/show_monthly_submissions).

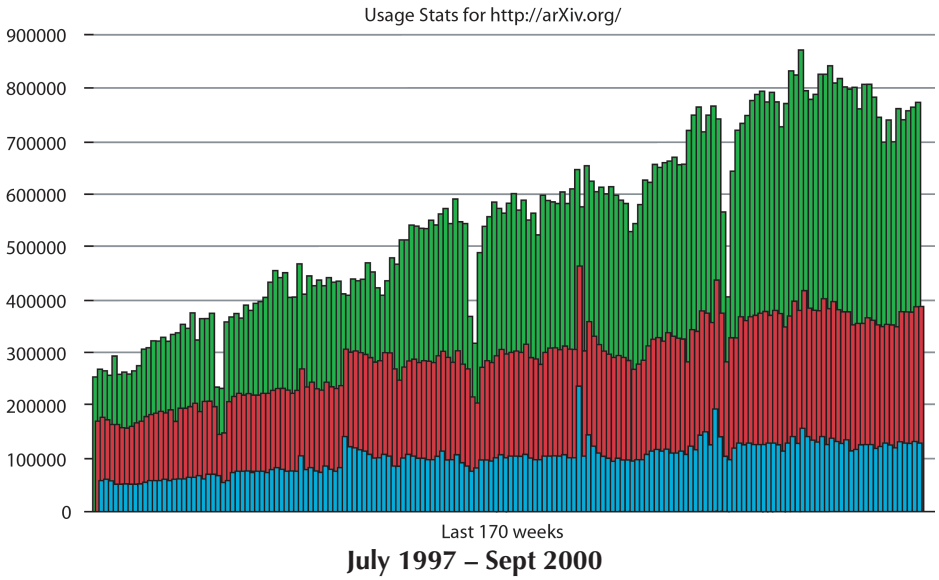


Fig. 2. The number of World Wide Web hits on http://arXiv.org/ per week, from July 1997 through September 2000 (not including the mirror sites), currently corresponding to over 3 million hits a month. Green (upper)—number of connections in each week; red (middle)—number of hosts that week (divide by 10 for correct number); blue (lower)—number of new hosts that week (divide by 10). (See http://arXiv.org/show_weekly_graph.)

The tremendous savings in time, energy, and real dollars and the dramatic increase in convenience and functionality have revolutionized the way research is conducted in many fields of physics and related areas.

The reach of the project even attracted the attention of the U.S. President's Science Advisor, Dr. Neal Lane. On a 1-day visit to Los Alamos on October 26, 1998, he requested to be briefed on the status of the archive. After my presentation, Lane concluded to the assembled management group that efficient and creative communication of scientific research results is a high priority for his office and that Los Alamos should be very proud of this "unique achievement" and should take advantage of this opportunity to continue leading the way in the development of this exciting new methodology. He expressed tremendous enthusiasm for the archive and for what it had done to revolutionize knowledge distribution in fields like physics, saw it as the prototype of an expanded global knowledge distribution system to cover a wider array of technical scholarly fields, recognized that it was ahead of the alternatives, and confirmed that this was part of the White House vision.

Technology and Sociology

The e-Print archive software provides an automated system for the archiving and distribution of electronic reports. Users can submit documents, obtain listings, search for author names or keywords, and retrieve abstracts or full texts of papers in a variety of choices of high-quality output formats (TeX source, hyperdvi, gzipped hyperPostScript with choice-of-font resolution or type 1 PS, or PDF). Users can also subscribe to the system and thereby receive a daily listing of titles and abstracts of new papers received, and the archive sends out many tens of thousands of messages every weekday to its dedicated worldwide subscribers. Authors who submit papers to the archives are permitted subsequently to submit revised versions of their manuscripts, and a cross-referencing feature permits linkages among e-Print archives serving different disciplines. The archive retains the full date-stamped revision history associated with any entry, so all earlier versions also remain available to readers. The system maintains extensive online help pages, with pointers to necessary software both for composing submissions and for viewing existing submissions on all current platforms. Use of the archives is free of charge.

The e-Print archives remain one of the largest and most active scientific databases on the Internet, and one of the largest open-document repositories. At the end of 2001, the database contained nearly 200,000 submissions. Additional mirror distribution sites in Australia, Brazil, China, France, Germany, India, Israel, Italy, Japan, Russia, South Africa, South Korea, Spain, Taiwan, the United Kingdom, and the United States have given better response times. In the long run, the mirrored distribution also provides a global backup system resistant to localized database corruption and/or loss of network connectivity. It will eventually be converted to a more fully peered server system in which the remote sites can also be used for submission upload.

The success of the e-Print archives to date reflects the major technological advances of the late 1980s and early 1990s. One such advance was the widespread standardization of scientific word-processing packages that enabled researchers to produce high-quality versions of their research documents. The online archives were also enabled by the widespread availability of low-cost, but high-powered, workstations and personal computers with high-capacity storage media. At the end of 2001, the archive source-document formats required about 16 GB to store and were growing at a rate of about 3 GB per year. An average paper (with figures and after compression) requires less than 100 KB to store. Hence, one of the current generation of rapid-access 72-GB disk drives costing a few hundred dollars can hold more than 800,000 articles at an average cost of much less than 1/20 cent per article. For the archive mirror sites, a high-powered Pentium box, including disk space and running Linux, can now be obtained for under $2,500 and is more than adequate as a fully functional archive web server. These technological advances made possible, if not inevitable, the development of e-Print archives and other access methods to the gray literature. More surprising has been the readiness of scientific communities to adopt this tool of information exchange and to explore its implications for traditional review and publication processes.

These e-Print archives are entirely scientist-driven and are flexible enough either to coexist with the pre-existing publication system or to help it evolve to something better optimized for researcher needs. They are an example of a service created by a group of specialists for their own use: when researchers or professionals create such services, the results often differ markedly from the services provided by publishers and libraries. It is also important to note that the rapid dissemination the e-Print archives provide is not in the least inconsistent with concurrent or *post facto* peer review and, in the long run, offer a possible framework for a more functional archival structuring of the literature than is provided by current peer-review processes.

The current methodology of research dissemination and validation is still premised on a paper medium that is difficult to produce, difficult to distribute, difficult to archive, and difficult to duplicate—a medium that therefore requires numerous local redistribution points in the form of research libraries. The electronic medium is opposite in each of the above regards. Therefore, if we were to start from scratch today to design a quality-controlled distribution system for research findings, it would likely take a very different form both from the current system and from the electronic clone that it would spawn without more constructive input from the research community.

Reinforcing the need to reconsider the existing methodology is that each article typically costs many tens of thousands of dollars minimum in salary and much more in equipment, overhead, etc. A key point of the electronic communication medium is that, for a minuscule additional fraction of this amount, it is possible to archive the article and make it freely available to the entire world in perpetuity.

This is, moreover, consistent with public policy goals for what is in large part publicly funded research.

It is also useful to bear in mind that much of the entrenched current methodology is largely a post-World War II construct, including both the large-scale entry of commercial publishers and the widespread use of peer review for mass-production quality control (neither necessary to, nor a guarantee of, good science). Ironically, the new technology may allow the traditional players from a century ago, namely the professional societies and institutional libraries, to return to their dominant role in support of the research enterprise.

On the one- to two-decade time scale, it is likely that other research communities will also have moved to some form of global unified archive system without the current partitioning and access restrictions familiar from the paper medium, for the simple reason that it is the best way to communicate knowledge and therefore to create new knowledge. One of the objects of this project is to research and experiment how such archives can be established, reliably maintained, and designed to interoperate, so that possible models for future research communications infrastructure can be assessed experimentally.

Impact on Scientific Information Exchange

The e-Print archives have already significantly altered the processes of research and information exchange in the fields they serve, with even more fundamental changes undoubtedly yet to come.

In a feature article, "Physicists in the New Era of Electronic Publishing," in the August 2000 issue of *Physics Today*, [2] then American Physical Society president James Langer begins as follows:

> Some of my colleagues in Santa Barbara—the string theorists, for example, and several of my coworkers in condensed matter theory as well—insist that they don't need *The Physical Review*. For research purposes, they don't need refereed print journals at all. They are producing remarkable results this way, so I take them very seriously.
>
> What they are doing is using the Los Alamos e-Print archive for all of their research communications. They check it every day for new information. They post all their papers there, cite references by archive number, use the search engine to find other papers, and need little or no other publication services. Publication on the archive is instantaneous. It costs the users nothing and is self-organizing—or at least it appears so. It's also far more democratic than the old system with which I grew up. Physicists all over the world can post their research results without being hassled by grumpy editors and referees. And they don't have to be part of some inner circle of accepted colleagues to be on the preprint mailing lists; they can find out what's new on the archive just as soon as everyone else does. This new system of scientific

> communication—to a large extent Paul Ginsparg's brainchild—is doing far more than just providing an ultraeffective mode of operation for scientists. It is forcing a complete re-evaluation of the role of scholarly journals and, inevitably, an equally thorough re-evaluation of the roles of organizations like the American Physical Society.

Electronic access to scientific research has been a major boon to developing countries, because the expense of connecting to an existing network is infinitesimal compared with that of constructing, stocking, and maintaining libraries. The e-Print archive administrators frequently receive messages from physicists in developing countries in eastern Europe, South America, and the Far East confirming how much better off they find themselves even in the short term with the advent of the current electronic distribution systems—no longer are they "out of the loop" for receipt of information. Others report the feeling that their own research gets a more equitable reading, no longer dismissed for superficial reasons of low-quality print or paper stock.

A recent article in *Science* [3] details some of the impact on astrophysics, with comments like "[Astro-Ph] has taken over [from the journals] as far as I'm concerned," says Nick Kaiser, an astronomer at the University of Hawaii, Manoa. "Every morning the first thing I do is read the Astro-Ph e-mail I get." The server will soon put traditional journals out of business, Kaiser predicts, and formal peer review will give way to some sort of electronic dialogue. That would not surprise Princeton University astrophysicist, David Spergel. "For me personally, publication doesn't matter," he says. "I've pretty much stopped reading the journals."

More conventionally refereed and moderated forums are of course neither obsolete nor necessarily undesirable. In some, though not all, disciplines, the refereeing process is regarded to play a useful role both in improving the quality of published work and in filtering out large amounts of irrelevant or incorrect material for the benefit of readers. The archives already benefit from an automatic form of peer review since users typically replace their submissions in response to direct feedback, and subsequent revisions frequently benefit as much or more from this feedback as from the conventional referee process. The archives have also implemented full "version control," meaning that any earlier version in a series of revisions can be obtained to aid adjudication of priority disputes or otherwise track changes.

Stimulated and Collaborative Projects

The Open Archives initiative (See [4] for details.) was initiated in conjunction with the Los Alamos library to generalize and distribute the e-Print project. The first meeting took place in Santa Fe in October 1999, bringing together interested parties from a number of universities (Harvard; Massachusetts Institute of Technology; Stanford; Cornell;

Virginia; Old Dominion; University of California; etc.) including both computer and information scientists and librarians. The Open Archives initiative has been set up to create a forum to discuss and solve matters of interoperability between author self-archiving solutions, as a way to promote their global acceptance. The results of this initiative have been extended to mediate not only the interchange of metadata between repositories but also the integrated processing of citations and dynamic linking. In 2002 the interoperable metadata transfer protocol became widely adopted, encouraging the creation of institutional archives and ensuring their aggregation.

The arXiv project is also continuing to work in a number of major national and international collaborations.

With the American Physical Society and the Cornell University Library, T-Division researchers have investigated how best the essential services pioneered by the e-Print archives can be assumed and transformed by existing professional societies and university systems. Work with these partners focuses on how to integrate these services into their very different existing operations and build much more powerful systems in addition to them. The underlying model is one of basic repository services with a generalized service level facilitated by the interoperability protocols described previously.

ArXiv personnel have also participated at great length in the initiation and development of the U.S. National Institutes of Health PubMed Central initiative and remain in close contact to take immediate advantage of their database work and work toward a more functional format for technical documents.

An international collaboration underway with the French Centre National de la Recherche Scientifique, a research funding agency trying to become involved in the international archiving process, has established a French mirror site that can also serve as an alternate upload site for European users.

Another ongoing collaboration is underway with the German Physical Society on metadata issues relevant to interoperability with the internal German physics network, together with an attempt to coordinate this with the European Physical Society.

The Future

The e-Print archives point the way to fundamental changes in communications infrastructure for the full spectrum of research disciplines, within both the natural sciences and the humanities. They model the discovery and retrieval capabilities for the electronic journals of the future and suggest a system for electronic communication of research results better optimized for the needs of researchers.

"The database should always retain its essential T-Division character, however, just as T-Division should somehow retain its identity as a place where such unique projects can be initiated and incubated."

The late 1990s was a turbulent and political time for Los Alamos, and suggested that the arXiv operation had achieved sufficient maturity to permit reestablishment within a broader-based academic culture. In 2001, after 10 yr of operation, the arXiv operation moved to Cornell University, where its second stage of development can take advantage of a wider range of academic input and participation. The database should always retain its essential T-Division character, however, just as T-Division should somehow retain its identity as a place where such unique projects can be initiated and incubated.

Acknowledgements

Many others from T-Division and elsewhere at the Laboratory made crucial contributions to the development of this project during the 1990s. Tanmoy Bhattacharya was a continual source of technical expertise; Erica Jen coordinated the first National Science Foundation funding starting in 1994 and organized the expansion into the nonlinear sciences; Joe Carlson set up the original gopher interface in late 1992 and the original World Wide Web interface in spring 1993 (the first at Los Alamos); Terry Goldman provided useful feedback and guidance; Dave Forslund (Advanced Computing Laboratory) and Rick Luce (Research Library) helped lobby for support from within the Laboratory; and the Laboratory Research Library provided essential logistical support toward the end of the project's first decade. Also, the group leader Geoffrey West of the T-Division endorsed the archives' establishment within the group, repeatedly promoted Los Alamos as an appropriate continued sponsor of this activity, and worked overtime to find support both from within the Laboratory and from government funding agencies. Former T-Division Director, Richard Slansky, endorsed the project's establishment within the division. Postdoctoral fellows who worked on the project during the 1990s included Mark Doyle, Rob Hartill, Brian Gough, Thorsten Schwander, and Simeon Warner. Paul Fendley participated as a limited-term staff member. The project also benefited from the consistent support of all members of the Elementary Particles and Field Theory Group, and use of group resources, through group and acting group leaders (Fred Cooper, Emil Mottola, Salman Habib, and Rajan Gupta).

Beowulf Computing

by Michael S. Warren

Beazley
Feng
Feynman
Germann
Goda
Hagberg
Lomdahl
Moulton
Neal
Salmon
Slansky
Sterling
Warren
Weigle

Using off-the-shelf computer hardware and the Linux operating system, Theoretical Division researchers built Loki, a supercomputer that performed at levels of machines that cost ten times as much. In 1997, Loki won the Gordon Bell Prize in the performance-per-dollar category at the annual supercomputer conference, SC97: High-Performance Networking and Computing, and was featured on the cover of the* Linux Journal. *Next, T-Division personnel built Avalon, an improved computer over Loki that also won the Gordon Bell Prize in the price/dollar category. Avalon was also 113th on the list of TOP500 fastest computers in the world. Both of these Beowulf-cluster computers have proved that supercomputers don't have to be enormous and expensive. T-Division researchers have partnered with the National Aeronautics and Space Administration to continue their work toward ever-faster and more space- and cost-efficient supercomputers.

Supercomputing at Los Alamos National Laboratory has come a long way since Richard Feynman solved diffusion problems by using a bunch of people in a room passing punched cards among their mechanical calculators. For five decades, each new generation of supercomputers led to higher performance and the solution of ever-larger problems. However, in the mid-1990s, this progression began to falter. The supercomputer industry was in transition from big-iron vector supercomputers typified by Cray, to massively parallel machines like the Connection Machine. In this unsettled environment, political change reduced the amount of money supporting high-end computing, and economic pressures began to take their toll. Cray Computer declared bankruptcy, other marquis names were bought by lower-flying competitors, and the fastest computer in the world (the ASCI Red System) did not use exotic technology at all: instead it used lots of ordinary Intel microprocessors.

It was with this backdrop that in 1996 I wrote a proposal for internal Laboratory research funding to develop the equivalent of a small parallel supercomputer using off-the-shelf personal computer hardware and the Linux operating system. The proposal was rejected by the committee reviewing it, but a memo to Theoretical Division Director Richard Slansky later that year quickly convinced him to provide the funding for the machine that was later named, "Loki."

The precursor of this work was a project at Goddard Space Center of the U.S. National Aeronautics and Space Administration (NASA), headed by Thomas Sterling, which built the first "Beowulf" cluster. The real advantage at Los Alamos was my parallel N-body software, which had been developed over the previous six years with John Salmon, his colleague at the California Institute of Technology (Caltech). The code had been thoroughly stress-tested on several generations of unreliable commercial parallel computers, so it provided a stable software infrastructure upon which one could experiment with different hardware solutions and be confident that if something was broken, it wasn't the software. This code and the astrophysics it could enable also provided a strong incentive to find a reliable and affordable computing platform to run on.

Loki

With the assistance of Patrick Goda, I soon had Loki computing at rates of over 1 billion operations per second. For a cost of $50,000, Loki was performing at levels typical of machines that cost ten times as much. Not long after being built, Loki was packed up and shipped to the yearly supercomputer conference. As anecdotal evidence of the flexibility of the architecture, at the conference, Loki was joined with a similar machine by running 16 Ethernet cables underneath the carpet to the neighboring Caltech research booth. The combined machine set a new Beowulf cluster performance record, and Caltech happily paid a fine to the conference organizers for having an "unauthorized network connection." At the same conference a year later (in 1997), Loki won the Gordon Bell Prize in the performance-per-dollar category, demonstrating that the new architecture really was competitive with the best machines that industry could offer. In January 1998, Loki was featured on the front cover of the *Linux Journal*.

Avalon

Building upon the experience gained with Loki, we built a larger machine called Avalon. Avalon used 140 Alpha processors and cost about $300,000. Helping me in this endeavor were Aric Hagberg and David Moulton of T-Division and David Neal of the Center for Nonlinear Studies. Avalon again won a Gordon Bell prize/performance prize, running my gravitational N-body code and the Scalable Parallel Short-range Molecular (SPaSM) molecular dynamics code developed by Peter Lomdahl, Timothy Germann, and David Beazley of T-Division.

Additionally, Avalon ranked in 113th place on the TOP500 list of the fastest supercomputers in the world, at 48.6 Gflops. At that time, only 12 countries in the world had a computer faster than Avalon, and only 10 commercial enterprises had ever developed a machine that could demonstrate performance in that regime. In the next couple of years, the Avalon cluster was used by researchers in T-Division and elsewhere at the Laboratory for results presented in over 50 scientific publications.

Since then, Beowulf clusters have filled a wide niche in high-performance computing. As renowned computer architect, Gordon Bell, says in a recent paper, "only five years after its introduction, Beowulf has mobilized a community around a standard architecture and tools. Beowulf's economics and sociology are poised to kill off the other two architectural lines and will likely affect traditional supercomputer centers as well."

Into the Future

We have been funded by NASA to use clustered computing technology to push the boundaries in the related area of data storage. Using distributed disk arrays, we have proposed to develop by 2005 the capability to store and process more than a petabyte (1 million MB) of data for a cost of less than $1 million. In another collaboration with Wu-chun Feng and Eric Weigle of the Los Alamos Computer and Computational Sciences Division, we put together a cluster with 240 processors that fits in a single rack (1 cubic meter). This new "Green Destiny" cluster sets new records for the amount of computing per square foot of floor space and per watt of power, which will soon become limiting factors for all supercomputers. Finally, our third-generation Beowulf cluster, the "Space Simulator," became operational in October 2002. Using 288 Pentium-4 processors and a gigabit Ethernet network, the Space Simulator currently ranks as the eighty-fifth fastest computer in the world.

In the final analysis, the work in T-Division on Beowulf clusters was not so much research as a proof-of-concept, which led to a much greater appreciation and acceptance of the Beowulf architecture as a bonafide high-performance computer architecture for scientific computing.

Beyer
Bivins
Devaney
Eckert
F. Evans
Everett
Fermi
Feynman
Frankel
Hamming
Lazarus
Louck
Mark
Mauchly
Metropolis
Mulliken
Pasta
Richardson
A. Rosenbluth
M. Rosenbluth
Rota
Rotenberg
Spack
Stein
A. Teller
E. Teller
Tuck
Turkevich
Ulam
Urey
von Neumann
Wells
Wooten
Worlton

The Nicholas C. Metropolis Legacy

by James D. Louck

The following information is excerpted from a presentation by James D. Louck at the Los Alamos National Laboratory Family Day celebration, April 13, 2001.

Nicholas C. Metropolis, the son of immigrant parents, grew up to make a major impact on the fields of computing and mathematics. He came to Los Alamos in the beginning, April 1943, and stayed on to help engineer numerous advances in the emerging computing field and to help produce innovative mathematical techniques (such as the Monte Carlo method and the Metropolis algorithm). Some of his research forms the bases for many scientific endeavors today. As an example of the importance of his work, the popular journal,* Computing in Science & Engineering*, selected the Metropolis algorithm as one of the top 10 algorithms that have the "greatest influence on the development and practice of science and engineering in the 20th Century."

Nicholas C. Metropolis was born June 11, 1915, in Chicago, Illinois, the third of seven children (three brothers and four sisters). His parents came from Greece and were working-class Americans in Chicago. He spoke only Greek until he entered elementary school, where he was soon identified as a precocious child. He was a member of the swim team in high school and set a school record, I believe in the butterfly stroke. I think he also swam in college. He also was an avid tennis player and skier throughout his life.

He went through undergraduate and graduate school at the University of Chicago, where he received his Ph.D. in 1941 in the then emerging field of quantum chemistry under Robert S. Mulliken, Nobel Laureate in Chemistry in 1966. After working for a short period in 1942 with Urey at Columbia University, Metropolis became a member of the Metallurgical Laboratory, which brought him to Los Alamos in April, 1943, as it did many other members of Enrico Fermi's group which had proved in the famous reactor experiment under Stagg's stadium that the multiplication of neutrons made feasible the explosive release of nuclear energy.

The War Years and Beyond

During the war years and the development of the atomic bomb, it became clear that accurate and fast numerical calculations would be required for continued development of the physics of nuclear weapons. The Marchant hand calculators that Metropolis and Richard Feynman maintained for use during the war would not do the job. The ENIAC, developed by John Mauchly and Presper Eckert at the Moore School of

Electrical Engineering at the University of Pennsylvania, with John von Neumann consulting (itself a complex and intriguing story), set the stage for modern computing stratagems, as set forth by the genius of von Neumann.

To my knowledge the first mention of Metropolis in connection with nonmechanical computing comes in early 1945 when he and Frankel traveled to Philadelphia to meet with Mauchly and Eckert to discuss the ENIAC and determine its suitability for a variety of physical problems going beyond the ballistic trajectories for which the ENIAC had been designed. Out of these post-World War II interactions, the fast-growing needs of Los Alamos to meet challenging defense postures, and at the urging of the likes of Nicholas Metropolis, John von Neumann, Stanislaw Ulam, Carson Mark, Foster Evans, Cornelius Everett, Marshall Rosenbluth, Edward Teller, Anthony Turkevich, and others, was born the first MANIAC (mathematical analyzer, numerical integrator, and computer).

MANIAC I became operational in 1952, and the opening of the Nicholas C. Metropolis Center for Modeling and Simulation for Family Day comes on the 50th anniversary of this achievement. Aside from the important use of these computers in bomb calculations, these early vacuum-tube machines kept Los Alamos at the forefront of computer development. This is true not only in the sense of building the machines themselves and their influence on commercial developments, but perhaps more significantly in the initiation of effective mathematical techniques for carrying out such calculations and methods for instructing (programming) the machine how to do its work.

It is remarkable that Ulam, Metropolis, and von Neumann in the late 1940s and early 1950s produced mathematical techniques, known as the Monte Carlo method and the Metropolis algorithm, of such penetrating depth that they are mainstream methods of calculating to this day—in the very supercomputer being installed in this building. The paper that set forth the technique known as the Metropolis Algorithm was authored by Metropolis, Marshall Rosenbluth and his wife, Arianna, and Edward Teller and his wife, Augusta, in 1953. This paper has been cited over 6000 times by other researchers and has been included by the popular journal, *Computing in Science & Engineering*, as one of the top 10 algorithms that had the "greatest influence on the development and practice of science and engineering in the 20th Century."

Impact of Metropolis Work

The initiation of computers into our society at the mid-20th Century was a profound technological development. Computers reach into our daily lives from modern health services, education, shopping, voting, information gathering, etc. They monitor and shape our society and thinking in many ways not yet understood. Not only have computers changed our daily lives, they have changed the very way in which we

do science itself. With its Turing machine and algorithmic approach to logic, the computer even challenges the meaning of mathematics. Richard Hamming, inventor of the Hamming codes and a lifelong friend of Metropolis, beginning with college days, debated with Metropolis for decades the issue as to whether only computable numbers have meaning. No doubt it was these concerns about the foundations of mathematics that led to the 30-year period of collaboration on the basis of mathematics between Metropolis and Gian-Carlo Rota, professor of combinatorial mathematics and professor of philosophy at Massachusetts Institute of Technology.

Why has Metropolis been singled out to be honored by having a building named after him—a first for the Laboratory and the University of California—and given a place of prominence in this mural (Fig. 1) on the wall? He was the first to admit that MANIAC I and II would not have been possible without the untiring efforts of individuals such as Mark Wells, Roger Lazarus, Marge Devaney, Paul Stein, Jim Richardson, Bill Spack, Jack Worlton, and others. He was the first to credit the early nonweapons accomplishments of the MANIAC to colleagues such as Fermi, Pasta, Ulam, Tuck, Beyer, Bivins, Rotenberg, Wooten, and many others. It is because over the entire landscape of the computer revolution, from their building to their logical operations to their multipurpose uses to their parallel unitary capabilities, one sees the imprint of Metropolis. Director John Browne, following Metropolis' death on October 17, 1999, wrote: "Nick's work in mathematics and the beginnings of computer science forms the basis for nearly everything the Laboratory has done in computing and simulation science."

Fig. 1. The idea and design of the mural originated with artist Donald Montoya of the Los Alamos Communication Arts and Sciences Group and a team of other group members, with some advisement from members of T-Division. The 12-ft × 30-ft mural, which hangs in the Metropolis Center for Modeling and Simulation, is an attempt to capture, through fleeting glimpses of events and ideas, the legacy that Metropolis left us.

History of Quantum Computation in T-Division

by Gennady P. Berman and Gary D. Doolen

Quantum computers offer the promise of efficient calculation of problems thought to be intractable on conventional classical computers. These computers replace classical ones (1s) and zeros (0s) with quantum states in individual atoms. These machines give incredibly more power to the concept of parallel computing. Many government agencies are interested in building quantum computers and are funding experimental and theoretical efforts in this field. Theoretical Division researchers have been and continue to be in the forefront of developing such computers.

A quantum computer can factorize large integers and search large databases in many fewer operations than a classical computer. It can also model the behavior of atoms and solve partial differential equations, including hydrodynamical equations. Research on quantum computers is stimulating progress in many technological applications including atomic-scale microscopy for a single-spin measurement, interferometer devices based on Bose-Einstein condensates, and nano-devices.

The simplest conceptual quantum computer is a one-dimensional chain of interacting spin-1/2 particles. These spins are placed in a magnetic field with a uniform gradient, causing each particle to have an individual-spin transition energy. Logic operations are carried out using a sequence of microwave pulses to flip the spins of the desired particles at the desired time interval. Reading the result of a given computation requires measuring the spin of each particle.

The idea of quantum computation was first seriously introduced independently in the early 1980s by Richard Feynman, one of the original members of Theoretical Division, and by P. Benioff. They demonstrated theoretically that quantum bits (qubits) can execute reversible computations, a capability that minimizes heat production.

Currently, research in quantum computation is a large international effort funded at the level of a few hundred million dollars per year. The main reason for this research investment is that quantum computers appear to be able to solve a set of problems that are intractable on digital computers. Many government agencies are interested in building quantum computers and are funding experimental and theoretical efforts in this field.

Benioff
Berman
Chuang
R. Clark
Cory
Doolen
Feynman
Grover
Gubernatis
Hammel
Hawley
Hughes
James
Kamenev
Kane
Knill
Kwiat
Laflamme
Lloyd
Milburn
Ortiz
Shor
Tsifrinovich
Yepez
Zalke
Zurek

One such problem is the prime factorization of large integers. In 1994, Peter Shor demonstrated a quantum prime-factorization algorithm that requires only operations that are polynomial in the number of digits of the number to be factored. The number of operations required to do this calculation on digital computers is exponential in the number of digits.

Another algorithm, discovered by Lov Grover in 1996, uses operations that grow only as the square root of the number of entries in the database, instead of the linear dependence on the number of entries required for digital computers. The main advantage of a quantum computer is its enormous parallelism. (Simultaneous operations can be done on 2^L states if there are L interacting spins.)

The recent history of quantum computation in the T-Division began before Peter Shor created his factorization algorithm. In 1993, Seth Lloyd, then a postdoctoral fellow in the Center for Nonlinear Studies and now at the Massachusetts Institute of Technology (MIT), proposed a quantum computer that uses electromagnetic pulses to induce resonant transitions in a chain of weakly interacting atoms. In 1994, T-Division's Gennady Berman and Gary Doolen, and Vladimir Victor Tsifrinovich of New York's Polytechnic University extended this idea to Ising spin chains for a specific quantum Hamiltonian. Now many quantum computer proposals are based on the Ising spin systems in which quantum operations are performed using electromagnetic pulses.

Two leading quantum computer approaches use nuclear spins as qubits for quantum operations.

(1) One approach is based on quantum computing with liquid-state nuclear magnetic resonance. In this approach, a single qubit is represented by an ensemble of nuclear spins in organic molecules. Issac Chuang and coworkers at International Business Machines (IBM) and Stanford University used a quantum computer to demonstrate prime-factorization of the number 15 in 7-qubit molecules. In T-Division, Ike Chuang, along with Raymond Laflamme (now in Canada) and Wojciech Zurek initiated research in this field. They were aided by E. Knill and coworkers of the Los Alamos Computer and Computational Sciences (CCS) Division.

This work is continuing in collaboration with L. Viola of the Los Alamos CCS Division and D. Cory and coworkers at MIT. T-Division's Christof Zalka, now in Canada, made significant contributions to the understanding of optimal properties of quantum algorithms. These researchers also derived important theoretical results for error-correction codes and thresholds for errors in quantum computation.

(2) The second approach uses nuclear spins as qubits and is based on a strategy proposed by Bruce Kane in 1998. In his approach, the qubits are the nuclear spins in phosphorus-31 atoms, implanted in a silicon-28 substrate.

A major obstacle for quantum computers is decoherence. Sound waves and most interactions with a thermal reservoir cause the initially coherent quantum states to decohere or to lose their relative spin alignment during the process of computation. Many years before quantum computation began as a scientific initiative, Wojciech Zurek achieved important results for understanding the effects of decoherence of quantum systems due to interactions with the environment. Zurek not only identified the main cause of decoherence, but he also was the first (1982) to estimate the characteristic time of decoherence. Research in this direction includes a number of papers that has grown exponentially in time. Several books have also been published on this topic.

An important theoretical problem for quantum computation is building a theory of quantum computation with many (1,000 and more) qubits. This theory is needed to understand optimal ways to perform scalable quantum computation. With Dmitry Kamenev, we developed a useful perturbation theory for quantum computation with many qubits and applied it to electron and nuclear spin chains with Ising nearest-neighbor interactions in an inhomogeneous magnetic field. Using electromagnetic resonant pulses, we also simulated quantum logic. This theory was used to calculate the optimal design parameters required for scalable quantum computation and for benchmarking complicated quantum operations involving 1,000 qubits. This theory is now being applied to self-assembled organic molecules in collaboration with experimental groups from Pennsylvania State, Rice, and Cornell Universities.

In 1998, Robert Clark of Australia and Chris Hammel (now at Ohio State University) of the Los Alamos Materials Science Technology (MST) Division began research on a silicon-based solid-state quantum computation. This project is based on Kane's proposal for solid-state quantum computation using nuclear spins of phosphorus donors in silicon. Daniel James, working with Berman, developed theoretical approaches for measuring a qubit state in this quantum computer device.

With our coworkers, we developed a theory of nuclear-spin measurement based on Kane's scheme of swap operations of information between the nuclear spin and electron spins of the same phosphorus atom. We estimated the optimal time for the nuclear spin measurement depending on the noise produced by metallic gates. James collaborated with Paul Kwiat (now at the University of Illinois) to develop theoretical approaches to measuring entangled states using optical techniques. This direction is very important for understanding complicated quantum states that have no classical analog. Another important direction of James's research is quantum computation with trapped ions, in which he achieved many important results in collaboration with the experimental group of Richard Hughes in the Los Alamos Physics Division.

Berman and coworkers developed a theory of a single-spin measurement using a new technique based on cantilevers. This magnetic resonance force microscopy (MRFM) approach is capable of detecting a force of the order 10^{-18} Newtons. With coworkers, Berman and Hammel proposed a quantum computer based on MRFM.

The group of Hughes, Berman, James, and coworkers developed a theory of quantum instabilities and chaos in ion traps. This research was needed to understand how to design stable trapped-ion quantum computers.

Recently, T-Division's Gerardo Ortiz and James Gubernatis, in collaboration with Knill and Laflamme, developed an approach for using a quantum computer to simulate fermions.

In 2000, Berman and coworkers, in collaboration with the experimental group of M. Hawley of the Los Alamos MST Division, initiated another important direction of quantum computation based on scanning-tunneling microscopy. This promising direction involves both quantum computation and single-spin measurement using a scanning-tunneling microscope (STM).

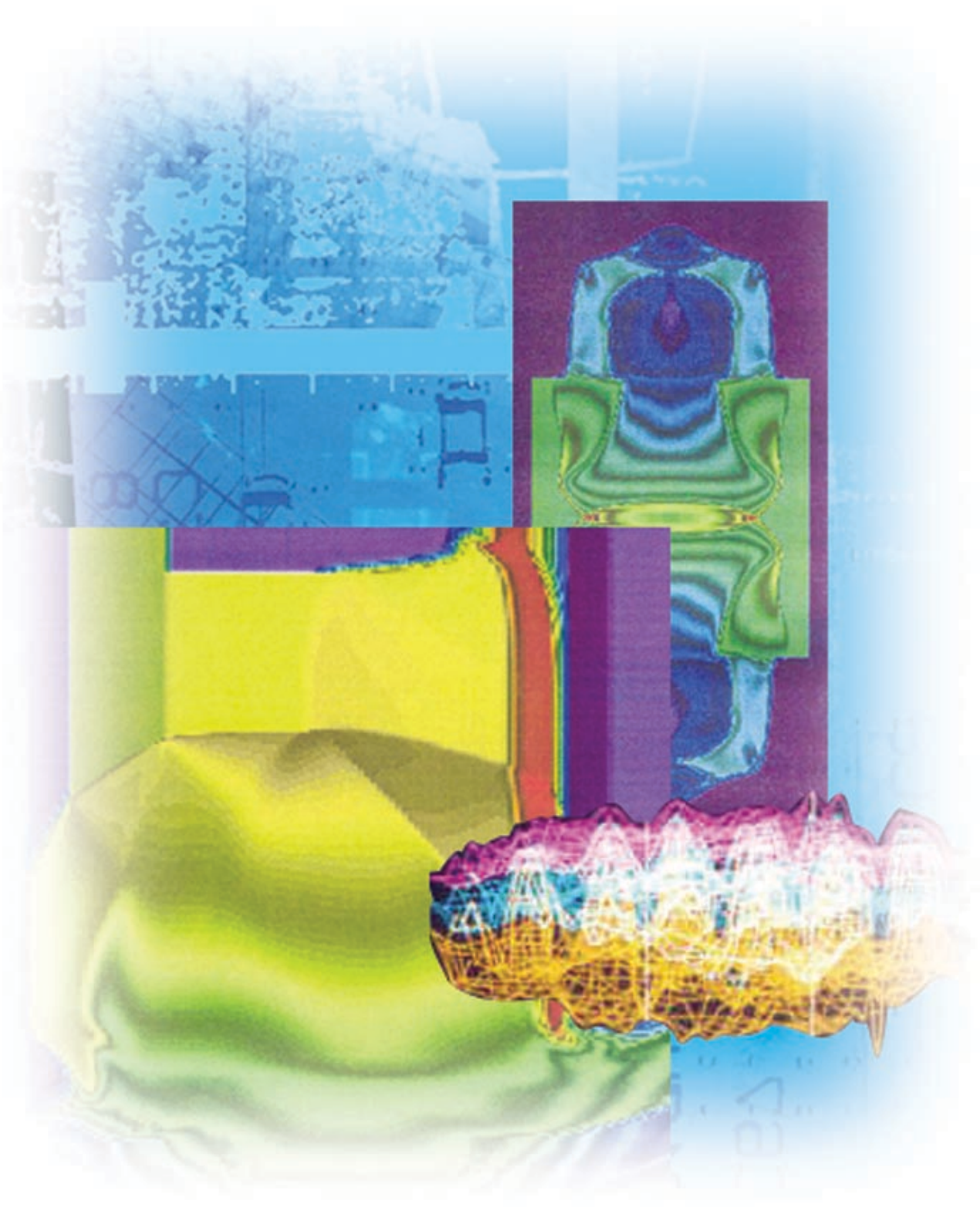
In 2001, Knill, Laflamme, and Milburn created a scheme for efficient quantum computation with linear optics.

Berman and coworkers initiated another important direction in quantum computation, the Type-II quantum computer, first proposed by Jeff Yepez at the Air Force Research Laboratory. A close collaboration now exists between T-Division and Yepez's group. A working Type-II quantum computer would be ideal for solving partial differential equations, including hydrodynamical equations. Cory and coworkers recently used the Type-II quantum computer to solve the diffusion equation in one dimension.

Now quantum computing research is an important part of the newly created Quantum Institute at the Laboratory.

When I woke up it was time to get ready for the explosion [Trinity test, July 1945, near Alamogordo, New Mexico]. . . . I was fascinated with the preparations which Fermi was making to measure the energy yield of the atom bomb by the horizontal distance which a paper strip traveled when released from a height of four feet. . . The explosion took place just before dawn with the sky still dark. All of a sudden, the night tuned into day, and it was tremendously bright. . . We stood there in awe as the blast wave picked up chunks of dirt from the desert soil and soon passed us by…Fermi's paper strip showed that, in agreement with the expectation of the Theoretical Division, the energy yield would be 20,000 tons of TNT. For a period of a few hours, this was the best measurement of the energy yield. It is noteworthy that Professor Rabi, a frequent visitor to Los Alamos, won the pool on what the energy yield would be—he bet on the calculations of the Theoretical Division! None of us dared to make such a guess because we knew all of the guesstimates that went into the calculations and the tremendous precision which was required in the fabrication of the bomb.

—Joseph O. Hirschfelder [former group leader in T-Division], "The Scientific and Technological Miracle at Los Alamos," in *Reminiscences of Los Alamos: 1943–1945*, pp. 76–77.



Introduction to Concentrated Energy

A central responsibility of the Los Alamos National Laboratory is to develop and understand the concentrated energy sources that arise from both chemical and nuclear processes. Nuclear processes are discussed in the chapters on "National Defense" and "Plasma Physics" and in this "Concentrated Energy" chapter on the topic of Inertial Confinement Fusion (ICF). Energy-producing chemical reactions also are described in this book: for example, internal combustion engines are included in the chapter on "Educational, Industrial, and Governmental Collaborations," and explosives, in this "Concentrated Energy" chapter.

Throughout the history of the Laboratory, understanding the behavior of high explosives (HEs) has been crucial to achieving the high material velocities and pressures required to drive nuclear weapons to a critical state. Researchers also need to understand the behavior of HEs in order to perform a broad array of experiments to study material behavior at high strain rates, as described in the chapter on "Material Properties."

The Los Alamos Theoretical Division has been heavily involved in planning and analyzing these experiments and in examining the processes associated with initiating and propagating detonations. In this "Concentrated Energy" chapter, we discuss some of the principal issues and how we have resolved them. The activities exemplify the effective experimental-theoretical collaborations that have contributed to the multidisciplinary capabilities for which the Los Alamos National Laboratory is famous.

Using basic physics principles, reading the literature, nabbing a few experts, and talking to the theorists, we became instant blast experts.

—John H. Manley [former associate director of Los Alamos Scientific Laboratory], "A New Laboratory is Born," in *Reminiscences of Los Alamos: 1943–1945*, p. 35.

The JTF Model for Shock Initiation of High Explosives

Forest
J. Johnson
Kober
Tang

by J. N. Johnson

Reaction-rate models are used by national defense laboratories and industry in calculations of explosive initiation for which the explosive component is designed either to detonate reliably (nuclear or conventional weapons) or is expected to remain inert under various transportation and storage scenarios (explosive safety). Successful prediction and control of each end of the reaction spectrum is obviously important to the success of any explosive device. Los Alamos National Laboratory and Theoretical Division, have long been the center of experimental, theoretical, and computational research involving explosive performance and initiation.

The "JTF" in the JTF model for shock initiation of high explosives comes from the initials of the last names (James N. Johnson, Pier K. Tang, and Charles A. Forest) of the authors of the 1985 paper introducing this multistep approach to shock initiation. [1] An important aspect of the collaboration is the interaction across divisional boundaries at Los Alamos: the authors represent Theoretical, Applied Physics, and Dynamic Experimentation Divisions, respectively. Technically, the JTF model might be thought of as a "next step" in the Forest Fire reaction-rate model. [2]

Hot Spots

The Forest Fire rate of reaction is the rate necessary to accelerate an impact wave along a single, experimentally measured path (initial shock pressure as a function of measured distance to detonation) and is expressed as a function of a single variable, the local pressure. The JTF model introduces an additional variable, the hot-spot temperature, that is created by the shock wave and evolves according to a standard law of chemical kinetics.

Hot spots are small, localized regions of elevated temperature (approximately 400°–500°C) that get the chemical reaction started in shocked liquids and solids. Hot-spot sizes are usually such that little heat is lost through thermal conduction in the extremely short times of the initiation process (typically a millionth of a second). Compared to our common, everyday experiences, hot-spot temperatures are quite high; we recall that a comfortable summer's day in France is approximately 25°C and that water boils at exactly 100°C.

The law of chemical kinetics expresses the rate of reaction in terms of the hot-spot temperature, a frequency factor (roughly, how many times per second the explosive molecule attempts the decomposition reaction before success), an activation energy (the energy barrier which resists spontaneous chemical reaction), and a heat of reaction (the energy we ultimately realize from the reaction). The initial hot-spot temperature (immediately following its creation in the shock front) is expressed as a function of shock pressure in the JTF model. Localized chemical reaction proceeds from that point leading eventually to full detonation.

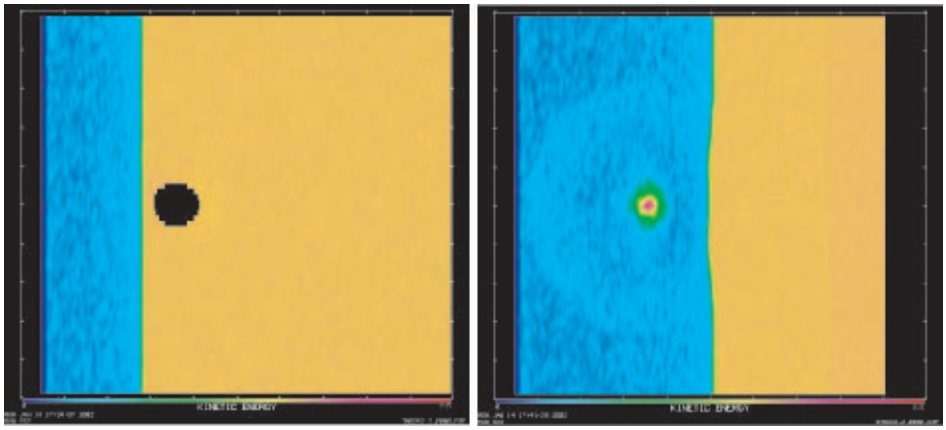


Fig. 1. Molecular dynamics calculation [3] illustrating how a void can produce a region of elevated temperature with the passage of a shock wave. On the left, conditions are shown prior to shock (blue) reaching the void (black). On the right, conditions are shown after the void (now red, indicating higher temperature) has collapsed due to shock. The yellow regions are uncompressed material ahead of the shock wave. (For a digital copy of Fig. 1, contact Edward M. Kober of T-Division and ask for "spasm 2d void" for J. N. Johnson.)

As a simple illustration of a hot spot, Fig. 1 shows a molecular dynamics (MD) simulation of a hot spot in the form of a collapsing void. [3] MD represents the atomic scale and involves many hundreds of thousands of individual atoms but is not yet practical for full-scale engineering applications. The total time encompassing an MD calculation is typically only a few millionths of a millionth of a second, although that time is currently being extended. In the diagram at the left a bubble (the void, here in black) is placed in a liquid explosive (yellow) at atmospheric pressure. A planar shock front impinges from the left (blue). When the shock front engulfs the void, it collapses rapidly and becomes very hot (red) due to extreme localized departure from uniform planar motion.

When the hot-spot temperature is low, the time for localized thermal explosion is long. When the hot-spot temperature approaches 400°C, the time for reaction decreases rapidly. For example, if we take typical values of reaction-rate constants for secondary explosives (those typically used in military applications), we find the time for thermal explosion t at two different temperatures to be [4]

$$t \text{ (77°C, a "lukewarm-spot" temperature)} = 2 \times 10^{9} \text{ s}$$
(approximately 63 yr) or

$$t \text{ (427°C, a "hot-spot" temperature)} = 2.5 \times 10^{-6} \text{ s}.$$

The JTF model incorporates important effects such as this in large-scale computational simulations of actual explosives in military and commercial applications.

Benefits

A particular benefit of including the physical phenomena of hot-spot reactions is that it allows the modeling of additional physical processes without ad hoc modifications or fixes of existing models. As one example, we found experimentally that low-amplitude shock waves can desensitize explosives. This simply means that subsequent shocks no longer have the same ability to drive the explosive to a sustained detonation as quickly as a single high-amplitude shock wave. An ad hoc fix of existing models to represent this effect might be simply to

(1) detect the presence of a low-amplitude shock,

(2) set a reminder in the computational code that this region has been visited by a weak shock wave, and

(3) "turn off" normal reaction if a large-amplitude shock were to come along later.

The JTF model includes the effect of shock desensitization automatically. This is explained as follows: the leading low-amplitude shock creates a "lukewarm spot" that takes too long to react. (See the thermal explosion times shown before.) This lukewarm spot corresponds to the red region on the right of Fig. 1 but with a temperature well below the 400˚C necessary for reaction on the time scale of a millionth of a second. In addition, the source of a potential hot spot is taken out of play for subsequent shocks since the original void has now been eliminated. Further compression of a lukewarm spot can take place only under more-normal conditions, and this process is very inefficient in getting local temperatures up to the 400°C–500°C necessary for reaction on the time scale of the shock process.

The lesson in this case is that more physics means better models. What this means for the engineering and scientific community is that advanced models more than "pay their way" when used in computer design simulations, as long as important physical processes are incorporated and automatically taken into account.

The Hydrodynamic Hot Spot Model

by Charles L. Mader

The Hydrodynamic Hot Spot Model has permitted evaluation of hazards in explosives that have been changed by aging or that have been damaged. The ability to evaluate these hazards results in greater public safety and awareness.

The development of Lagrangian and Eulerian one-, two-, and three-dimensional reactive hydrodynamic codes between 1960 and 1980 by members of Theoretical Division permitted the solution of the problem of shocks interacting with the density discontinuities in explosives. The resulting Hydrodynamic Hot Spot Model, using the Arrhenius rate law with experimentally determined activation energies and frequency factors, allowed calculation of the sensitivity of various explosives to shocks and the determination of the major factors that resulted in a shock-sensitive or shock-insensitive explosive.

The Hydrodynamic Hot Spot Model reproduced the observed variation of the shock-initiation behavior of explosives with different sizes of density discontinuities. This understanding permitted evaluation of hazards in explosives that have been changed by aging or that have been damaged.

The hydrodynamic hot spot study resulted in a model for the experimentally observed desensitization of explosives by weak shock waves and even the quenching of a propagating detonation by weak shock waves.

Shock Initiation of Heterogeneous Explosives—Forest Fire Model

Mader
Popolato

by Charles L. Mader

The Forest Fire Heterogeneous Shock Initiation Model resulted in the solution of many of the explosive initiation, safety, and vulnerability problems that had not been previously understood. The model for describing the shock initiation of heterogeneous explosives was developed by members of Theoretical Division between 1965 and 1975. The model is enhanced by the HOM equation of state, also developed in T-Division, that is used throughout the world.

The Forest Fire Heterogeneous Shock Initiation Model for describing the shock initiation of heterogeneous explosives was developed by members of Theoretical Division between 1965 and 1975. This model permitted for the first time the numerical modeling of the process of a shock-wave-initiating-propagation detonation that results from the decomposition caused by the interaction of shock waves with density discontinuities (such as small holes or voids) of heterogeneous explosives.

The model requires the experimentally measured distance of run to detonation as a function of shock pressure called the POP plot, named after its discover, Alphonso Popolato of the Explosive Engineering Division of the Los Alamos National Laboratory. The model also requires an equation of state (EOS) of the solid explosive, its detonation products, and partially decomposed explosive. This information is arrived at by the HOM EOS developed by members of T-Division and used throughout the world. The model and its many applications are described in my monographs, Numerical Modeling of Explosives and Propellants and Numerical Modeling of Detonations.

The forest fire model solved many explosive initiation, safety, and vulnerability problems that had not been understood previously. Scientists had been seeking answers to these problems for more than a century.

A recent example of the application of the forest fire model is the numerical modeling of colliding/diverging PBX-9502 detonations to a proton radiographic study of the complicated flow that results in a large nonreactive zone in the explosive as the detonation attempts to turn a corner. PBX-9502 is an explosive that contains the shock-insensitive explosive triamino trinitro benzene (TATB).

The shot geometry is shown in Fig. 1 and seven radiographs taken 0.357 microseconds (μs) apart are shown in Fig. 2. Figures 3a and 3b show the two-dimensional (2-D) density contours and mass fraction of undecomposed explosive plots for the recently developed Eulerian

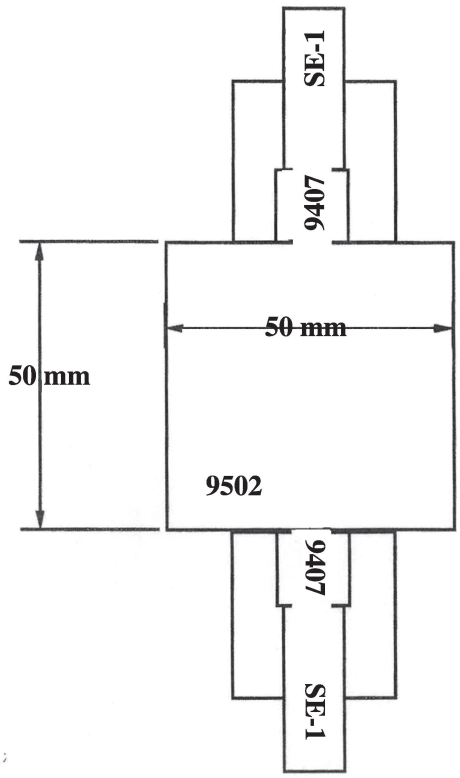


Fig. 1. The geometry of the Proton Radiographic shot designed to study diverging, colliding detonations of shock insensitive high explosives that result in large amounts of unreactive explosive.

NOBEL calculation with forest fire heterogeneous shock initiation burn at the same time as the radiograph. The numerical model with the forest fire model reproduced the experimental observations of the very complicated reactive hydrodynamic flow.

We also used the forest fire model to experimentally observe desensitizing of explosives by a weak shock and quenching of the detonation wave.

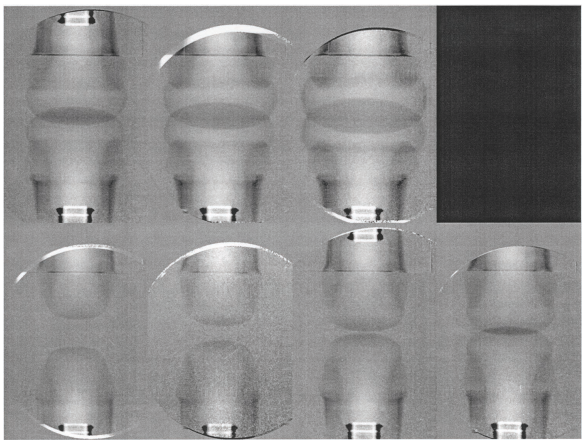


Fig. 2. The seven proton radiographs at intervals of 0.357 μs.

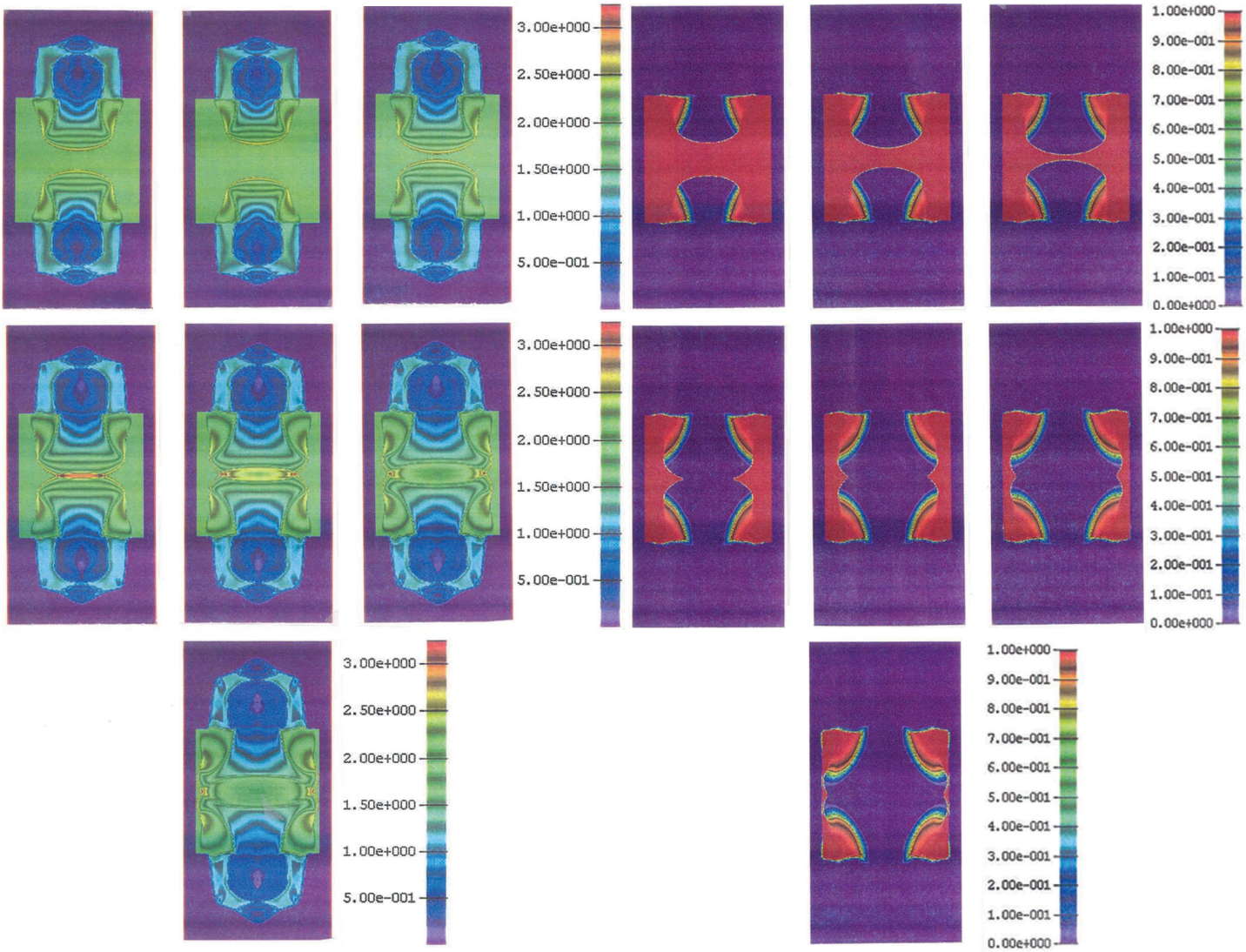
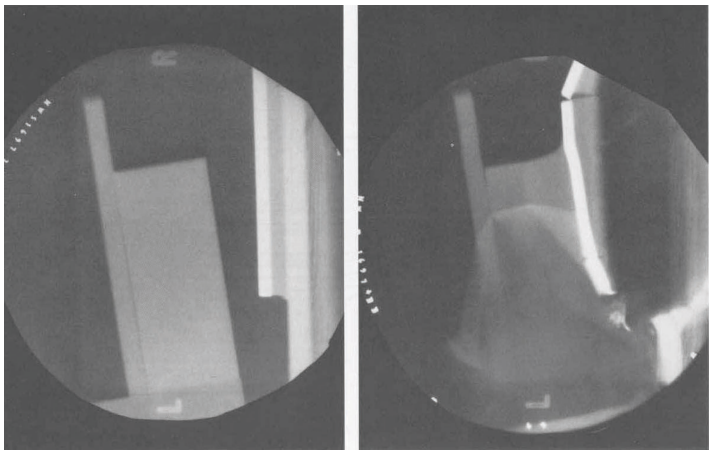


Fig. 3. The 2-D density contours and mass fraction of undecomposed explosive plots for the Eulerian AMR NOBEL calculation with forest fire heterogeneous shock-initiation burn at the same times as the radiographs in Fig. 2.



The PHERMEX radiograph in Fig. 4 shows the desensitization of the shock-insensitive explosive PBX-9502 by a 40-kbar shock wave generated by a flying plate. The propagating detonation is quenched as the explosive is desensitized by the 40-kbar shock wave. The calculated density profile is shown in Fig. 5. We performed this calculation using the new AMR Eulerian reactive hydrodynamic code NOBEL, which includes the desensitization of explosives by preshocking.

Fig. 4. PBX-9502 (95/5 wt% TATB/Kel-F at 1.894 g/cm^3) was shocked by a 6.35-mm-thick steel plate going 0.8 mm/s and initiated by 25.4 mm of trinitrotoluene (TNT) and a P-040 lens after a delay of 53.58 s. The resulting detonation failed to propagate in the preshocked explosive.

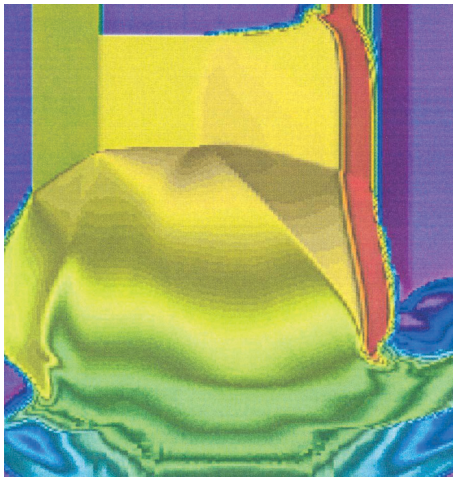


Fig. 5. The numerical model density contour for the experiment described in Fig. 5.

Nonideal Explosives Model

by Charles L. Mader

Mader

Nonideal explosives do not behave as high explosives usually do in the way that they release available energy. In the 1960s and 1970s, Los Alamos researchers developed the aquarium test and a calibration model to describe the unusual behavior of these nonideal explosives. The model is now being applied to uses of commercial explosives throughout the world.

Some explosives do not exhibit the usual properties of high explosives, such as releasing most of the available energy upon detonation and propagating at a fixed detonation velocity and maximum pressure (C-J) over a wide range of geometries. Often these explosives are popular commercial explosives, such as ammonium nitrate-fuel oil (ANFO) used in large quantities for mining and other applications.

Understanding the behavior of nonideal explosives and developing a modeling capability were important problems studied by Los Alamos experimental groups and members of Theoretical Division in the 1960s and 1970s.

The joint efforts of experimental and modeling groups at Los Alamos resulted in the development of the aquarium test, which furnished experimental data for the determination of the chemistry required for calibrating models for describing the nonideal behavior of explosives and calibrating their equations of state.

The aquarium test is used to evaluate the performance of explosives. A cylinder of nonideal explosive is detonated in a tank of water; and an image-intensifier camera records the progression of the detonation wave, the interface between the water and detonation products, and the water shock wave generated by the detonating explosive.

This ability to model the nonideal behavior of commercial explosives is being applied to many applications of explosives throughout the world, such as the extraction of oil from oil shale.

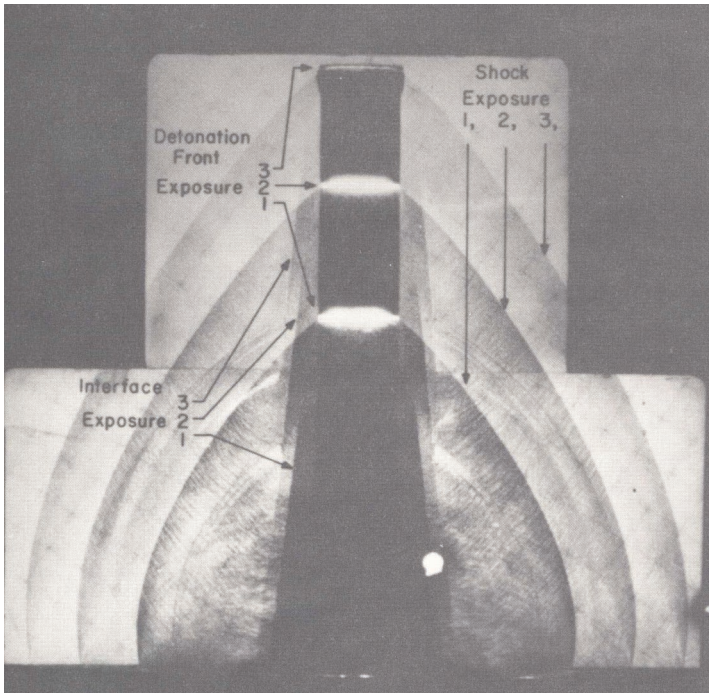


Fig. 1. A photograph with three exposures obtained for the detonation of a cylinder, 10 centimeters (cm) in diameter, that contains the nonideal explosive ANFO (6% oil in ammonium nitrate).

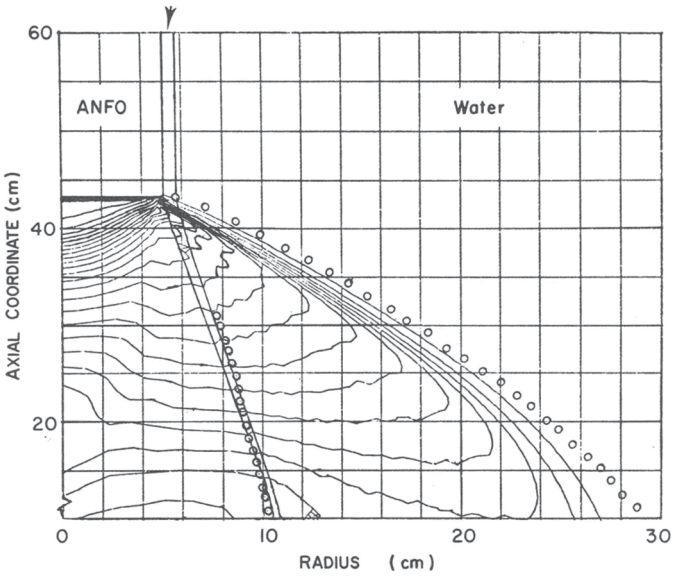


Fig. 2. The calculated and experimental water shock and explosive-water interface positions for a cylinder, with a 5-cm radius, that contains ANFO confined by water in a tank of water. In this geometry, 55% of the ammonium nitrate is unreacted.

The Heterogeneous Explosive Reaction Zone Model

Mader

by Charles L. Mader

Different explosives have reaction zones with varying thicknesses and pressures. Theoretical Division scientists developed a model that analyzes and helps understand these variations in the field of weapons physics.

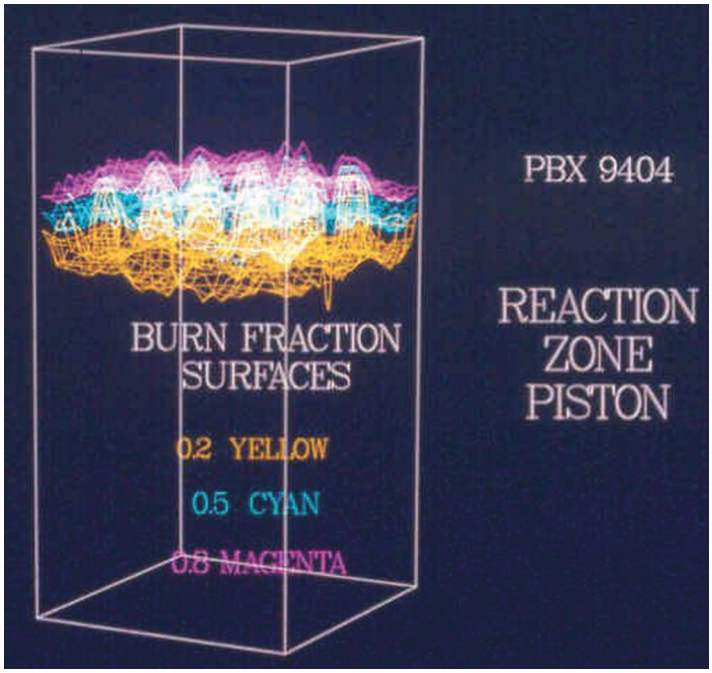


Fig. 1. The 3-D plot of the heterogeneous reaction region of PBX-9404.

The development of Lagrangian and Eulerian one-, two-, and three-dimensional (1-, 2-, and 3-D) reactive hydrodynamic codes between 1960 and 1980 by members of Theoretical Division researchers permitted the modeling of the reaction zone of heterogeneous explosives. (See Fig. 1.) We discovered that heterogeneous explosives exhibit reaction zones of varying thicknesses and state values that result from the interaction of the reacting shock front with the heterogeneities.

Over several decades, experimental studies of the reaction zone of heterogeneous explosives resulted in large differences of the measured reaction-zone thickness and pressure for different methods of measurement. The Heterogeneous Explosive Reaction Zone Model showed that these differences were a consequence of how the different experimental methods averaged the multidimensional heterogeneous thickness and state parameters in determining some mean thicknesses and pressures at the front and end of the reaction zone.

The heterogeneous explosive reaction zone is a dominant feature in detonation-wave propagation, detonation wave failure, and explosive interaction with other materials, such as metal plates, and produces complicated flows at the shock metal interfaces and surfaces.

This topic is important to many problems in weapons physics. The development of massively parallel computers and the AMR Eulerian reactive hydrodynamic code NOBEL will permit the determination of many effects caused by the complicated 3-D heterogeneous explosive reaction zone.

Documentation

by Charles L. Mader

Detailed information about the models discussed in this "Concentrated Energy" chapter are available in several locations and formats.

The Forest Fire Heterogeneous Shock Initiation Model, the Hydrodynamic Hot Spot Model, the Heterogeneous Explosive Reaction Zone Model, the Nonideal Explosive Model, and the unique reactive hydrodynamic codes developed for these studies are described in detail in the Los Alamos monograph, *Numerical Modeling of Detonations*, published in 1978 by the University of California (UC) Press and the second edition, entitled *Numerical Modeling of Explosives and Propellants*, published by CRC Press in 1998. (See Fig. 1.)

Computer movies on these topics and most of the studies described in the monographs are available on the Theoretical Division Detonation and Application Group web site at http://t14web.lanl.gov/Staff/clm/movie/movie.htm.

T-Division started and led a multiyear, multidivision project to collect and publish all the dynamic material properties experimentally determined at Los Alamos National Laboratory over its first 40 yr. This effort resulted in seven data volumes that were published by the UC Press over the years.

T-Division established the Los Alamos Data Center for Dynamic Material Properties that resulted in the following publication, which the Los Alamos Research Library has preserved on their web site http://lib-www.lanl.gov/ladcdmp/. The publication includes the following files that are separate Portable Format Documents (PDFs): *LASL Shock Hugoniot Data*, http://lib-www.lanl.gov/ladcdmp/shd.pdf; *Los Alamos Shock Wave Profile Data*, http://lib-www.lanl.gov/ladcdmp/swp.pdf; *LASL Explosive Properties Data*, http://lib-www.lanl.gov/ladcdmp/epro.dpf; *Los Alamos Explosives Performance Data*, http://lib-www.lanl.gov/ladcdmp/eper.pdf; and three LASL PHERMEX data volumes, http://lib-www.lanl.gov/ladcdmp/ph1.pdf, http://lib-www.lanl.gov/ph2.pdf, and http://lib-www.lanl.gov/ph3.pdf. The data volumes have become the "star catalogs" of their respective fields, preserving the wealth of experimental data generated at Los Alamos over the Laboratory's first 40 yr.

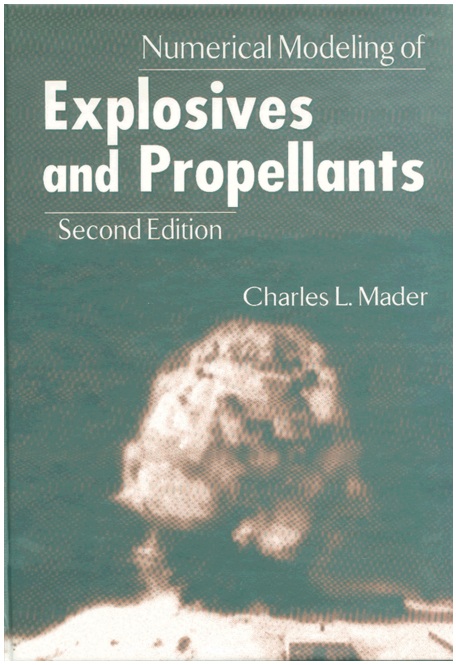


Fig. 1. The cover of the Los Alamos monograph, *Numerical Modeling of Explosives and Propellants,* published in 1998 by CRC Press.

Inertial Confinement Fusion Program

Bethe
Hanson
Kushner
Moliere

by David E. Hanson

One of the ultimate goals of the Inertial Confinement Fusion Program was to enhance commercial power generation. Therefore, understanding the electrical efficiency of lasers was important. Theoretical Division personnel were major participants in this U.S. Department of Energy program from the 1970s into the 1990s, primarily in the area of laser modeling but also in the study of the physics of capsule implosion.

To achieve a thermonuclear burn with Inertial Confinement Fusion (ICF) or any other method, three conditions must be met.

(1) The target fuel material (usually assumed to be a mixture of deuterium and tritium) must be heated to a sufficiently high temperature,

(2) compressed to a high-enough density, and

(3) these conditions must be maintained long enough for the fusion reaction to initiate and proceed.

Satisfying these requirements is a complex challenge. In practice, the process works as follows.

(1) spherical target capsule, a few millimeters in diameter, is suspended inside a larger metal cylinder with openings at each end to allow the laser beams to enter. The enclosing cylinder is commonly referred to as a hohlraum.

(2) The laser beams are configured to illuminate the inside of the cylinder where the light is absorbed by the metal wall, and the energy is transformed to x-rays. The x-ray flux provides a very intense and uniform illumination of the target capsule.

(3) The outermost layer of the capsule absorbs this energy and is ablated.

(4) By Newton's third law of motion, there is a reaction force accompanying the change in inertia of the ablated material. It is this reaction force that accelerates the outer layers of the target capsule radially inward, compressing and heating the fuel to the extreme temperature needed to create nuclear fusion.

The current estimate for the amount of laser energy that must be delivered to a hohlraum to initiate fusion is on the order of one million joules.

In the 1980s, Los Alamos National Laboratory adopted the electron-beam-excited krypton fluoride (KrF) laser as its choice for an ICF driver. Very large double-pass laser amplifiers were needed to satisfy the large energy requirements. The laser medium, typically a one-atmosphere mixture of argon (Ar), krypton (Kr), and fluorine (F) in the ratio

90:9.7:0.3, was excited by beams of high-energy electrons injected transverse to the optical axis. The electron beam currents lasted between 1/2 and 1 microsecond. During this excitation, a series of laser beams sequentially entered the amplifier window, traversed the laser medium, reflected off of the back mirror, and traversed the medium again, exiting by means of the front window. The initial laser pulse, with the required temporal pulse shape, was generated in a master oscillator, and a chain of amplifiers was needed to generate a beam that was sufficiently powerful to efficiently extract energy from the final large amplifier modules. Theoretical Division was given the task of developing and assembling a comprehensive suite of models to predict the performance of the laser amplifiers.

Developing the Models

Since one of the ultimate goals of the ICF program was its application for commercial power generation, understanding the electrical efficiency of the lasers was important. To understand how the energy efficiency of a laser depended on its design and operation parameters required several specific computer models. Three-dimensional (3-D) power deposition profiles in the laser medium, caused by the high-energy electron beams, were modeled using a code developed in the Los Alamos Applied Physics Division. It employed a Monte Carlo technique and used an empirical, shielded Coulomb formula (by Moliere) for the electron-scattering cross sections with a mean loss rate normalized to Bethe's formula. Prof. Mark Kushner at the University of Illinois provided a model for the local production rates of ions and excited-state species of the laser gas atoms and molecules due to the power deposited by the electron beam.

No analytic expression for the time-resolved amplification of a laser pulse is possible for a double-pass amplifier, since the pulse can overlap itself as it is reflected by the back mirror. Consequently, a numerical model is required that provides the source terms for another model, the laser kinetics code. Approximately 70 coupled chemical-reaction-rate equations were solved by numerical methods to track the flow of energy among the approximately 25 chemical species from initial excitation of the argon, krypton, and fluorine to the metastable lasing molecule, krypton fluoride (KrF). Also included in the kinetics model was the stimulated-emission process, in which a KrF molecule interacts with the time-varying electric field of the laser-cavity photons and dissociates into krypton and fluorine atoms plus a coherent laser photon. The amplifier output was computed by integrating the stimulated emission along the optical axis for the duration of the electron beam.

To model an entire amplifier chain from the master oscillator through the final large-power amplifiers required another model that integrated information from the previous models and also considered laser-beam propagation effects. Laser amplifiers were resolved into two spatial dimensions plus a time-dependent electron-beam excitation. Laser kinetics code calculations were parameterized by electron-beam pump power and laser-cavity flux, reducing the number of equations and

unknowns from approximately 70 to 2: one for the medium gain (essentially the local KrF density) and one for the laser cavity flux. The code also included the transverse spatial variation of the electron-beam pump power, laser-beam focusing effects, beam angle with respect to the amplifier axis, and the temporal profile of the initial (master oscillator) pulse. Using laser gains computed with the kinetics code (parameterized on electron-beam deposition powers), realistic system studies could be performed.

This code was also used to simulate experiments performed at Los Alamos laser facilities; e.g., Aurora and the downscaled follow-on, Mercur. Comparisons with experimental gain and absorption measurements were used to test the predictions of the kinetics code. In general, the kinetics code was found to be in agreement with small-signal gain measurements on the smaller amplifiers. In these small-signal gain experiments, the laser intensity was purposefully kept low enough to prevent a significant reduction of the KrF density by stimulated emission, allowing a direct comparison with gain and absorption predicted by the kinetics code.

One issue never resolved was the substantial underprediction of absorption by as much as a factor of two. This was an important parameter with respect to power generation because it imposed a limit on the ultimate efficiency that was attainable by a KrF laser. The code was also used to predict the performance of a proposed ICF laser facility at the Naval Research Laboratory Nike research program, which was subsequently built. This laser began routine operation several years after the predictions were made. Recently, the measured laser-performance specifications were published by the Nike team, and the observed energy output was in good agreement with our model predictions.

Undertanding the KrF Molecule

Since the initial electron beam excitation reactions were so important, and experimental data was not available, a considerable effort was made to compute the electron-collision cross sections for some of the species.

Quantum mechanical methods, such as the Born distorted-wave approximation, were used to compute the energy-dependent cross sections for a number of important excitation and de-excitation reactions. The dissociative excitation of fluorine and the quenching (de-excitation) of KrF are examples. Because the number of excited states for most species was far too great to treat individually with the kinetics code, the multiple states for a particular chemical species were collapsed into a few composite states, such as ions and metastables.

Considerable emphasis was placed on understanding the molecular physics of the KrF molecule. From published fluorescence data, it was known that the spectroscopy of KrF is complex. There are three upper states that can be stimulated to lase, all at different wavelengths. Each of these is composed of a multitude of vibrational states, which form a cascade, allowing the internal energy to relax from the initial state of

formation to the ground state by collisions with free electrons and gas molecules. Computer models were written to compute these relaxation rates and the frequency-dependent stimulated-emission cross sections. With these results, we were able to predict the change in the shape of the laser spectrum as a function of amplification.

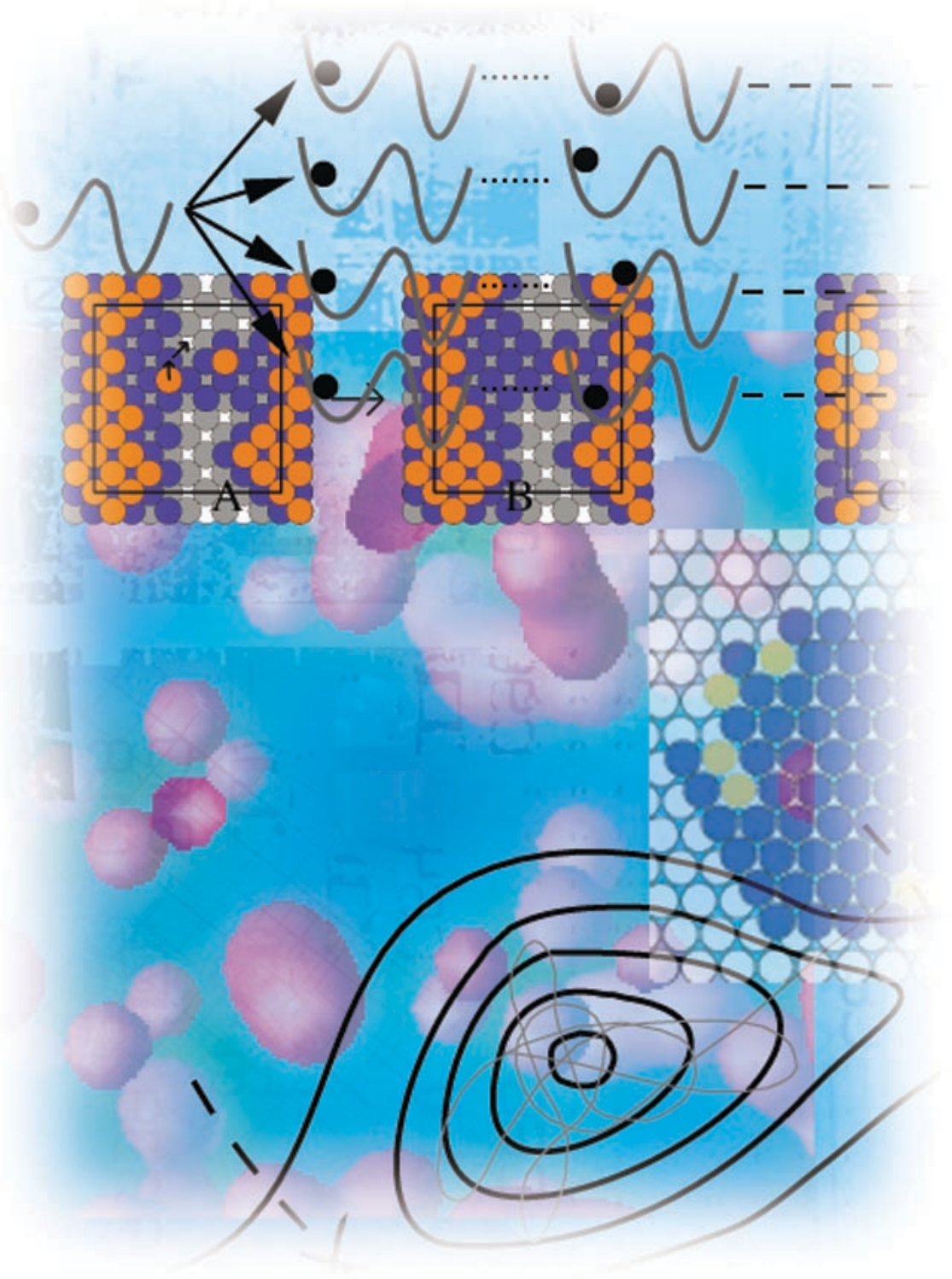
The frequency shape tended to narrow during amplification because the stimulated-emission cross section enters as an exponential of the expression for gain. Keeping the bandwidth as great as possible was an important requirement for target physics reasons; broad bandwidth tended to reduce the microscopic spatial and temporal intensity variations of the illumination of the capsule hohlraum. We predicted that a KrF amplifier chain could easily satisfy the minimum bandwidth requirement of 200 cm^{-1}.

An important consideration in ICF target performance is the material mixing induced by interfacial instabilities (e.g., Richtmyer-Meshkov) and enhanced heat transport induced by turbulence. If instabilities develop during the implosion caused by the ablating outer shell, the target may not reach the desired density and geometry. In order to estimate these effects computationally, one can employ simple turbulence models, like k-epsilon, or more complicated models, like BHR, which is described in the article, "Understanding Fluid Instability and Turbulence," in the "Fluid Dynamics" chapter of this book. BHR, a variable-density turbulence model developed in T-Division in collaboration with French researchers, was implemented into several hydrodynamic codes within T-Division and the rest of the Laboratory. It was used to assess the amount of shell/gas mixing expected in various ICF configurations. These tools have helped the ICF capsule design team estimate the effects of mixing on target performance.

Collaborations with Livermore and Rochester

In 1994, we began a collaboration with a group at Lawrence Livermore National Laboratory to better understand the frequency tripling of 1.06-micron glass-laser radiation for its use as a laser-fusion driver. It was known that target performance could be dramatically improved for laser wavelengths by a factor of two-to-three smaller than the 1.06-micron wavelength of glass lasers. After developing the appropriate theory and computer codes necessary for this study, [1–3] we obtained excellent agreement between our predictions and experiments at Livermore. We were able to explain observed variations in time of the frequency-tripled radiation intensity as being due to spatial variations of the phase of the 1.06-micron radiation: phase "ripples" were found to translate into temporal intensity variations. [4,5] Experiments at Livermore corroborated these predictions. [6,7]

We also studied ways to efficiently frequency-triple broadband radiation, which is desirable for laser fusion. Our studies showed that this could be done by using several crystals in tandem, each having its optic axis oriented appropriately. [5,6] Experiments at the University of Rochester's Laboratory for Laser Energetics later confirmed our predictions.


A
B

Molecular-Dynamics Simulations in T-Division

by Brad Lee Holian, Timothy C. Germann, and Peter S. Lomdahl

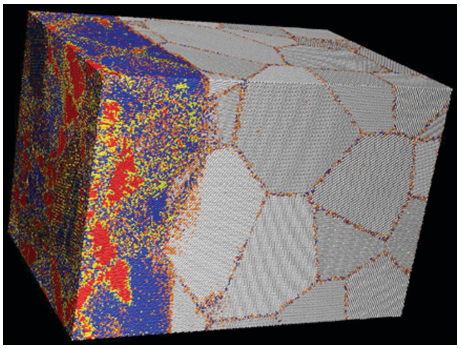
Alder
Ashurst
Beazley
Erpenbeck
Germann
Holian
Hoover
Lomdahl
Voter
Warren
B. Wood

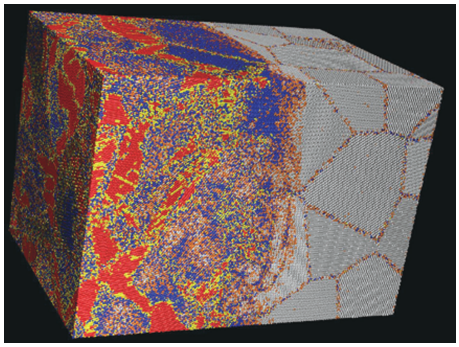
For several decades, molecular dynamics (MD) simulations have been an important tool in statistical mechanics. Applications have ranged from equilibrium equation-of-state calculations to nonequilibrium behavior, such as transport coefficients. Theoretical Division researchers developed a new framework for performing and interacting with large-scale MD simulations by using interpreted scripting languages. Their approach has gained substantial and widespread acceptance in the computational science community. As larger simulation sizes became feasible, scientists have been able to study phenomena that they could not analyze previously. They have gone beyond the "single-crystal era" to study polycrystalline effects in solids and to understand fluid instabilities.

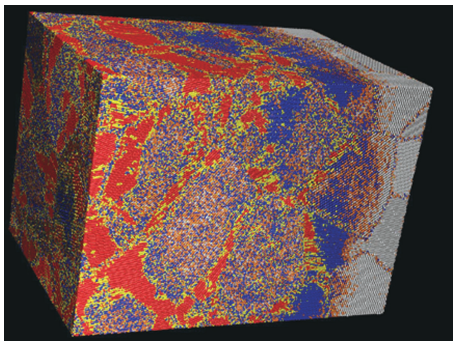
For several decades, molecular dynamics (MD) simulations have been an important tool in statistical mechanics, with applications ranging from equilibrium equation-of-state (EOS) calculations to nonequilibrium behavior such as transport coefficients.

The idea behind an MD simulation is very simple: Newton's classical equation of motion, $F = ma$, is solved directly to evolve the positions and velocities of a large collection of atoms that the fluid or crystal of interest is composed of. Although conceptually simple, this task can present a formidable computing problem. If all atoms interact by means of pairs, or two-body, forces (such as is the case for gravity or for Coulombic interactions between charged particles), the direct solution of this general N-body problem will require the calculation of $N(N-1)/2$ forces, i.e., the computational cost will grow as the square of the number of particles, $O(N^2)$.

To complicate matters even further, many of the empirical force fields, which have been (and are being) developed for materials ranging from elemental silicon to transition metals and hydrocarbons, involve more complicated many-body potentials with explicit or effective bond-bending and torsional terms, whose cost in theory grows as $O(N^3)$ or $O(N^4)$, respectively. Such direct calculations quickly overwhelm the computing capabilities of even the fastest supercomputers; therefore, so clever algorithms have been developed to bring the computational cost under control. For charged systems, where Coulombic interactions are unavoidable, or for astrophysical simulations that require gravitational interactions (as currently carried out by Michael Warren and colleagues in Theoretical Division), the fast multipole method developed in the late 1980s has become the tool of choice, reducing the cost to $O(N)$ by approximating the long-range interactions in a controllable way.







Three snapshots from a 24-million atom molecular dynamics simulation of a shock-induced phase transition (bcc → hcp) in a polycrystalline iron sample. The phase transformation is nucleated at grain boundaries and proceeds inwards towards the center of each grain. Atoms are color-coded by the number of neighbors *n* within 2.75 Ångstrom. The unshocked bcc region (*n* = 8) is gray, the uniaxially compressed bcc (*n* = 10) is blue, the transformed closed-packed hcp grains are red (*n* = 12), and the twin boundaries are yellow (*n* = 11).

Fortunately, the situation is even better for most nonionic materials because the absence of Coulombic interactions makes the neglect of any long-range forces a good approximation. The resulting short-range potentials can be naturally computed with $O(N)$ cost (even for many-body potentials) if the range of the potential is much shorter than the size of the simulation box, because each atom only interacts with a small number of neighbors.

Simulations Have Been Limited

Despite these efforts, until recently (the past decade) most simulations have been limited to a few hundred thousand atoms and a relatively small number of time steps. To carry out realistic simulations of material properties in three dimensions (3-D), however, the inclusion of tens of millions to billions of atoms may be required. (Even a billion-atom MD simulation would be a "small" system, considering that a speck of dust can contain more than a billion atoms.) In addition to this length-scale concern, there is also a time-scale issue to be aware of. A single time step in most MD simulations could be on the order of a few femtoseconds, yet for a realistic experiment it may be desirable to follow the dynamical evolution for a minimum of several microseconds, i.e., a billion MD time steps. Fortunately, the sound, shock, and dislocation velocities in solids are sufficiently fast that processes involving these types of motions are typically complete within a nanosecond for a submicron MD-length scale. (It should be mentioned that methods for reaching even longer MD time scales for rare-event systems have been developed in recent years by Art Voter and coworkers in the Theoretical Chemistry and Molecular Physics Group in T-Division.) [1]

When molecular dynamics was first pioneered by Berni Alder and Tom Wainwright of Lawrence Livermore National Laboratory in the mid-1950s, [2] the emphasis was on the EOS of hard spheres (disks in 2-D), and the systems were limited to only 100 particles, at most. In the 1960s, once it had been established by Alder and Wainwright, using MD, and by Bill Wood at Los Alamos, using Metropolis Monte Carlo, that the thermodynamic limit was achievable by atomistic computer simulations, [3,4] whereby the melting curve for hard spheres was determined for the first time, the work in MD turned toward the establishment of the hydrodynamic limit. Alder and Wainwright showed that the velocity autocorrelation function (VACF) for a hard-sphere fluid demonstrated a persistence of memory far longer than the exponential decay invoked by Boltzmann (molecular chaos).

In the 1970s, Bill Hoover at Livermore and his Ph.D. student, Bill Ashurst, pioneered the method of nonequilibrium MD (NEMD) for a variety of transport properties in fluids. [5] That decade and the one to follow were dedicated to resolving apparent discrepancies between the earlier Green-Kubo (GK) evaluations of transport coefficients (integrals of appropriate autocorrelation functions; self-diffusion given by the VACF, for example). Driving conditions mimicking laboratory experiments are set up in NEMD simulations, and the response as a

function of the force is measured and extrapolated back to zero force to get the GK limit. Because the systems were so small and thermal fluctuations made the signal difficult to separate from the noise, the driving forces used in NEMD were enormous compared to real-world experiments. Thus, when it was subsequently found that NEMD results were correct, but some of the equilibrium fluctuation (GK) measurements had been done incorrectly, the passage of two decades of doubt led many to suspect that MD (and NEMD) were flawed tools for discovering the underlying truth about hydrodynamic behavior. In fact, the two approaches proved ultimately to be highly convergent. [6,7]

Alder and Wainwright's [8] picture of the double vortex set up by a moving fluid element in a hydrodynamic simulation was verified at the atomistic level, where the velocity field around an atom in the fluid was time-averaged in an equilibrium MD simulation. Later, Jerry Erpenbeck and Bill Wood [9] did a quantitative comparison of the VACF from equilibrium MD with mode-coupling (hydrodynamic) theory. As a postdoctoral student in T-Division, Brad Holian showed that NEMD, which is conceptually more straightforward and computationally more efficient than the equilibrium GK method, gives identical results for the diffusion coefficient.

Los Alamos Simulations

In the late 1970s and early 1980s, we at Los Alamos carried out a series of NEMD simulations of shockwave phenomena, including a paper on solitons in the Toda lattice [10] that spawned a number of papers in (nonlinear) mathematical physics. Solid shock waves in 3-D revealed the production of permanent plastic flow above a critical shock strength. [11] The connection to hydrodynamic behavior was made most clearly in an NEMD study [12] of a strong shock wave in 3-D fluid Lennard-Jonesium, where the profiles of density, pressure, energy, and temperature could be compared to Navier-Stokes (NS) predictions, which use the EOS and transport properties measured in equilibrium MD simulations. The NEMD results were surprisingly close to NS, with deviations primarily explained by anisotropy in temperature (higher in the direction of shock propagation). These NEMD calculations were heroic for the time, namely 4,000 atoms.

In the late 1980s, simulations with 10^6 atoms were being performed on conventional vector supercomputers, [13] like the Cray XMP. However, it was the parallel multicomputers that seemed to hold the greatest potential for reaching beyond 10^9 atoms. Coarse-grained transputer systems were being used in successful 2-D simulations with 10^6 atoms [14] at Livermore and data parallel single instruction, multiple data (SIMD) implementations on the 64 K processor Connection Machine CM-200 at Los Alamos demonstrated 2-D simulations for 1.6×10^7 atoms. [15]

In 1992, we began work on a parallel MD code that would allow us to simulate systems of tens to hundreds of millions of atoms. This work was in part inspired by the growing needs of the materials science community to simulate systems large enough to be experimentally relevant; and in part by the availability of the 32 GB, 1024 processor Thinking Machines CM-5 parallel computer at the Advanced Computing Laboratory in Los Alamos. Our work resulted in a computer code, Scalable Parallel Short-range Molecular dynamics (SPaSM), that has been described in detail elsewhere. [16,17] The SPaSM code was ported to several multicomputer platforms supporting both message-passing (on distributed memory computers) and multithreading (on shared-memory systems), as well as clusters of inexpensive personal computers (Beowulf systems).

Although our work was at least partly successful in terms of code speed and size of systems we could simulate, it also opened a whole new set of problems which we had to solve. In particular, utilizing the overwhelming amount of data generated by MD simulations with hundreds of millions of atoms is a major challenge to this day. A typical MD simulation with 100 million atoms that runs for maybe 100 picoseconds (ps), producing a "snapshot" of the system every 2 ps (typically consisting of three coordinates, and a couple of scalar values, say a total of 20 bytes/atom) results in fifty 2-GB files, i.e., a total of 100 GB. While this may not seem like a lot by today's terascale computing standards, it was beyond typical disk-space allocations on the supercomputer systems of the 1990s. A new approach was clearly needed in order to efficiently analyze and interact with data sets of this size.

This realization led us to develop a new framework for performing and interacting with large-scale MD simulations. [18] Rather than using compiled languages for everything, we explored the use of interpreted scripting languages for controlling and interacting with the MD applications. The idea was really quite simple—write performance-critical code in C and use a scripting language for gluing components together and providing a high-level interface.

The approach is essentially identical to that used in commercial packages such as MATLAB or interface definition language (IDL). The critical ingredient in this scheme was the automated code-generation tool simplified wrapper and interface generator (SWIG) (http://www.swig.org), developed by our collaborator David Beazley while he was a graduate student in T-Division. SWIG would automatically provide all the necessary interface code to integrate a conventional C code with a scripting language. By using this model, users could control an MD simulation by interactively entering commands, writing scripts, and adding additional functionality in the interface language, without having to modify the underlying C code. This approach allowed us to perform MD simulations of unprecedented size; but, more important, we could interact with the simulation and analyze the results on the fly—change parameters in the code and proceed with the simulation—all without going through the traditional edit/compile/

run/analyze cycle. This approach has gained substantial and widespread acceptance in the computational science community today, but it was all pioneered in T-Division in the 1990s.

More Studies Now Possible

With this capability, a whole new range of materials science problems became amenable to multimillion-atom NEMD studies. These have provided insight into the atomistic-level details of dislocation nucleation, propagation, and interactions [19] in solids under various forms of loading: tensile (as in crack propagation), [20] compressive (e.g., shock loading), and shear (sliding friction), [21] to name a few examples.

As larger simulation sizes became feasible, new phenomena have been opened up to study. Take shock compression of solids as an example: whereas plastic deformation of 10^4-atom systems studied in the 1980s contained one, or at most two, parallel stacking faults, 10^7-atom simulations in the 1990s were able to exhibit an intricate pattern of intersecting stacking faults that were lying on different slip systems. [22] Advancing computer power has enabled not only larger systems, but also more complex interaction potentials, so that in the first years of the 21st Century, we have studied the shock-induced solid-solid phase transformation of iron, [23] and the shock-to-detonation transition in model energetic materials. [24] With 10^8 to 10^9 atom simulations currently possible, we can now go beyond the "single-crystal era" of MD simulations and begin to study mesoscale (polycrystalline) effects in solids. MD simulations have also reached the mesoscale in the study of fluid instabilities: we are currently simulating the Rayleigh-Taylor problem, with more than 100 million atoms in 3-D.

Simulations of Warm, Dense Matter

by Lee A. Collins, Joel D. Kress, and Stephane F. Mazevet

Theoretical Division scientists have aided such disciplines as planetary science, nuclear weapon development, and high-energy density physics by developing sophisticated quantum molecular dynamics simulation methods. These integrated, sophisticated techniques have shown great promise in modeling a diverse set of environments in which matter exists under extreme conditions.

The environments that characterize the following diverse systems have many features in common: the interiors of giant planets, the initial implosion of an inertial confinement fusion (ICF) capsule, the compression phase in high-energy density devices such as Z-pinches and

Bickham
Collins
Kress
Kwon
Lenosky
Lynch
Magee
Mazevet
Merts
Schneider
Troullier
Younger

explosive-magnetic generators, the insulator-metal transition in systems as varied as hydrogen liquid and rare-gas solids, shocked hydrocarbons in detonations, and various stages in primary and secondary nuclear weapons. In general, these processes span temperatures from 10^2 K to 10^6 K and densities from about 1/100 to 100 the density of the solid. The medium in each system resembles a "soup" of various species, including atoms, molecules, ions, and electrons, that exhibits distinctly nonclassical behavior in the interaction of all the particles. This material state that includes shock-compressed liquids and dense plasmas is now generally referred to as "warm, dense matter."

Quantum Molecular Dynamics

To model such systems, we have developed quantum molecular dynamics (QMD) simulation methods that treat the rapidly moving electrons quantum mechanically and the sluggish nuclei classically. In order to provide a diverse and systematic representation of the quantum mechanical effects, we have treated the electrons at various levels of sophistication from a simple semiempirical tight-binding (TB) model to a state-of-the-art finite-temperature density functional theory (DFT) approach. Since these techniques begin with only the most basic assumptions on the nature of the microscopic particle interactions from which all the macroscopic properties derive, the techniques are designated as *ab initio* or "from first principles."

Through this QMD prescription, we can represent very complex structural and dynamical quantum processes that dominate these media. These methods currently allow for the treatment of a few hundred particles; however, scaling tests for massively parallel computers indicate that simulations with thousands of atoms will shortly become routine. An additional advantage of the QMD methods comes from their integrated nature. Having established the elemental particle interactions, all the static, dynamical, and optical properties rest on an internally consistent set of principles. Whereas, in many models of dense media, the representations of these processes arise from different approaches at different levels of approximation.

The characterization of a warm, dense system as an evolving sample of particles interacting through quantum mechanical forces has become possible only within the last two decades with the development of supercomputing capabilities. Los Alamos has pioneered in this endeavor. As early as 1985, prototypical models, [1] based on a semiempirical determination of the quantum forces in representative snapshots of atomic configurations derived from classical molecular dynamics (MD) simulations, indicated the potential of such integrated approaches. By the mid-1990s, density functional methods had matured to the extent of effectively treating atomic samples of approximately 100 particles, a threshold for obtaining statistically significant macroscopic properties. This development initiated a renewed effort to employ these techniques together with semiempirical approaches to model warm, dense systems. [2]

The ensuing years have witnessed the steady improvement of the basic quantum mechanical methods, providing great accuracy in the characterization of these media. Using warm, dense aluminum plasma as a case study, [3] we have progressed from the insights provided by the initial semiempirical study in 1985 to our 2002 quantitative predictions of electrical conductivity from full QMD simulations. In direct comparison with experimental results obtained from a tamped, exploding-wire technique, we found agreement in the computed electrical conductivities for temperatures between 6000 K and 3000 K and for densities near solid to 1/100 of the solid density.

The large variety of systems and environments explored by QMD attests to the great flexibility and applicability of the approaches. Such applications include isotopic plasma mixtures of dense hydrogen, highly compressed rare-gas solids, [4] alkali metals near melt and along the vapor-liquid coexistence boundary, [5] impurity atoms in dense hydrogen plasmas, [6] shock-compressed liquids of atoms and hydrocarbons, [7,8,9] disorder in semiconductors, [10] and ultracold plasmas. [11] As a demonstration of the efficacy of these methods, we consider several representative examples.

Hydrogen

As a first example, we focus on hydrogen (H), both for its deceptive simplicity and for the considerable controversy that rages over its equation of state (EOS). From an atomic physics standpoint, hydrogen, being composed of one proton and one electron, is the simplest element known. Surprisingly enough, its phase diagram is extremely complex. For a temperature range between 10^4 K and 10^5 K and density from 0.1 g/cm^3 to 1 g/cm^3, hydrogen starts as a dense diatomic fluid. As the temperature and density increase, the fluid undergoes continuous dissociation and ionization to become a fully ionized plasma that consists solely of electrons and protons. For this regime, the challenge in obtaining a meaningful EOS lies in accurately describing the evolution of the delicate balance of atomic, molecular, and ionized species that constitute the fluid.

The complicated nature of hydrogen becomes evident in Fig. 1, which displays the electronic probability density around the nuclei for a snapshot within a QMD simulation. [7]

The EOS of hydrogen and its isotope deuterium (D) has received new attention due to recent laser experiments that seem to call into question older models. Figure 2 displays the current status of this controversy and depicts the pressure as a function of density for the shock compression of an initially cryogenically cooled molecular deuterium (D_2) sample.

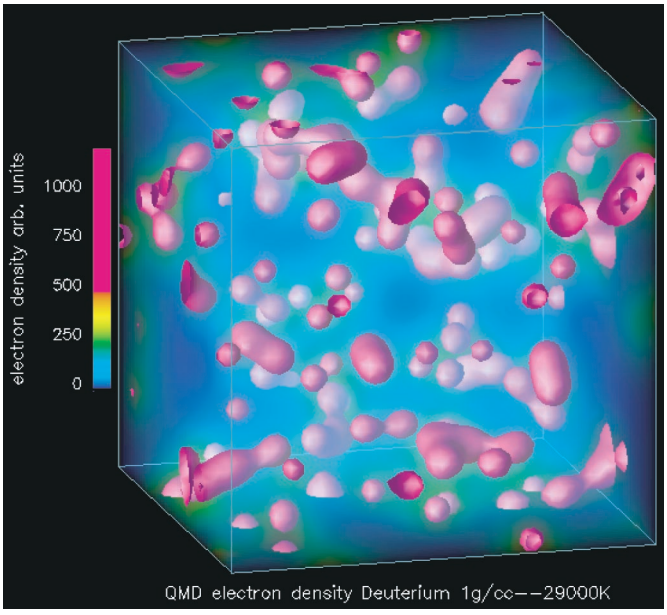


Fig. 1. Electronic probability density of hydrogen at 1 g/cc and 29,000 K.

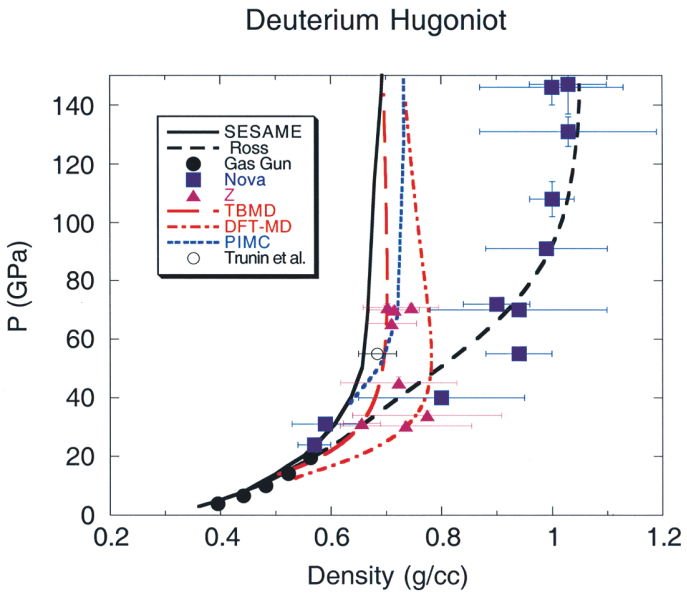


Fig. 2. Comparison of pressure as a function of compression ($\eta = \rho/\rho_0$) for deuterium along the principal Hugoniot starting from the cryogenically cooled liquid molecular state. The comparison includes various theoretical and experimental results.

Until the laser experiments, the established path, or Hugoniot, came from a chemical model in the SESAME tables compiled at Los Alamos in the 1970s and gave a maximum compression η, given by the ratio of the density ρ to the initial density of the sample ρ_0, of four. Experiments at the Nova laser facility of the Lawrence Livermore National Laboratory indicated a far more compressible medium with $\eta = 6$. This difference has profound ramifications for such diverse fields as planetary interiors and nuclear weapons.

To gain insight into this experimental disagreement, we performed QMD calculations using the simple semiempirical tight-binding method (TBMD) and the very sophisticated density function approach (DFT-MD). Our QMD values agreed much better with the SESAME results and with subsequent, similar *ab initio* calculations such as Path Integral Monte Carlo (PIMC). In 2000, a new set of experiments at the Sandia National Laboratories, employing a flier plate accelerated by a pulse-power machine (Z), produced results in close accord with the *ab initio* methods. Finally, recent Russian experiments with explosively generated converging shock waves support the Sandia findings. The final verdict on the EOS of hydrogen awaits further experimental trials; however, the rather good agreement among the *ab initio* MD simulations and the Z and Russian experiments gives a strong penchant for favoring the stiffer compressibility of hydrogen.

Opacity

Before concluding our discussion of hydrogen, we should remark upon the importance of these integrated *ab initio* approaches to needed improvements in current opacity libraries. Opacity, a measure of the absorption of radiation in matter, is an important quantity in the modeling of diverse phenomena in astrophysics and weapons design. Many of the opacity libraries commonly used for standard macroscopic modeling programs, such as in hydrodynamic codes, employ physical models that have not seen significant revision in decades.

During these years, developments in a wide variety of fields, including weapons, ICF, high-energy density, and astrophysics, have required an extension of the libraries into new and complex regimes. Such an extension necessitates a careful validation, either from experiments or from more sophisticated theoretical methods, of the physical models that produce the opacity data. Because experiments have proved difficult within these new realms, as witnessed by the continuing controversy over the EOS of compressed deuterium, *ab initio* simulation techniques provide the best venue for making meaningful critiques of these models.

As indicated, these *ab initio* approaches produce from the same simulation a consistent set of material and optical properties. In contrast, the opacity libraries consist of a collection of approximate models. Therefore, an understanding of the differences in the opacities between the libraries and *ab initio* approaches requires a detailed examination of the underlying material properties, such as EOS, and optical properties, such as the absorption coefficient. To this end, we have performed large-scale QMD simulations of hydrogen and compared representative properties with the results from standard opacity libraries, in particular, the Light Element Detailed Configuration Opacity Code (LEDCOP) from Los Alamos.

The absorption coefficient $\alpha(\rho,T,\nu)$, which gives the attenuation of radiation as a function of frequency ν (photon energy) at a given density ρ and temperature T, is the fundamental physical quantity determined by both the QMD and LEDCOP. This quantity has a direct relationship to the ac electrical conductivity of the medium; and its inverse, integrated in frequency over the derivative of the normalized Planck function, yields the ubiquitous Rosseland mean opacity κ_R.

Figure 3 shows a representative comparison between the QMD and LEDCOP for the hydrogen absorption coefficient as a function of photon energy. For this regime, in which the medium becomes fully dissociated and partially ionized, the two methods show rather good agreement for the region giving the largest contribution to k_R, marked by the green bar, and for larger photon energies. The disagreement at low photon energies arises from the representations of the fluid. LEDCOP indicates a weakly interacting atomic system in which the free-free absorption dominates. However, the QMD result correctly goes to the limit at zero frequency of the dc conductivity, a condition that indicates a strongly correlated fluid in which density effects still play a predominant role. For lower densities and temperatures, the differences in the mean opacity become greater between the two approaches. This example demonstrates that the QMD approach offers considerable opportunities to validate and improve current opacity libraries commonly used in many macroscopic modeling programs of the nuclear weapons program.

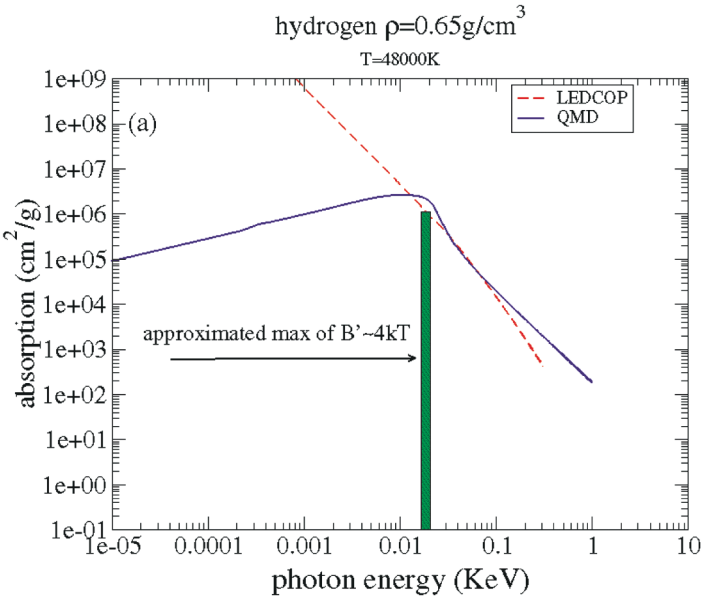


Fig. 3. Comparison of the absorption coefficient as a function of photon energy between QMD and LEDCOP for hydrogen.

Nitrogen and Oxygen

As a second example [8] of the broad applicability of these approaches, we present comparisons of QMD simulations for compressed nitrogen and oxygen, which have certain similarities to hydrogen. They all have molecular liquid states at very low temperatures, large dissociation energies, and moderate ionization potentials. For nitrogen and oxygen, though, several gas-gun experiments have probed a larger part of the Hugoniot than for hydrogen.

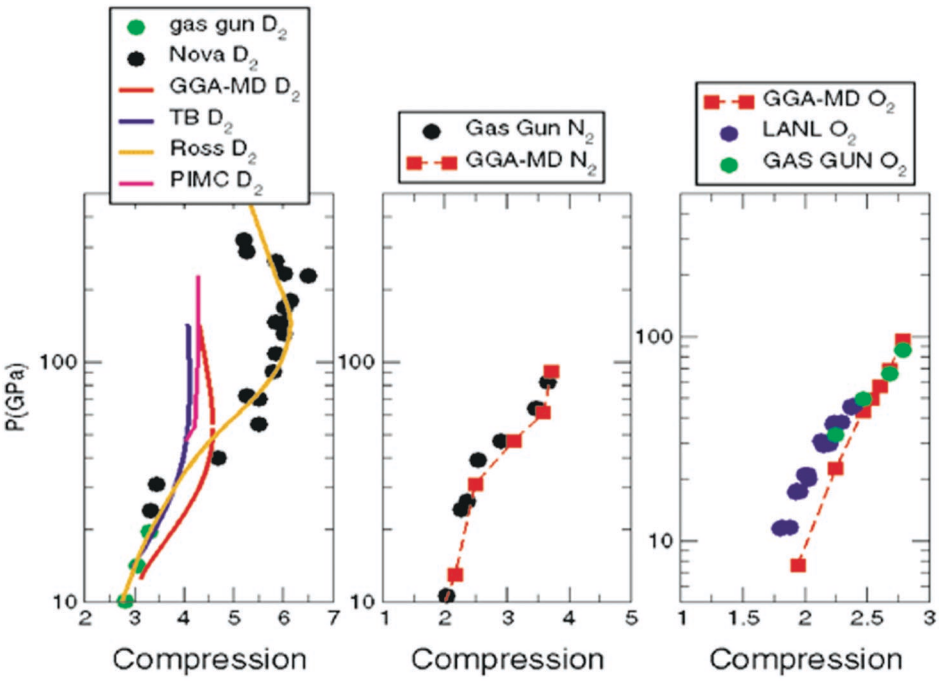


Fig. 4. Pressure as a function of density along the principal Hugoniot for hydrogen, nitrogen, and oxygen. The comparison includes QMD results and gas-gun experiments.

Figure 4 displays the excellent agreement obtained between the QMD simulations and the experiments for nitrogen and oxygen along the principal Hugoniot. This indicates that the QMD approach can accurately characterize the progress of a very complex medium through many different stages.

Hydrocarbons

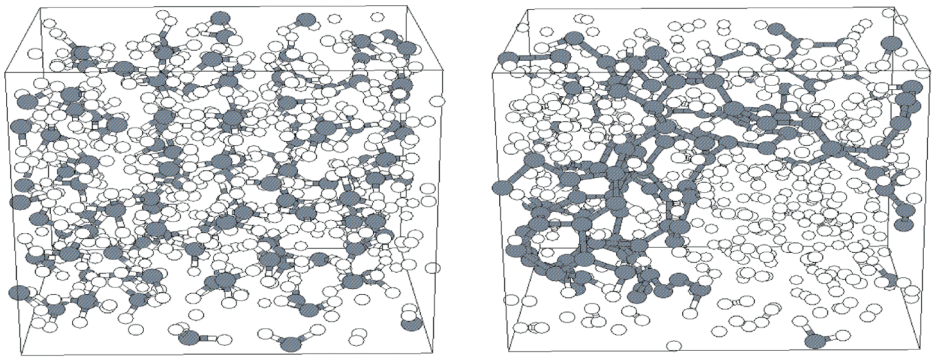
Finally, as a last example, we describe semiempirical TBMD simulations of shocked hydrocarbon fluids. Hydrocarbons are found in various extreme environments such as the atmosphere of Jupiter, and methane is an important constituent of the middle ice layers of the outer gas giants, Uranus and Neptune. Determining the transformations induced in hydrocarbons by nonequilibrium phenomena, such as shock waves and detonations, presents a challenging problem. The accepted interpretation of shock-wave experiments holds that these materials ultimately decompose into elemental carbon (probably amorphous carbon and diamond powder), molecular hydrogen, and perhaps other gases.

There is clearly interest in determining what chemical transformations occur on the intermediate time scales, as well as examining the formation of transient species under nonequilibrium conditions. TBMD simulations on dense methane at a specific volume of 10.04 cm^3/mol were carried out for temperatures between 1,000 K and 8,000 K. Between 4,000 K and 5,000 K, a change in slope was observed in the plot of pressure as compared with the temperature. This slope change is indicative of a phase change or reaction.

In Fig. 5, we present snapshots from the simulation at a temperature below (2,000 K) and above (6,000 K) the change in slope point. At 2,000 K (Fig. 5a), no reaction has occurred because the methane (CH_4) molecules are all still intact; and no carbon-carbon bonds or hydrogen molecules have been formed. At 6,000 K (Fig. 5b), the situation has drastically changed because the methane molecules have dissociated and reacted to form hydrogen molecules (hydrogen-hydrogen bonds) and carbon-containing clusters (indicated by the formation of carbon-carbon bonds). Such a drastic change in molecular bonding is nearly impossible to capture using traditional classical interatomic potentials. Thus a quantum mechanical description of the electronic structure is needed.

Summary

In summary, QMD simulations have proved an effective and versatile theoretical and computational approach to treating a large variety of warm, dense systems of particular interest to a broad number of research programs by providing a systematic, integrated technique for probing matter under extreme conditions.



(a) (b)

Fig. 5. Snapshots from a simulation of dense methane. The black balls are carbon atoms and the white ones are oxygen. The temperatures are 2,000 K (5a) and 6,000 K (5b).

In addition to the authors, many people have made direct contributions to this continuing project. Those closely associated with Los Alamos, either as employees or collaborators, include Al Merts, Norm Troullier, Diane Lynch, Inhee Kwon, Scott Bickham, Tom Lenosky, Barry Schneider, Norm Magee, and Steve Younger.

Modeling and Molecular Theory of Liquids

by Lawrence R. Pratt

The condensed phase that falls on your head many mornings each week is a liquid. Los Alamos Theoretical Division researchers have made decisive contributions to the theory and modeling of liquids throughout the lifetime of the Laboratory. Here I note some of those highlights (for example, the development of quasichemical theories that resolved ion-hydration problems), give personal views of how the theory of liquids fits into a big picture of scientific problems (aids molecular biophysics, nanotechnology, planetary research, chemical engineering, and physics), and give a brief personal indication of current research.

The theory of liquids is predominately physics that is done principally in chemistry groups. The reason is that liquids are the condensed phase most used for chemical purposes. Liquids permit simple concentration, transport, and mixing of chemicals.

Bethe
Guggenheim
Kirkwood
Metropolis
Pratt
M. Rosenbluth
E. Teller
Thiele
Wertheim
B. Wood

Often regarded as more difficult than the theory of either gases or solids, the molecular theory of liquids combines the difficulties of those two other research areas, each already difficult enough. Liquids are both dense *and* disordered. This "jointly difficult" characteristic is exhibited again by the requirements of chemical fidelity for the molecular constituents and of physical scope in describing the supermolecular laws that emerge. The molecular theory of liquids is a primordial multiscale problem.

Beginning in the 1950s, Laboratory scientists established an important new foothold on the problems of liquids by the development of particle simulation methods of Monte Carlo and molecular dynamics.

Looking Backward

The story of a first epoch in the theory of liquids at Los Alamos has been charmingly recounted by Bill Wood. [1,2] Goals of that initial work were equations of state (EOSs) for detonation products. Simulation efforts to treat simple models of liquids were enabled by Monte Carlo methods and the Metropolis algorithm. The first efforts (1953) using those ideas were due to Metropolis, Rosenbluth, and Teller, in combinations. The work ongoing from there clearly captured the active attention of J. G. Kirkwood, [1,2] who is regarded in many circles as the father of the molecular theory of liquids. By 1957, these simulation techniques had been firmly established. Helpful in convincing the broader scientific community of the validity of simulation was the successful cross-checking of alternative molecular dynamics calculations carried-out at Lawrence Livermore National Laboratory. [1,2]

Quantitative results for simple-model liquids can also spur classic theories. Proposed in 1957, the Percus-Yevick theory was solved analytically for the hard sphere case in 1963, independently by Wertheim and Thiele. [3,4] Though intimately familiar with T-Division research, both Wood and Wertheim were associated with different organizations when the historic works noted here were carried out. Wood worked in what became the Los Alamos Design Engineering Division, and Wertheim was at the Courant Institute of Mathematical Sciences on leave from T-Division.

Looking Forward

Subsequent work has included more complicated liquids. *Ab initio* molecular dynamics, in which interatomic forces are obtained from an "on-the-fly" solution of a relevant Schrödinger equation, is an important more recent development. But which real liquids present important research issues and why? Here I sketch a personal response to such questions. My personal estimate of rank, from more to less important, is as follows.

Water and Aqueous Solutions. This ranking is due to the ubiquity of water in our physical world and our culture. Water has a high-curiosity value because it is peculiar among liquids and often participates in chemical reactions. Molecular biophysics (aqueous) and nanotechnology are closely related because molecular biophysics offers numerous examples of sophisticated, molecular-scale mechanical, electrical, and optical machines. The example of selective ion channels is discussed in more detail subsequently. Water participates directly in the structure and function of these machines. And even though these machines are "self-assembled," the issues of concentration, transport, and mixing, already mentioned, are still fundamental. Liquid ammonia, water's cousin and perhaps relevant to questions of life on the moons of Jupiter, is often associated with aqueous solutions and so should be put in this category.

Petroleum-Derived Liquids. This category includes the vast majority of solvents used in chemical processes, most nonbiological polymeric solutions, and most nonbiological liquid crystal systems. Polymers, liquid crystals, and nanotechnology overlap the previous item. One view is that the chemical engineering profession, historically dominated by the science of petroleum fluids, will in the future significantly include bio- and nanotechnology problems. This view reveals the perspective that chemical engineers are the sophisticated audience for good theories of liquids.

Plasmas, Earth's Interior, the Sun, High-Density Reacting Fluids. In these cases, "fluid" typically seems a more relevant appellation than "liquid." Fluid metals are typically legitimately liquids, and chemically complex "room-temperature ionic liquids" (low-temperature molten salts) are of high current interest.

Quantum Liquids, Particularly Helium, the Gas Giant Planets. Both this and the previous item have high-curiosity values. These categories are associated with a physics subset of the theory of liquids.

Simple Liquids. First and foremost, this has meant liquid argon. But other liquified simple gases, such as neon (Ne), nitrogen (N_2), oxygen (O_2), and methane (CH_4), occupy this category too. Work on this category is foundational: to establish that the simplest cases can, in fact, be well solved.

Current Example

This catalogue identifies a local research emphasis: chemically complex aqueous liquids, biophysics, and nanotechnology. The currently *hot* problems of biological ion channels combine all these elements.

Biological ion channels are membrane proteins that govern ion conduction into and out of nerve cells. They can be gated, and they can discriminate between chemically similar ions for transmission, e.g., select potassium $K^+(aq)$ over sodium $Na^+(aq)$.

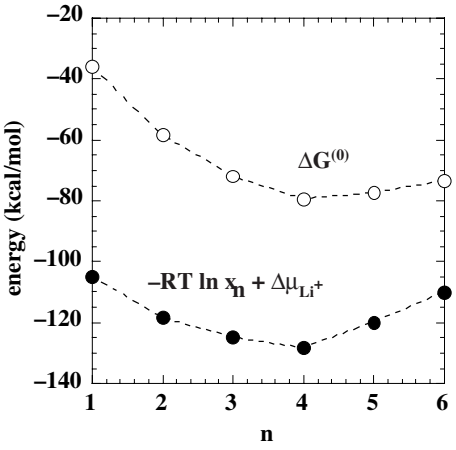


Fig. 1. (lower curve) Quasichemical approximate results for lithium $Li^+(aq)$. x_n is the fraction of Li^+ ions with *n* inner-shell water-molecule neighbors in liquid water under standard conditions. This work concluded that the most probable inner sphere hydration number was four (n = 4) and clarified inconsistent neutron-scattering results on these solutions. (upper curve) Electronic structure results without consideration of the liquid environment. [5] **(See explanation on next page.)**

In working on related problems, we developed a *quasichemical* theory that permits the use of modern computational chemistry. [6] This theory has roots in 1930's problems of cooperative phenomena in crystals. For the lattice models studied there, the present theory is related to the works of Guggenheim and Bethe; "quasichemical" is due to Guggenheim and the simplest instances would be the "Bethe approximation" for many physicists. But conventional lattice models are inappropriate as molecular models of liquids. The advance was to see how the same physical ideas offered practical theories of molecular liquids without the lattices.

Initial applications of these quasichemical theories were to ion hydration problems as suggested by Fig. 1. The interactions of near-neighbor water molecules with metal ions involve chemical features. And these hydration structures play a role in the selectivity of ion channels: partial dehydration of an ion will sometimes be required for successful invasion of a channel.

Accelerated Molecular Dynamics Methods

by Arthur F. Voter

Germann
Hamilton
Kim
Mazevet
Montalenti
Rudd
Sholl
Sørensen
Sprague
Stuart
Swaminarayan
Uberuaga
Voter
Windl

The molecular dynamics (MD) simulation method, pioneered in part at Los Alamos in Theoretical Division, is now well established as a powerful tool for studying atomic-level behavior in chemistry, physics, materials science, and biology.

For a given interatomic potential, atom positions are evolved with no other assumptions than classical mechanics, so the dynamical behavior in a simulation is "truth" for the system under study. Direct and meaningful comparisons to experiment can be made, with discrepancies attributable solely to the inaccuracy of the potential. A major drawback of molecular dynamics, however, is that it is limited to time scales of nanoseconds or less. Even massively parallel computers, which have extended the size scale to millions of atoms, have had little impact on the accessible time scale. Theoretical Division scientists have developed accelerated molecular dynamics methods that can solve problems that are otherwise inaccessible to MD simulation.

For many systems, the long-time dynamics can be characterized as "infrequent event" behavior (See Fig. 1.): there are occasional transitions from one potential basin to another, separated by long periods of uninteresting vibrational motion. For these systems, the time scale limitations of molecular dynamics can be overcome using transition state theory (TST), in which the rate constant for escape from one state (basin) to another is approximated as the flux through the dividing

surface separating the two states. The appeal of TST is that dynamics are unnecessary; the TST rate is an equilibrium property of the system and can be evaluated directly once the dividing surface is defined. Although TST is not exact (due to correlations among the surface-crossing events), it is usually an excellent approximation for solid-state defect dynamics.

As powerful as TST is, however, there is a catch: to use it requires knowledge of the dividing surfaces. If the states to which a system might evolve are not known in advance, TST cannot be applied (or may be applied incompletely, missing important pathways). Many such situations occur in materials science. Examples include the healing events after ion implantation, the dynamics of a dislocation core, the advance of a crack front, or diffusion events during the growth of a film.

In an effort initiated in 1995 under Laboratory-Directed Research and Development (LDRD) funding and currently funded primarily by the U.S. Department of Energy Office of Basic Energy Sciences (DOE/BES), a group of us have been exploring an alternative approach to these types of problems. The team has included staff members (Tim Germann, Sriram Swaminarayan, and I), a number of postdoctoral researchers and fellows (Mads Sørensen, Francesco Montalenti, Stephane Mazevet, Blas Uberuaga), and a growing number of visitors and collaborators (Walter Rudd, James Sprague, Steve Stuart, Wolfgang Windl, Jeongnim Kim, John Hamilton, David Sholl). We are developing a new class of methods, largely based on TST, that advance the dynamics of complex infrequent-event systems on time scales that are currently inaccessible (and may never be accessible) to direct MD simulation. We term this approach accelerated molecular dynamics. We briefly describe three methods in this class and give a couple of examples of the power of this emerging approach. A recent review can be found in Ref. [1].

The key concept underpinning the accelerated dynamics methods is, in a phrase, that trajectories are smarter than we are. While previous approaches for reaching long time scales have involved applying physical or chemical intuition to specify the states the system might evolve to, in fact the system often does something unexpected, something outside our intuition. For example, only recently was it realized that diffusion on metal surfaces typically involves complicated many-atom mechanisms. As described in Fig. 1, a way around this problem is to simply let the trajectory find the escape path for us. Each of the three accelerated dynamics methods described here is built on this conceptual foundation.

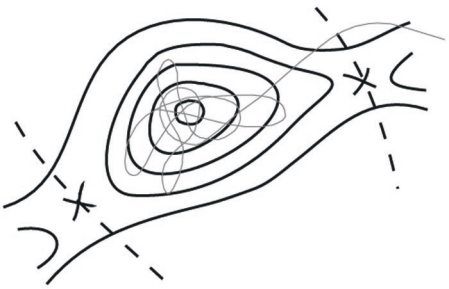


Fig. 1. Schematic illustration of an infrequent-event system. The trajectory wanders around in the basin over a time scale of many (perhaps a huge number of) vibrational periods. At some point in time, when enough energy has been localized into a reactive mode, the trajectory passes through a dividing surface, entering another state. In essence, it "accidentally" exits the state. During this brief period of excitation, it may recross the dividing surface, but shortly thereafter it settles into the new state (or the original state if it recrosses), beginning a new session of vibrational wandering, with no memory of how it arrived in that state. Although it may never again visit this state and "sees" only a single dividing surface as it exits, it nonetheless chooses an escape direction (relative to other possible escape directions) with the correct probability. This property of any infrequent-event system, that a trajectory will automatically choose an appropriate escape path with no prior information, is the basic concept exploited in the accelerated dynamics methods. The key is to coax the trajectory into making this choice more quickly without corrupting the relative escape probabilities (i.e., the rate constants for escape to various states).

The Hyperdynamics Method

In the hyperdynamics method, [2] a bias potential is designed to raise the energy of the system in regions other than at the TST dividing surfaces, as shown schematically in Fig. 2. Running a simulation on the biased potential gives accelerated evolution from state to state because

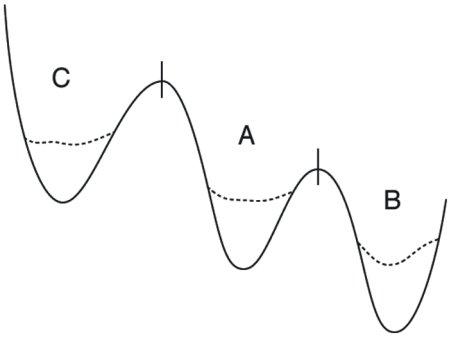


Fig. 2. Schematic illustration of the hyperdynamics method. A bias potential is added to the original potential (solid line). Provided that the bias potential meets certain conditions, primarily that it be zero at the dividing surfaces between states, a trajectory on the biased potential surface (dashed line) escapes more rapidly from each state without corrupting the relative escape probabilities. The accelerated time is estimated as the simulation proceeds.

the basins are not as deep. The elapsed time becomes a statistical property of the system, with the instantaneous gain (or "boost") in the rate at which time advances (relative to direct MD) depending exponentially on the bias potential strength. In the long-time limit, the dynamics on the biased potential are exact, provided that TST holds and the bias potential meets certain requirements. In essence, this method ("hyperdynamics" or "hyper-MD") captures the advantages of both molecular dynamics and transition state theory into a single simulation approach.

The key to hyperdynamics is designing a computationally tractable bias potential that is zero at the dividing surfaces (without explicit knowledge of them) and does not introduce correlated dynamical events. We have shown that such a bias potential can be constructed from local properties of the potential by means of the Hessian and gradient. [3] Moreover, the derivative of the bias potential can be computed using only first derivatives of the original potential, just as for regular molecular dynamics. Other forms of bias potential have also been suggested, [1] and the design of optimal bias potentials will likely remain a research topic for many years.

Parallel-Replica Dynamics

The next accelerated dynamics method, parallel-replica dynamics, [4] harnesses the power of parallel computing to achieve longer time scales. In principle, the large-scale parallel MD methods (See "Molecular-Dynamics Simulations" by Holian, Germann, and Lomdahl in this "Condensed Matter " chapter.) could be applied to a small system (e.g., 1,000 atoms) to reach long times, but in practice this is severely limited by the high cost of interprocessor communication between the too-small spatial domains. Those methods become efficient only when there are thousands of atoms per processor. For an infrequent-even system, however, we have discovered there is a simple alternative. As shown in Fig. 3, the procedure consists of replicating the entire system on each of M processors and running an independent trajectory on each one. When a transition is detected on any processor (e.g., "processor i"), all processors are stopped and all the trajectory times are summed. The new state found by "processor i" is then replicated on all processors and the procedure is begun again. Parallel replica dynamics gives the exact state-to-state dynamics for any system obeying Poisson statistics (i.e., that the probability distribution function for the time until the next transitions, after a suitable correlation time has passed after entering each new state, is an exponential). As such, we expect it may be useful in more general situations than just atomistic systems. [1]

Figure 4 shows an example of a parallel-replica simulation. An Ag(111) island-on-island structure decays over a period of 1 microsecond at T = 400 K. [1] Many of the transitions involve concerted mechanisms. It is interesting to note that the boost on just 32 processors exceeds what we could currently obtain from either hyperdynamics or temperature-accelerated dynamics, because of the relatively high temperature

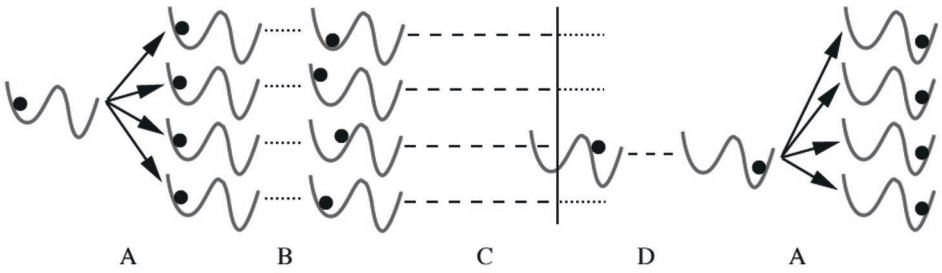


Fig. 3. Schematic illustration of the parallel replica method. The four steps are (A) replication of the system into M copies, (B) dephasing of the replicas, (C) independent trajectories until a transition is detected in any of the replicas, and (D) brief continuation of the transitioning trajectory to allow for correlated events, such as recrossings or follow-on transitions to other states. The resulting configuration is then replicated, beginning the process again.

(400 K) of this simulation. While these other two methods offer dramatically increasing boosts as the temperature is reduced, parallel-replica dynamics provides a consistently predictable acceleration unless events are so frequent that the dephasing overhead (See Fig. 3.) dominates.

Temperature-Accelerated Dynamics

Last, we describe the temperature-accelerated dynamics (TAD) method. [5] It has been long understood that raising the temperature of a system causes the activated events to occur more frequently. However, because raising the temperature changes the relative rate constants of the different events, this is not a useful approach for accelerating the overall dynamical evolution, except in very simple cases. For example, raising the temperature to accelerate surface diffusion during a simulation of film growth can change the resulting film morphology, which may be the very property under study.

In the TAD method, we harness the acceleration that comes from raising the temperature, while remaining faithful to the state-to-state evolution that would take place at the normal temperature. Making a few reasonable approximations, we have derived a way to "filter" the transitions from basin to basin, allowing only those transitions that would have occurred at the normal temperature. We run the high-temperature simulation in such a way that each attempted transition is detected, but not allowed to proceed. In this "basin-constrained" simulation, the transition state for each attempted event is found, using modern saddle-point search methods, after which the trajectory is reversed and continued in the original state. In this way, a survey of the possible transitions is accumulated. After a certain length of time, this state-constrained trajectory can be terminated, and the "correct" transition for the normal temperature can be selected.

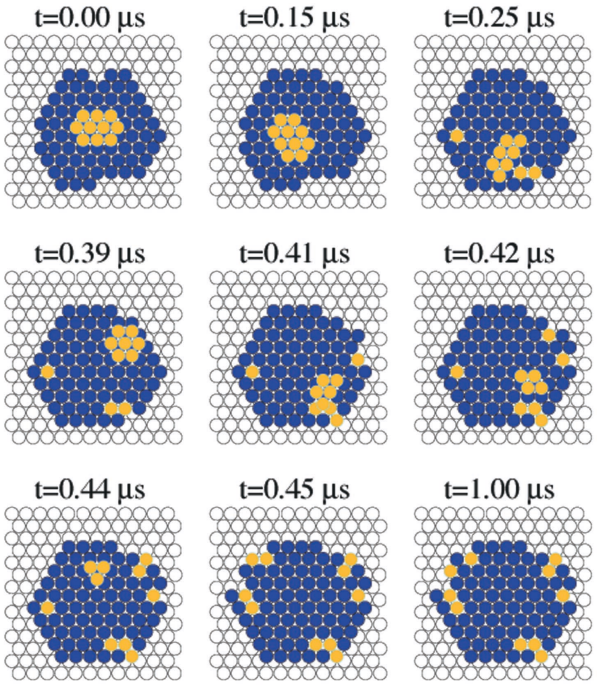


Fig. 4. Nine snapshots from a parallel-replica simulation of an island on top of an island on the Ag(111) surface at T = 400 K. On a microsecond time scale, the upper island gives up all of its atoms to the lower island, filling vacancies and kink sites as it does so. This simulation took 5 days to reach one microsecond on 32 1-GHz Pentium III processors.

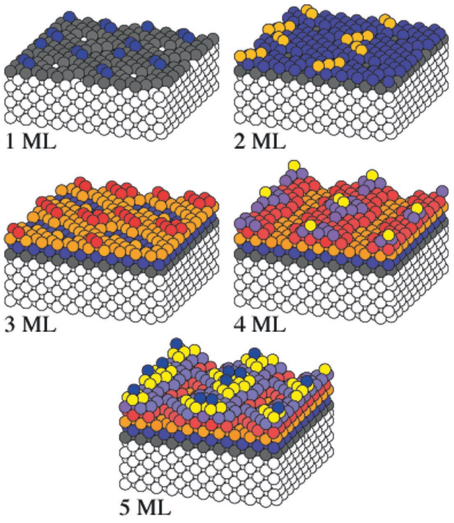


Fig. 5. Snapshots from a TAD simulation of the deposition of 5 monolayers (ML) of Cu onto Cu(100) at 0.067 ML/s and T = 77 K, matching the experimental conditions of Egelhoff and Jacob. [6] Deposition of each new atom was performed using direct molecular dynamics for 2 ps, while the intervening time (0.3 s on average for this 50-atom/layer simulation cell) was simulated using the TAD method. This simulation would take ~107 yr using direct MD.

With a prechosen confidence level (e.g., 99.9%) one can say that no other transition that would occur after that time could replace the chosen transition. The system is then moved to the basin corresponding to the chosen transition, and the TAD simulation clock is advanced by the exact amount that a normal-temperature trajectory would have accumulated before making this transition. The procedure is then started again. As with the hyperdynamics method, [1] the "boost factor" for this approach (the simulation speedup relative to a direct MD simulation) is system dependent and increases (sometimes dramatically) as the normal temperature is lowered. In favorable cases the results are impressive.

One of the fields where accelerated MD is desperately needed is vapor-deposited crystal growth, where the typical time scale can exceed minutes. Figure 5 shows an example of the power of the TAD method for this problem. We simulate vapor-deposited growth of a Cu(100) surface at a deposition rate of one monolayer per 15 s and a temperature T = 77 K, exactly matching (except for the system size) the experimental conditions of a well-known experiment. [6] Each deposition event is simulated using direct molecular dynamics for 2 ps, long enough for the atom to collide with the surface and settle into a binding site. A TAD simulation with a high temperature of 550 K then propagates the system for the remaining time (0.3 s on average for this system size and deposition rate) until the next deposition event is required. The overall boost factor is ~10^7, although it is not as large at higher temperatures. Even at this low temperature, many events accepted during the growth process involve concerted mechanisms. A few examples of concerted mechanisms observed during growth at T = 77 K and T = 100 K are shown in Fig. 6.

These accelerated dynamics methods are still young, and we expect them to continue to develop. Nonetheless, this approach is already showing great potential for solving problems that are otherwise inaccessible to molecular dynamics simulation.

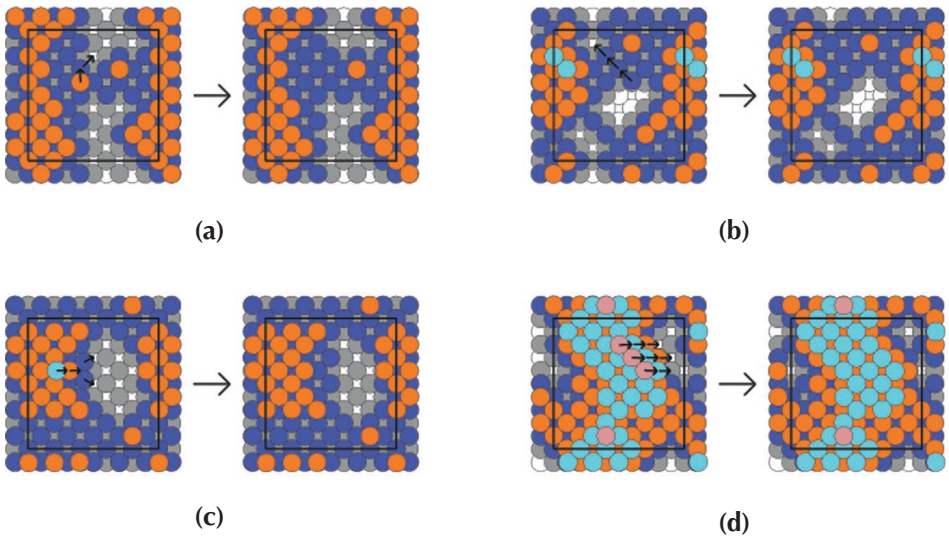


Fig. 6. Examples of concerted mechanisms observed during TAD simulations of Cu/Cu(100) epitaxial growth, such as those shown in Fig. 5. Atoms are deposited at an average flux of one monolayer per 15 s, corresponding to the experimental rate of Egelhoff and Jacob. [6] (a) A two-atom downward exchange, with activation barrier Ea = 0.255 eV, that occurred after 3 ms at T = 100 K. (b) Three-atom row sliding, Ea = 0.241 eV (28.3 ms at 100 K). (c) Downward exchange pushing two additional atoms in a metastable off-lattice site, Ea = 0.056 eV (88.9 ps at 100 K). (d) Sliding of an eight-atom cluster down a {111} facet, Ea = 0.046 eV (45.2 ns at 77 K).

Unifying the Quantum Mechanical Description of Matter

by Cristian D. Batista and Gerardo Ortiz

In Theoretical Division, we developed a unified framework to study and unveil order in the quantum description of matter. The concepts of* language, dictionary, *and resulting* universality *are central to our framework. We showed how to connect different (spin-particle-gauge) languages by means of dictionaries (exact mappings) and proved a fundamental theorem that establishes when two arbitrary languages can be connected. In this way, we uncover the hidden unity behind seemingly unrelated physical phenomena, thus establishing exact connections between them.

As science progresses, the diverse phenomena encountered in nature tend to be successfully described with fewer underlying principles. Indeed, the search for the unifying principles behind the fundamental laws of physics is a common theme and has a very simple reason, which is simplifying the understanding of the universe in which we live. Even if we knew the ultimate laws that govern the universe, could one predict the complex behavior observed in nature? This has been the subject of numerous works by very eminent people, like Philip W. Anderson, who rightfully argued that the whole is not necessarily the sum of its parts and thus "more is different."

Of particular relevance is the description and prediction of complex phenomena occurring in matter. Matter with its many phases and behaviors is full of unrelenting surprises. For instance, the last two decades have witnessed the discovery of fractional charges and skyrmion excitations in quantum Hall liquids (i.e., electrons confined to two space dimensions in the presence of strong external magnetic fields). Another example of current interest is provided by materials displaying a multiplicity of competing and coexisting complex quantum phases, like some heavy fermion compounds or high-temperature superconductors. What all these examples have in common are their extreme quantum and collective behavior and the presence of strong correlations, involving nonlinear couplings, among its constituents.

Describing the structure and behavior of matter entails studying systems of interacting quantum constituents (bosons, fermions, spins). In the quantum-mechanical description of matter, each physical system is naturally associated with a *language* (for example, quantum spin-1/2) and thus to a set of operators realizing this language (e.g., the Pauli spin algebra generated by a family of commuting quantum spin-1/2 operators). We think that crucial to the successful understanding of the mechanisms driving complex behavior is the realization of *dictionaries* (isomorphisms) connecting the different languages of nature and therefore linking seemingly unrelated physical phenomena.

The existence of dictionaries provides not only a common ground to explore complexity but leads naturally to the fundamental concept of *universality*, meaning that different physical systems show the same behavior. In this way, there is a concept of physical equivalence hidden in these dictionaries.

In 1928, Jordan and Wigner (JW) related the spin quantum mechanical degree of freedom to spinless particles with fermion statistics. We have recently generalized this one-to-one mapping to any spin, spatial dimension of the lattice, and particle exchange statistics. [1] These generalized JW mappings constitute a quantum version of the well-known classical spin-lattice-gas transformations. This result motivated us to raise the following question: "Can we connect the different (spin-particle-gauge) languages of nature within a single algebraic framework?" The answer to this question, with the conditions stated by the following theorem, is *YES*. Basically, the theorem states (See Fig. 1.):

> *Fundamental theorem* ("On the equivalence of languages"). Given two arbitrary languages having the same finite dimension D of their local Hilbert spaces, H_i, the generators of one of them can be written as a polynomial function of the generators of the other language and vice versa.

The plethora of dictionaries, or connections between different languages established by this theorem, not only allows one to relate seemingly unrelated phenomena, but also opens the possibility to develop new approaches (exact or approximate) to solve open problems in physics. [2–5]

Within the class of possible languages that can be used to describe a given problem, there are those whose generators are directly related to symmetry operations. (They are generators of a Lie algebra.) These special languages (dubbed *hierarchical*) are very useful to classify possible local-order parameters (related to broken symmetries) that may occur in the system considered. In this way, the use of adequate languages, based on symmetry generators, helps not only to predict new possible states of matter but also to determine the quantum phase diagram of models of strongly correlated matter.

These dictionaries also have important applications in the field of quantum information theory. [6] A key fundamental concept in information theory is the realization that a model of computation is intimately connected to a physical system through a closed-operator algebra. In other words, each physical system is associated to a certain language (e.g., spin S = 1/2) and thus to an algebra realizing it (e.g., Pauli algebra), and that particular algebra may become a possible model of computation. An immediate consequence is that an arbitrary physical system can be simulated by another physical system (e.g., a quantum computer) whenever there exists an isomorphic mapping between the different operator algebras representing the systems.

In summary, there are several reasons for our algebraic framework to constitute a powerful method to study complex phenomena in matter. It can be used to

(1) connect seemingly unrelated physical phenomena (e.g., high-temperature superconductors, heavy fermions, or electronic ferroelectricity, [7] and quantum spin theories);

(2) identify the general symmetry principles behind complex phase diagrams;

(3) unveil hidden symmetries (and associated order parameters) to explore new states of matter;

(4) obtain exact solutions of relevant physical models that display complex ordering at certain points in Hamiltonian space; and

(5) find new approximations that do not privilege any of the competing interactions.

The power of the present approach is reflected in the unlimited number of potential applications where they can be used that range from condensed matter and statistical mechanics to lattice gauge theories and quantum information and computation. For a comprehensive review, see Ref. [5].

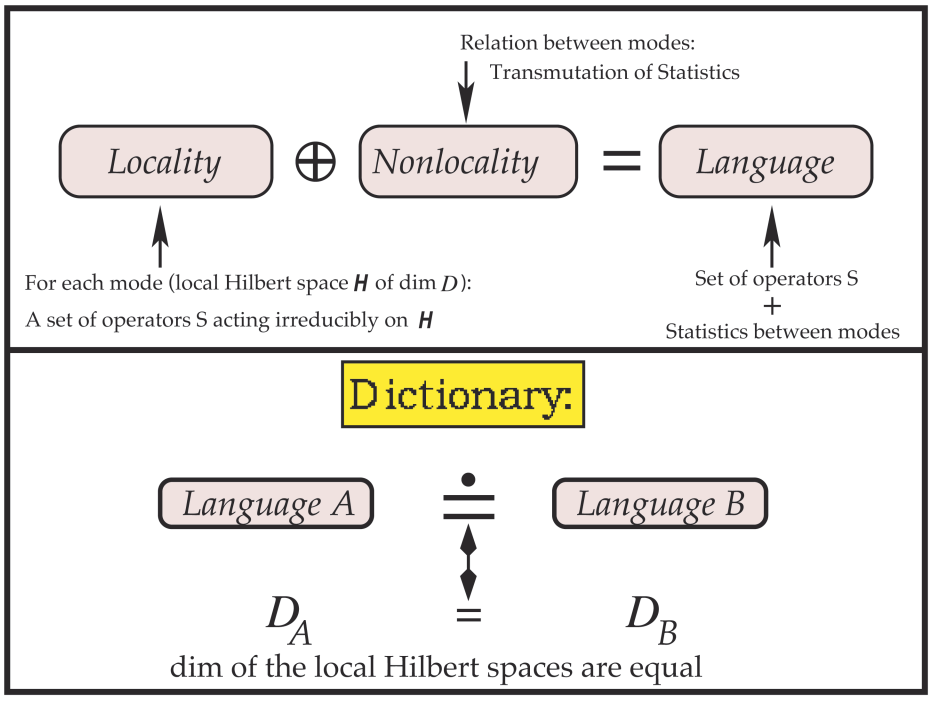
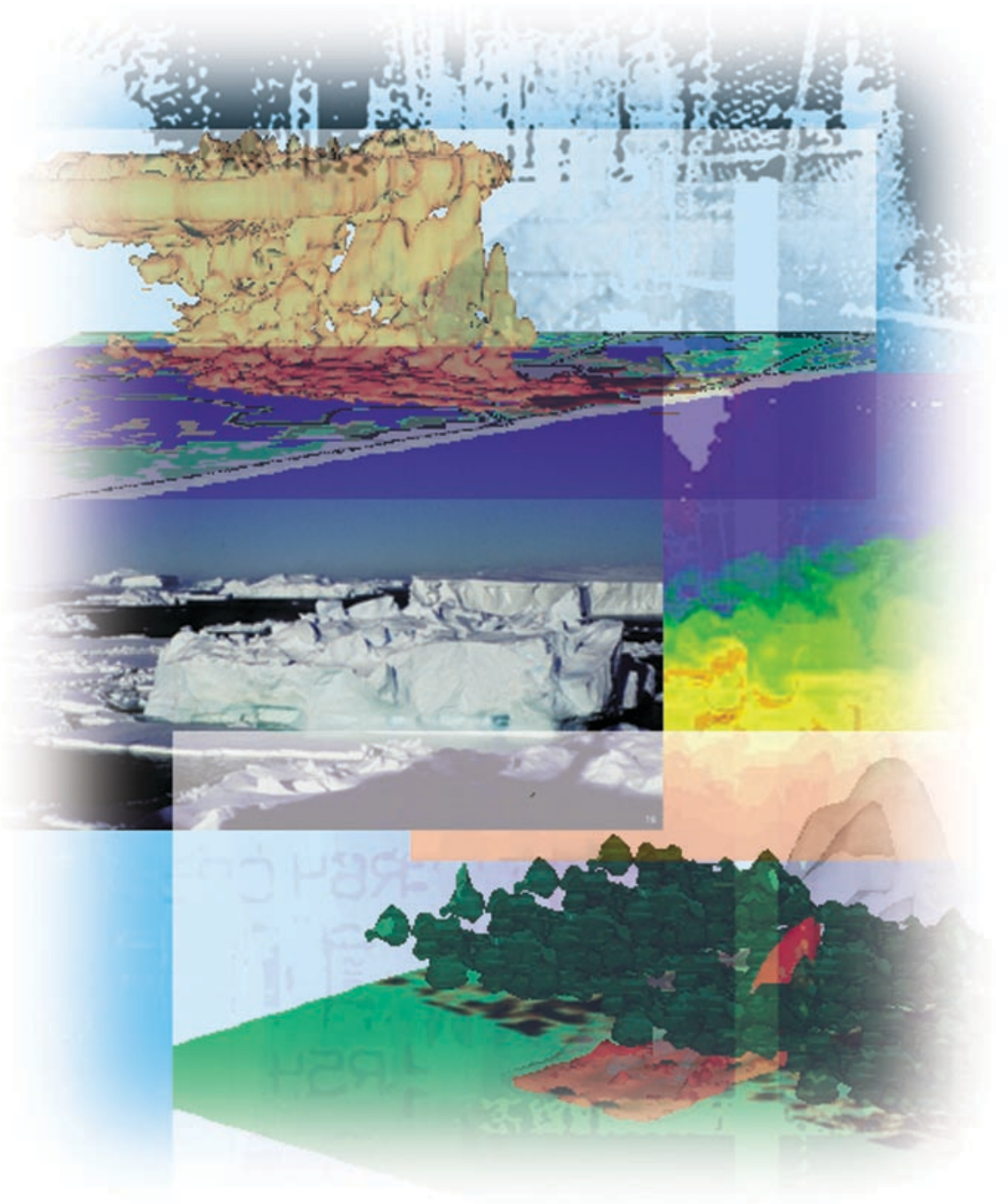


Fig. 1. What is a language, and when can two languages be connected? The figure is a summary of the main content of our fundamental theorem in conjunction with the concept of transmutation of statistics.



Environmental Studies

Problems created by both local and global modifications to the Earth greatly concern citizens of the United States and, indeed, people worldwide. The Los Alamos Theoretical Division has long been active in environmental studies.

In this "Environmental Studies" chapter, we especially highlight issues of ocean and atmosphere dynamics on both short-range and long-range time scales. T-Division also has given considerable study to pollutants (both chemical and radioactive), including transport and mitigation processes.

We have also used a coalescence of chemistry, turbulence transport, and atmospheric dynamics to analyze the complex processes that occur in a wildfire; and we have translated the knowledge we gained into a three-dimensional computer code that describes fire propagation through vegetation in any type of terrain.

Atmospheric "scrubbing" for the removal of carbon dioxide is another important process that was investigated in T-Division.

In this "Environmental Studies" chapter, we describe a sampling of these activities; and, in the last section of this book, we indicate plans for long-term T-Division involvement in solving environmental problems.

One of the reasons I didn't like the S-site was that to get the raw explosives there they had to be trucked all the way over the mesa, right through the center of the Los Alamos project, with the whole theoretical division sitting in offices on one side of the road, Oppenheimer's office (and mine) on the other. . . . I can assure you that a truck loaded with five tons of high explosive (H.E.) going off there would have wiped out 90% of the brains in those temporary buildings.

—George B. Kistiakowsky [chief or the Explosives Division in World War II Los Alamos], "Reminiscences of Wartime Los Alamos," in *Reminiscences of Los Alamos: 1943–1945*, p. 51.

Accelerator Transmutation of Waste Effort

by Edward D. Arthur

Transmutation is the modern equivalent of alchemy. However, instead of turning lead into gold, transmutation can be aimed at transforming long-lived components of nuclear waste into short-lived or stable byproducts. These components—plutonium and other actinides (neptunium, americium, and curium) and certain products from fission (technetium, iodine)—are the drivers behind long-term hazards associated with nuclear waste produced by nuclear energy generation. Transmutation occurs through the use of nuclear processes to change one isotope or element into another. It occurs naturally (during the decay of radioactive nuclei) or in nuclear reactors. (Fission is a form of transmutation.) Theoretical Division researchers have worked in transmutation since the early 1990s and are developing key technologies to eliminate nuclear waste.

In the early 1990s, Los Alamos National Laboratory researchers (including those in Theoretical Division) recognized the role that transmutation technology could play in dealing with long-term issues associated with nuclear waste. Although transmutation's applicability to nuclear waste had been proposed in the 1960s, significant progress in areas such as high-power accelerator technology was a major factor in reawakening interest in transmutation. Specifically, high-power accelerators could be used to generate the neutrons necessary for transmutation to occur.

In contrast to reactors (which also generate copious amounts of neutrons), accelerator-driven systems could do so in "subcritical" environments. Fission reactions in such subcritical systems would cease once the beam of the accelerator that drove the nuclear assembly was turned off. Reactors, on the other hand, operate in a self-generating or "critical" configuration so that other means must be used for their control. The subcritical operational mode of accelerator-driven transmutation systems was seen as a significant safety advantage when sizable amounts of exotic actinides were present in a transmutation system.

In addition to safety enhancements, accelerator-based systems brought other advantages over reactors, specifically when low-energy, or thermal, neutrons were used to transmute actinides and fission products. Thermal neutrons have high nuclear cross sections for transmutation, thereby requiring smaller inventories of materials to be present in a transmutation system. At the same time, certain nuclear properties associated with the use of thermal neutrons are less advantageous so that the accelerator-produced neutron source was necessary for efficient system operation.

Thus, in the early 1990s, the Laboratory created the Accelerator Transmutation of Waste initiative (ATW). T-Division played a major role in the initiation of the program, in terms of both defining the system concept and providing key technical information. An accelerator-based transmutation system requires nuclear data at energies and for nuclei not normally associated with nuclear reactor operation. Theoretical Division played (and still plays) a major national and international role in determining such information.

Complicated chemical separations are required to prepare specific materials for transmutation. Material separation processes have to be defined, and performance must be calculated for instances where experimental data are not available. Such chemical separation processes and operations are combined into an overall "flowsheet," which provides information on separation efficiencies and wastes produced during operation. In the early phases of ATW, the Theoretical Chemistry and Molecular Physics Group was a major resource.

Over the past decade, the ATW effort has continued to be an important national technical initiative. In recent years, ATW was joined with the Accelerator Production of Tritium effort to create the Advanced Accelerator Applications (AAA) program. AAA has in turn transitioned into the present Advanced Nuclear Fuel Cycle Initiative, which today involves several national laboratories and a current national budget of more than $60 million. The Advanced Nuclear Fuel Cycle Initiative is now an important part of national strategies to develop and deploy advanced nuclear energy systems that can play important roles to meet future energy needs without generation of carbon dioxide and other harmful emissions.

ATW had its genesis at Los Alamos, and T-Division played a significant role in its creation. Now workers in successor programs to ATW are developing key technologies for eliminating nuclear waste as a major impediment to nuclear energy development and application needed for the 21st Century.

A Theoretical Study of Geothermal Energy Extraction

Harlow
Pracht

by Francis H. Harlow and William E. Pracht

Efficient extraction of geothermal energy from a dry well depends on the ability to establish a closed pressurized circuit of water through a large zone fractured in hot impermeable rock. (See Fig. 1.) Long-term perpetuation of significant power extraction depends, in addition, on the ability to extend the initial fracture zone through the effects of thermal-stress cracking of the adjacent hot rocks. In support of an experimental program to test the feasibility of using this type of energy source, we have solved numerically the combined equations describing the coupled processes of fluid flow, heat transport, and rock fracture. The results show a strong dependence on the extent to which underground pressure can be maintained and the fracture zone continuously extended. They indicate that under favorable, but perhaps not unreasonably exotic, circumstances the extraction of significant thermal power from each well can be expected to continue for several decades. [1]

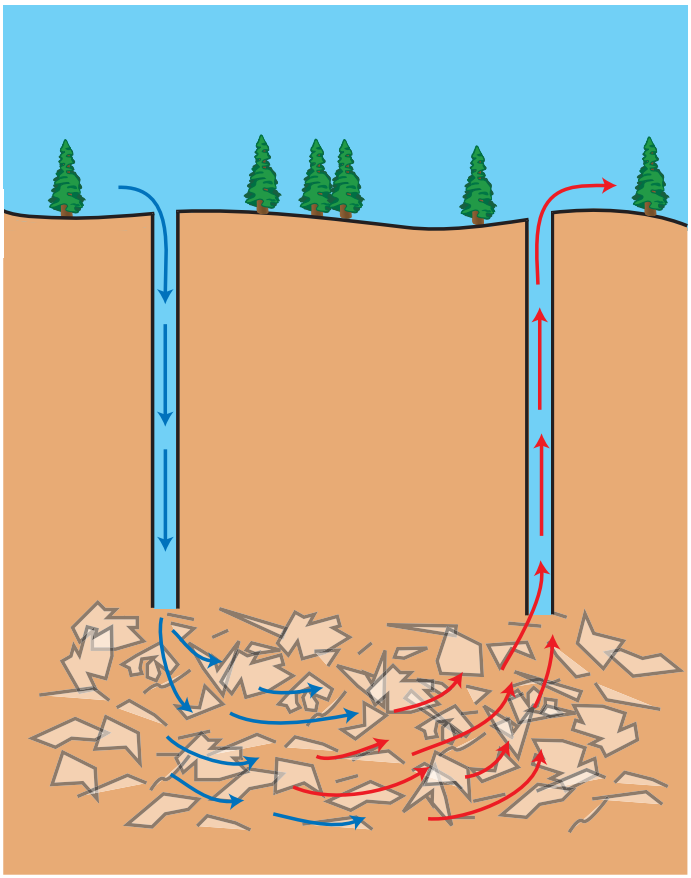


Fig. 1. The figure shows a pumping circuit through fragmented hot rocks.

Computer Analysis of Wildfire Behavior

by Rodman R. Linn, Jon M. Reisner, and Francis H. Harlow

Harlow
Linn
Reisner

Wildfires are a major threat to human life and property, as demonstrated in the last several years by major catastrophic fires in many parts of the United States. Wildfires such as the Cerro Grande Fire adjacent to Los Alamos, which burned 200 private structures, displaced more than 400 families, and consumed 47,000 acres (some of it Laboratory property), can be disastrous. Many areas of extreme fuel loads have resulted from a century of fire suppression, while the growth of the "urban-wildland interface" is continuing to put increasing numbers of people and property at risk. To examine the behavior of wildfires, the Los Alamos National Laboratory Theoretical Division (in conjunction with the United States Forest Service and Los Angeles County) has developed a new computer code for the self-determination of wildfire propagation in any terrain, vegetation, moisture, and meteorological circumstances.

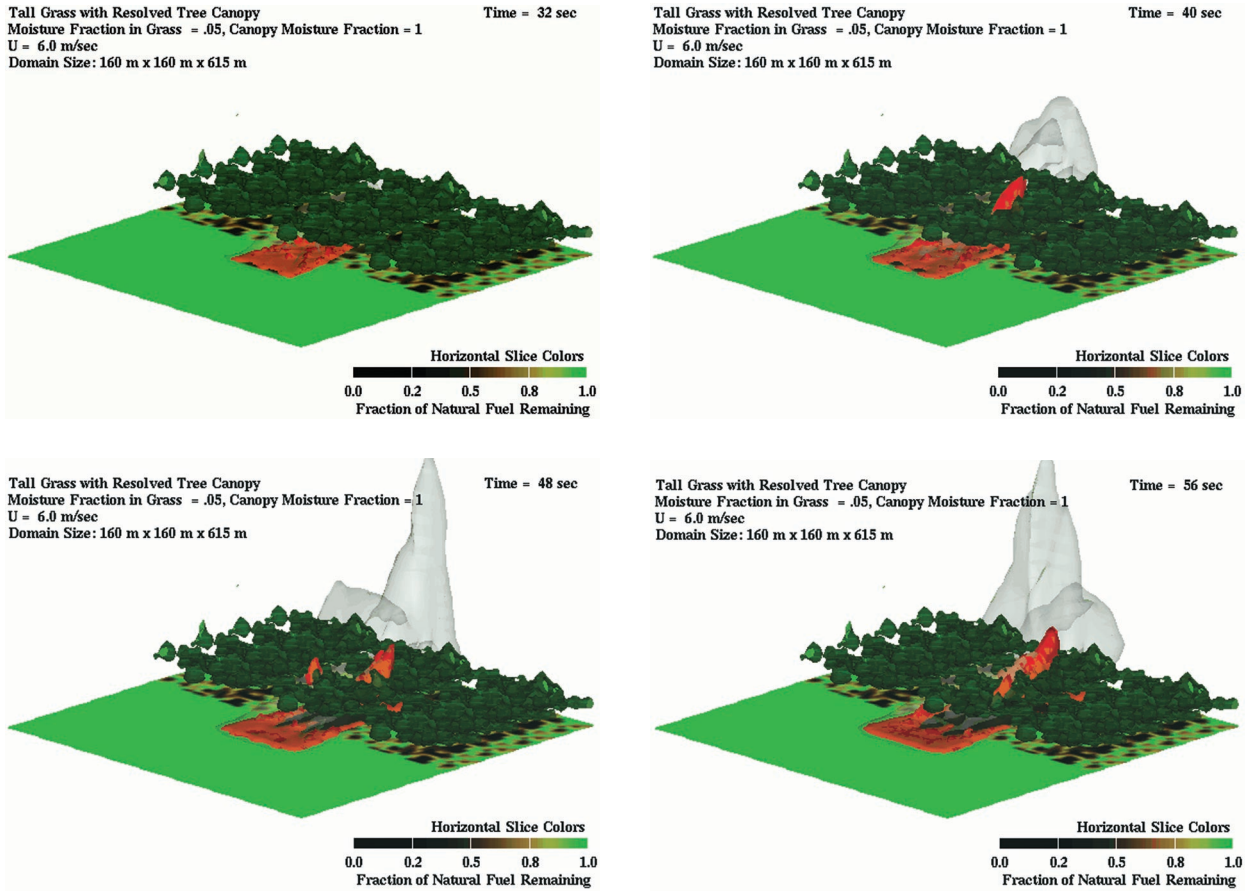


Fig. 1. Transition from a ground fire through grass land to a crown fire in the forest. At first the fire in the forest continues to burn only along the ground, but it soon jumps up into the upper foliage of the trees and propagates as a much more intense and potentially dangerous crown fire.

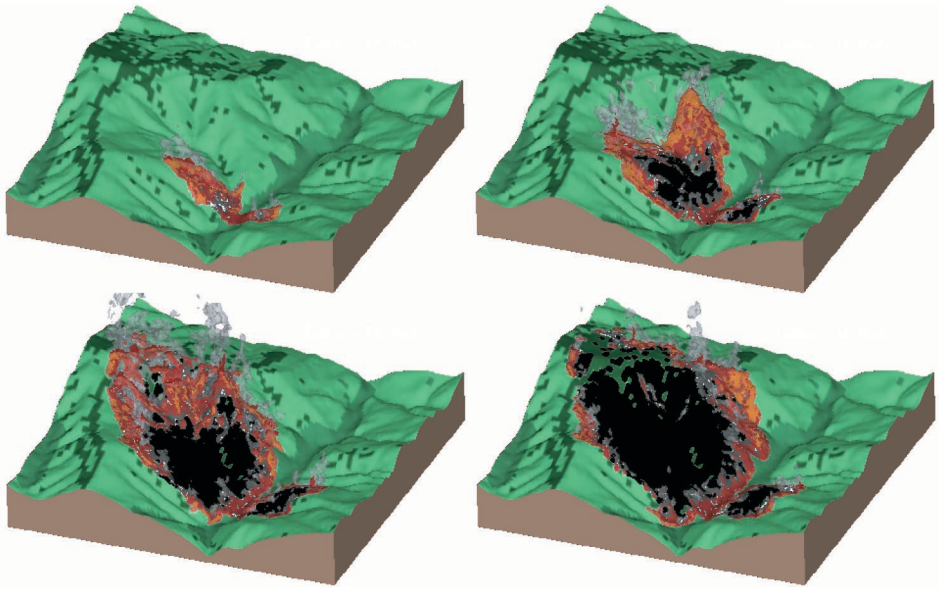


Fig. 2. Calculation of the Calabasas Incident (1996) near Los Angeles, California.

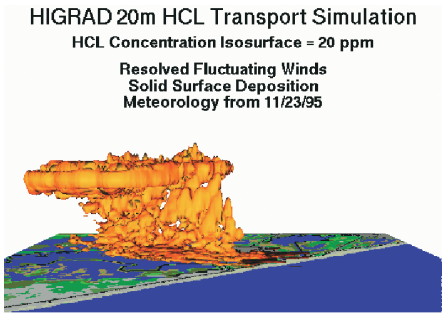


Fig. 3. Transport of hydrochloric acid from burning solid rocket fragments from a hypothetical accident at Kennedy Space Center.

In contrast to empirically based computer models for wildfires, our new code, called FIRETEC, is based on the fundamental laws of material dynamics and chemistry, incorporating subgrid representations of the unresolved details: individual leaves, branches, trees, and the turbulent eddies that swirl through the burning vegetation. The calculations are fully three-dimensional (3-D) and are coupled to the HIGRAD atmospheric dynamics code for computing the interactions between the regional winds and the fire-induced modifications to those winds. Figure 1 shows a FIRETEC calculation of the transition between a ground fire and a crown fire.

Numerous applications of FIRETEC demonstrate the validity of these calculations in real circumstances, from laboratory-scale burns to such major conflagrations as the Calabasas canyon fire near Los Angeles, California (Fig. 2). Simulations of potential fires in vulnerable locations in complex terrain including calculations of fires on Laboratory property have illustrated the value of FIRETEC as a planning and risk-analysis tool. As requested by the National Aeronautics and Space Administration (NASA), we also calculated the consequences of noxious material transport away from burning solid rocket propellant in a hypothetical accident at Kennedy Space Center (Fig. 3).

The applications of FIRETEC are especially in the areas of planning controlled burns, determining the optimal placement of green areas and fire breaks around new building developments, assisting fire marshals to formulate remedial strategies during the real-time unfolding of an uncontrolled wildfire, making value judgments as to best strategies for ensuring human safety during fires, studying the effectiveness of remedial techniques (slurry, water, firebreaks) in fierce wind and rugged-terrain circumstances, and developing risk-assessment strategies for insurance companies.

Climate, Ocean, and Sea Ice Modeling

by Mathew E. Maltrud, Elizabeth C. Hunke, Philip W. Jones, Richard D. Smith, John K. Dukowicz, and William H. Lipscomb

Dukowicz
Hunke
P. Jones
Lipscomb
Maltrud
R. D. Smith

In recent decades, the impact of the human population on Earth's climate has become a hotly debated topic in both the popular press and scientific journals. Theoretical Division scientists have developed state-of-the-art climate model components that enjoy international recognition and are used throughout the global climate-modeling community. This work has set a new standard for ocean and sea ice modeling and currently represents the highest-resolution simulations ever carried out on a global scale. This on-going research is aimed at understanding how the human race will be affected by climate change in the 21st Century.

In recent decades, the impact of the human population on the Earth's climate has become a hotly debated topic in both the popular press and scientific journals. Twenty-five years ago, [1] some feared that our climate was on the brink of another ice age, while others predicted a human-induced greenhouse effect that would warm the atmosphere, a prediction now being realized in atmospheric temperature measurements. Today's scientists, however, still do not fully understand the climate system (which includes the oceans, land masses, and cryosphere in addition to the atmosphere), nor can we predict with certainty the consequences of human activity on the planet.

In the early days, "climate" generally referred only to average properties of the atmosphere; but some climate phenomena (for example, the El Niño/Southern Oscillation) depend critically on interactions between the atmosphere and other components of the climate system. Researchers have since recognized the importance of the ocean in the climate system, and greater emphasis was given to coupling of atmospheric models with increasingly realistic models of the ocean for climate studies. Now it is standard practice to build full general circulation models (GCMs) of the atmosphere and ocean that also include the land surface and sea ice (ice that freezes on the ocean surface, in contrast to glacial ice on land).

The polar regions provide an example of the complex interactions between components of the climate system. The Arctic Ocean is a basin surrounded by continental land masses, while the Southern Ocean surrounds Antarctica with no distinct dividing line between it and the Atlantic, Pacific, and Indian Oceans. Primarily because of this topographical difference, the ocean circulations in the two regions are quite different, resulting in a large-amplitude annual cycle of sea ice coverage in the Antarctic, while sea ice in the Arctic Ocean covers an extensive area year-round. (See Fig. 1.)

Fig. 1. Ice that forms on the polar oceans plays an essential role in the Earth's climate system. Sea ice drives thermohaline circulations in the ocean by rejecting brine when freezing, and it regulates energy fluxes between the polar atmosphere and the oceans. We have developed comprehensive sea ice and ocean models for use in global climate simulations. (Photo taken by Elizabeth Hunke in the Weddell Sea, Antarctica, January 1998.)

The presence of sea ice significantly alters the amount of solar radiation that reaches the ocean surface, and thus a growing ice cover will further enhance ice growth by reflecting solar radiation. On the other hand, a melting ice cover exposes more of the ocean surface to the atmosphere. The ocean warms through solar heating, which then warms the atmosphere and further melts the sea ice. A similar feedback process warms the land surface as snow and glacial ice coverage decrease, which also contributes to warmer regional atmospheric temperatures. Without the thermal inertia of large continental ice sheets (as in the Antarctic), the "permanent" ice in the Arctic Ocean totters on the brink of disappearing altogether, should it begin to retreat. Because of this sensitivity, the Arctic system is being monitored as a possible early detector of climate change signals, whether human-induced or natural.

Los Alamos Gets Involved

The U.S. Department of Energy is interested in modeling climate to assess the effects of fossil-fuel emissions and provide insight into future changes in energy use and regulation. Theoretical Division, with help from other Los Alamos divisions, first became involved in climate modeling in 1979 through a collaboration with the National Center for Atmospheric Research. Prior experience with atmospheric GCMs, particularly while carrying out "nuclear winter" modeling studies during the mid-1980s, made it clear that there were already more than enough atmospheric GCMs and too many institutions developing them to make that a viable direction. Ocean modeling, on the other hand, was an ideal match for the expertise available in Los Alamos and T-Division in fluid dynamics modeling and associated numerical methods.

Numerical modeling of the ocean had lagged behind atmospheric modeling in the U.S. for several reasons.

- Progress in atmospheric modeling was driven strongly by the country's numerical weather prediction effort rather than by concern about climate change.
- Observations of the atmosphere far surpass those available for the oceans.
- Of vital importance for the proposed effort at Los Alamos, accurate numerical simulations of the ocean involve a combination of high resolution and long periods of time, requiring massively parallel computers.

High resolution is needed to accurately represent eddies, fronts, and topographic features that influence the transport of heat and other properties in the ocean, while simulations of the ocean over long periods of time are necessary because the deep ocean responds to surface forcing on time scales of hundreds to thousands of years. In contrast, the atmosphere adjusts to changes in months.

This lag in ocean modeling and the realization that massively parallel computers were essential for climate modeling played a key role in the decision at Los Alamos and T-Division to focus on ocean modeling when the DOE Computer Hardware, Advanced Mathematics and Model Physics (CHAMMP) program was being established in 1991. In the ensuing decade, Los Alamos developed a strong program in numerical modeling of the oceans and sea ice, with special emphasis on parallel computing. One measure of our success lies in the broad dissemination of our models, which enjoy international recognition and use throughout the global climate modeling community.

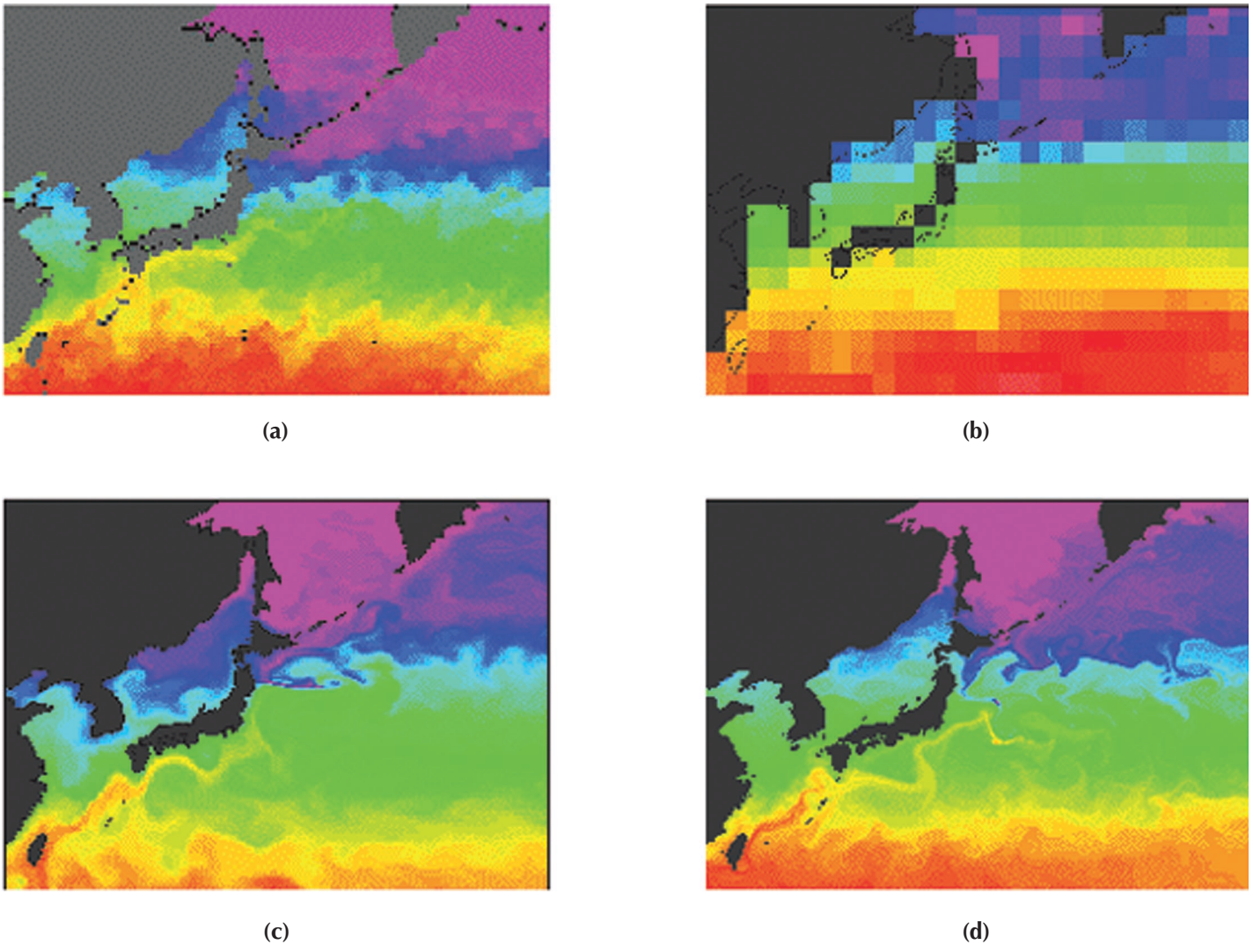
Parallel Ocean Program

Our focus on ocean modeling under the CHAMMP program resulted in the Parallel Ocean Program (POP). [2] POP was derived from earlier ocean models that solved the 3-D equations for fluid motions on a sphere under the "rigid lid" (flat surface) approximation and used a formulation for the fastest waves that required smooth topography and limited the number of islands or continental boundaries.

Our initial work on ocean models involved a new formulation of the fast-wave equations to allow a free surface, unlimited number of islands, realistic bottom topography, and a much faster method for solving these equations. The new model was designed to run efficiently on parallel computers, and further changes to the model numerical algorithms allowed us to use time steps four times larger than previous models. Following our initial successes, we have continued to improve POP by adding new mixing parameterizations and a better treatment of fresh water input to improve long-term climate simulations at coarse resolution. New horizontal grids, such as displaced-pole grids with the North Pole singularity shifted on to land points, have enabled a better simulation of the Arctic Ocean. We continue to improve the model's computational performance as computer architectures evolve.

Since the resulting POP model offered greater physical realism and an order-of-magnitude improvement in computational performance, we could consider undertaking modeling studies that had not been possible before. Using state-of-the-art parallel computational resources available to us at Los Alamos, we chose to pursue high-resolution simulations of the ocean circulation with the goal of explicitly resolving the most important scales in the ocean. The resulting realistic simulations of the ocean could then be used to investigate the role of the ocean in the Earth's climate, including mechanisms for transporting heat and salt around the globe, the role of turbulent eddies in the general circulation, and even short-term prediction.

When we performed our first major global ocean simulation in the early 1990s, global runs were typically configured with an average horizontal grid spacing on the order of several hundred kilometers. We immediately set a new standard with 25-km grid spacing and are now running the highest-resolution simulations ever carried out at global scale, with an average grid spacing of about 7 km (Fig. 2). As hoped, higher resolution has resulted in much improved fidelity in the simulation, where now we can compare model results directly with a variety of observations. Our focus on high-resolution ocean modeling has enabled us to become one of the few world leaders in this area and an important resource in helping the general oceanographic community understand ocean circulation.



(a) **(b)** **(c)** **(d)**

Fig. 2. Sea surface temperature in the northwest Pacific from (a) satellite data, (b) 2 × 2-degree POP simulation, (c) 0.28-degree POP simulation, and (d) 0.1-degree POP simulation. Note that the detail becomes much more realistic for the two high-resolution simulations (c and d). However, the warm Kuroshio current (off the eastern coast of Japan) separates much too far north in (c). The highest resolution grid is necessary to get the separation correct.

Sea Ice Model

Another important accomplishment of the Climate, Ocean, and Sea Ice program at Los Alamos and T-Division was the development of a new and innovative approach to sea ice dynamics. Our elastic-viscous-plastic formulation [3] greatly improves the accuracy and efficiency of sea ice models, particularly in the case where ice floes are driven into each other and begin to "lock up," forming a rigid mass. In addition to ice dynamics, the Los Alamos sea ice model (CICE) contains state-of-the-art components for ice ridging, thermodynamics, and horizontal advection. The thermodynamics model accurately simulates the atmosphere-ice-ocean energy budget and ensures conservation of energy. For better estimation of ice strength and growth and melt rates, CICE resolves a distribution of ice thicknesses within each grid cell into "thickness categories," including thin first-year ice and thick pressure ridges. A new linear remapping scheme is used to transfer ice between thickness categories, as ice forms and melts. A similar remapping scheme is used for the horizontal advection of ice and is very efficient for the transport of many ice categories. CICE also supports general horizontal grids, making it a natural companion for POP in coupled models because it can share the displaced-pole grids, eliminating the need for interpolation of information being exchanged between these closely coupled components.

CICE has been applied to a variety of interesting problems with a focus on the interaction of sea ice with the ocean and atmosphere. Figure 3 shows satellite imagery of the sea ice concentration in the Weddell Sea, Antarctica, in early 1998, and results from a CICE simulation. A large area of the Weddell Sea that is usually ice covered became ice free during the summer season. Model simulations revealed that the event was due to an unusual wind pattern during the spring and summer months, when solar radiation was available to melt ice. Ice disintegration accelerated as the ocean surface warmed, until temperatures dropped and the entire area quickly refroze in the fall.

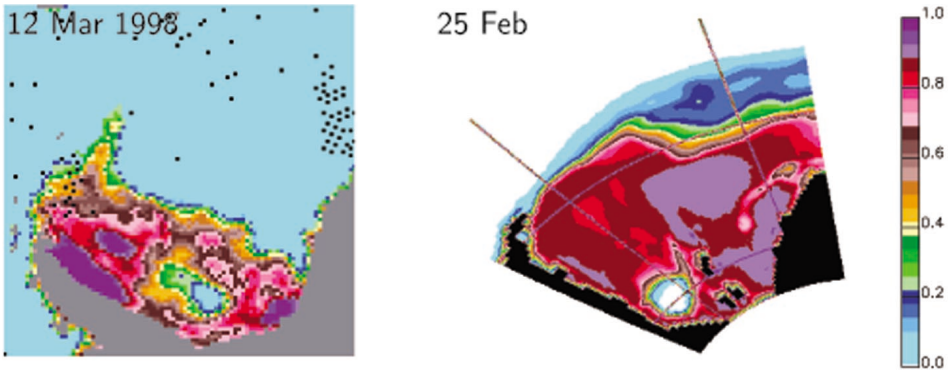


Fig. 3. Satellite imagery of the sea ice concentration in the Weddell Sea, Antarctica, in early 1998 (left), and results from a CICE simulation (right). The units are fractional area coverage, ranging from zero (no ice) to one (total coverage). An area of open water (a polynya) had appeared in the southern Weddell Sea during the previous summer melt season that was many times larger than any polynyas observed in the area since satellite observations began in the early 1970s. Numerical simulations of this event, using the CICE model, indicated that the cause was an anomalous wind pattern that advected the sea ice north, away from the Antarctic continent. As more of the ocean surface area was revealed to the atmosphere, the upper ocean layers absorbed greater amounts of solar radiation and warmed, thus further melting the sea ice. The figure shows the ice concentration at the end of the summer, as the ice cover began to reform. In both the satellite imagery and the simulation, new ice in the southern Weddell appeared first along the outer margins, leaving a "hole" in the pack, seen here, that gradually filled in over the course of about 2 weeks. This behavior indicates that the ice pack reformed by means of in situ freezing rather than by ice being advected from elsewhere.

Community Climate System Model

For a complete simulation of the Earth's climate system, ocean and sea ice models must be coupled to atmosphere and land surface models. To accomplish this coupling, we have collaborated with scientists at the National Center for Atmospheric Research (NCAR) in the development of the Community Climate System Model (CCSM). In the CCSM, atmosphere and land models (developed at NCAR and elsewhere) are coupled to the POP and CICE models through a "flux coupler" that keeps the models synchronized in time and passes information about each model's state to the other models. Fluxes of momentum, heat, water, and important chemical species such as carbon dioxide (CO_2) are also passed through the flux coupler. Because the models use different computational grids, the coupler must remap fields between these grids. Heat and water fluxes must be remapped in a manner that conserves the total heat and water in the climate system. The Spherical Coordinate Remapping and Interpolation Package (SCRIP) was developed in T-Division to perform these needed transformations and is an integral part of the CCSM flux coupler.

Once all of these components are coupled together, very complex simulations of the climate system may be carried out. The first step is to determine whether the model can reproduce important aspects of the observed climate system. When confidence in the model is high enough, simulations of future climate can be undertaken. For example, the effect of increasing atmospheric CO_2 and the potential for global warming can be investigated. In fact, the CCSM (which, again, includes POP, CICE, and SCRIP as essential components) is one of the models used by the United Nations' Intergovernmental Panel on Climate Change (IPCC) to make international policy assessments on the possible effects of global warming.

Global Warming

One potential issue associated with global warming involves an interaction between land ice, sea ice, and the thermohaline circulation of the ocean (the large-scale ocean circulation driven by both temperature and salt differences). Ice shelves, portions of continental ice sheets that extend over the ocean, may disintegrate due to a variety of glaciologic and climatic factors, including decrease in ice inflow from continental ice sheets, increase in basal melt rates due to ocean warming, and enhanced fracturing due to increased surface melt.

If a general retreat of the West Antarctic ice shelves should occur, several million square kilometers of additional ocean area would be exposed to sea ice production. Since brine rejection during sea ice freezing is believed to be an important mechanism driving the thermohaline circulation, thermodynamic consequences of the West Antarctic ice-shelf disintegration could be felt globally. In fact, it has been postulated that such an exposure of additional ocean surface could trigger a switch from the current interglacial to a glacial mode of thermohaline circulation. [4]

We are preparing a numerical experiment aimed at testing the sensitivity of global thermohaline circulation to disappearance of the two largest West Antarctic ice shelves, the Ronne-Filchner and Ross ice shelves. Disintegration of these ice shelves may have a significant influence on the global ocean and climate system.

The next big step we are taking in modeling the full climate system is the incorporation of ocean biogeochemistry into POP. This involves modeling of two important processes that affect the global carbon cycle: ocean ecosystem dynamics and the transport of chemical species (such as CO_2) throughout the ocean. As the amount of CO_2 increases in the atmosphere, there will certainly be an effect on the phytoplankton that make up the base of the food chain. Also, it is unknown how much CO_2 may be absorbed in the ocean and how long it might stay there. We hope to soon use our models to better understand these kinds of very important details that will affect the human race in the coming century.

Could Carbon Fuel the Twenty-First Century?

by Klaus S. Lackner

Lackner
Ziock

Predictions of innovative energy technologies for the next century usually include everything from fusion to photovoltaics with the one notable exception of fossil fuels. Because of fears of diminishing supplies, pollution, and climate change, people are reluctant to consider fossil carbon for the energy needs of the 21st Century. An energy strategy for the new century, however, cannot ignore fossil fuels. Contrary to popular belief, these fuels are plentiful and inexpensive. While it is true that fossil fuels are limited by their environmental impact, new technologies to eliminate environmental concerns are currently being developed in the Theoretical Division. These technologies would make it possible to sustain the fossil-fuel option for more than a century.

Over the next century, the amount of fossil carbon that could be extracted from under the ground could be too large to be absorbed by the world's ecosystems. It is tempting to grow more trees or other biomass to increase carbon storage, to augment soils with additional carbon, to let the oceans absorb the carbon, or to simply leave the carbon in the atmosphere. However, the quantities of fossil carbon that could be used far exceed the uptake capacity of these natural sinks. For example, the amount of fossil carbon that could be produced in the next 100 yr is several times the entire biomass carbon on Earth. The amounts that could be produced would be enough to make the ocean so acidic that coral growth would be stunted. Excess carbon dioxide (CO_2) in the air causes climate change. (See Fig. 1.)

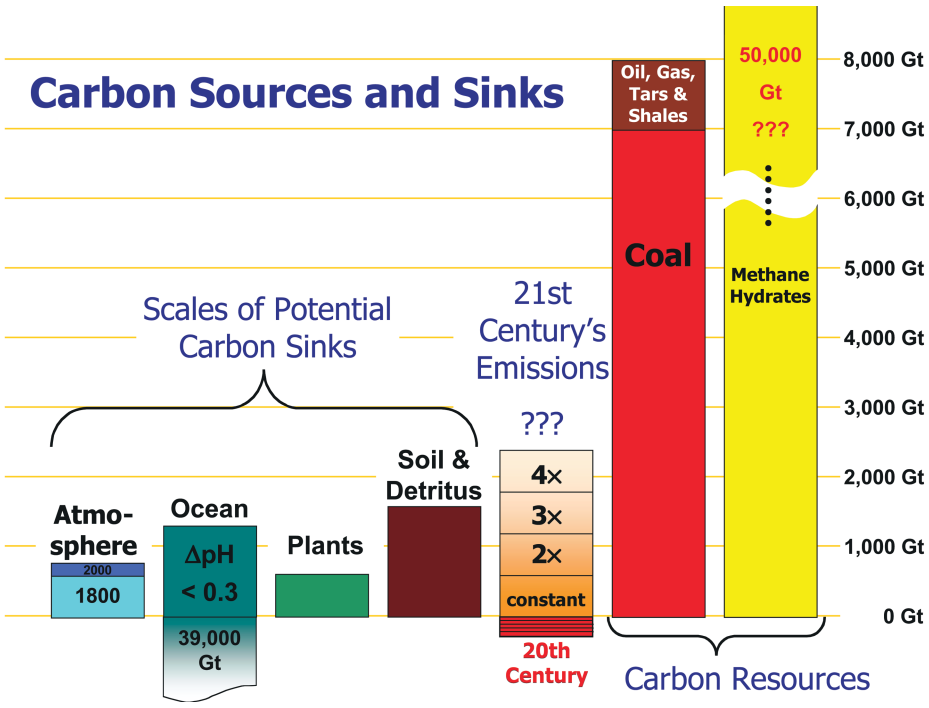


Fig. 1. The figure shows the size of carbon reservoirs on Earth. Potential emission of the next century (in multiples of current emissions per 100 years) are compared to the CO_2 content of the atmosphere, the buffer capacity of the ocean for carbonic acid, the size of the biomass pool and the amount of organic carbon stored in the soil. On the other side of the diagram, estimates of the available fossil carbon resources indicate that fossil are likely to last for centuries. The use of carbon is limited by environmental constraints not resource constraints.

Clearly CO_2 emissions must be curtailed to prevent buildup of excess carbon in the world's ecosystems. Efforts in conservation are important and will extend the time period available to develop new energy technologies, but these cutbacks will be inadequate in the long term. For example, if a future world of 10 billion people were to share equally in a CO_2 emission budget one-third the size of today's world emission, the per-capita emission allowance would be only 3% of the actual per-capita emission in the U.S. today. Nonfossil forms of energy, such as wind energy, solar energy, or nuclear energy, are technically possible; but they are too small in scope, too inefficient, too expensive, or too unpopular. Furthermore, eliminating fossil carbon, which is the world's largest energy source, from world energy markets would likely cause a major political crisis. Instead of focusing solely on replacing fossil energy, technological options for net-zero carbon emissions should be explored.

Environmental Solutions

The engineering obstacles in achieving net-zero carbon emissions are less daunting than the political obstacles. The effort to reach a nationwide consensus on the necessity of paying for carbon management will be greater than the effort involved in building a zero-emission power plant. Fortunately, the price of carbon management appears small when compared with recent swings in the energy market. Furthermore, the incremental cost of using abundant resources such as coal, oil shale, or tar instead of more limited ones, such as oil and gas is small. All of the above options are less expensive than solar, wind, or nuclear energy.

In order to prevent the accumulation of CO_2 in the air, one needs to collect CO_2 from the source or from the air and then dispose of it permanently and safely. Biomass, soils, and the ocean are limited in their uptake capacity. Two less-limited disposal options are the chemical fixation of CO_2 as a carbonate and the injection of carbon into underground reservoirs. I prefer the former because it ensures that the CO_2 is stored safely and permanently.

Mineral Sequestration

Solid mineral carbonate formation occurs spontaneously, albeit slowly, in rock weathering. Aboveground mineral carbonation accelerates natural carbonation in an industrial process. By mining readily accessible and abundant magnesium silicates (e.g., peridotites or serpentinites), grinding them up and letting them react with CO_2, one produces magnesium carbonate, silica, and usually some water. The process is a net energy producer. Mining, crushing, and grinding are similar in scale to conventional mining and contribute less than $10 to the disposal cost of a ton of CO_2. That amount would add between 0.5 cents and 1 cent to the price of a kilowatt-hour of electricity, and it would add 8.5 cents to the price of a gallon of gasoline.

The engineering challenge lies in the chemical process. We have developed a successful implementation, but further improvements are required to drive costs down. In a recent independent engineering study, Nexant, Inc. suggested that implementing the current laboratory designs on an industrial scale would drive the cost of mineral carbonate disposal to $70 per ton of CO_2. According to the same study, streamlined implementation of the same process could hold the price below $30 per ton. Our long-term goal is $15–$25 per ton.

While providing less certainty as a permanent disposal method than the carbonation process, the injection of carbon underground is already possible with today's technology and, indeed, is performed routinely in enhanced oil recovery. As the scales of disposal get larger, establishing good injection sites will become more challenging; and the carbonate process will probably become more popular. However, if the political will existed, CO_2 could be stored permanently and safely starting today at a cost below $30/ton of CO_2.

Capturing Carbon Dioxide

Before one can dispose of CO_2 it needs to be captured. Capture from existing power plants would entail retrofitting costs that are best avoided. However, there are new plant designs that would capture CO_2 directly at the source. These plants could use abundant low-grade energy sources such as coal to produce hydrogen or electricity and deliver CO_2 and other wastes in separate concentrated streams ready for environmentally acceptable disposal.

There are a number of designs for zero-emission power plants. At Los Alamos, Hans Ziock and I have put forward one particular design. Ours may be some years from commercial availability, but it will be exceptionally efficient while producing clean power. Our technology, pursued by the Zero Emission Coal Alliance (ZECA), combines CO_2 capture and the elimination of all emissions to the air with an energy conversion efficiency of 70%. If, however, one temporarily would accept reduced efficiency, one could move to zero-emission power plants virtually immediately. This would provide a path forward. Later on, plants like the ZECA plant could efficiently provide clean electricity and clean hydrogen, which, in many people's opinion represent the future of energy.

There are no hydrogen wells, and the energy to produce it must come from somewhere. The ZECA process is well suited to generate hydrogen from water with energy from coal. Coal energy is five-to-ten times cheaper than oil or gas. Electricity is even more expensive than oil, effectively ruling out electrolysis as a source of hydrogen. Consequently, gasification processes like the ZECA process would provide a far more cost-effective basis for a hydrogen economy than other energy sources. Potentially, hydrogen could become cheaper than natural gas. Central plants could generate electricity or hydrogen and safely dispose of all waste products including CO_2.

Carbon dioxide capture may be most effective at the source, yet a large fraction of all emissions are from small, distributed sources such as cars and planes. It is difficult and expensive to collect CO_2 on board these mobile sources. Consequently, we have looked at the possibility of capturing CO_2 directly from the air, using wind as the transport agent. This is seldom considered because CO_2 in the air is very dilute. At first sight, this low concentration seems to be a huge obstacle; but in the case of wind energy we have learned to capture a far more dilute commodity. At a reasonable wind speed of 6 m/s, the CO_2 flow through an aperture the size of a small television screen (0.2 m^2) equals the per capita production of CO_2 in the U.S. The aperture required for the kinetic energy flux to equal the U.S. per capita energy consumption is 80 m^2.

The Benefits of Air Extraction

We have sketched out a process to capture CO_2 from the air using lime as the capture agent. Still far from optimal, the process suggests a long term price goal of $10 to $20 per ton of CO_2 captured. This translates into a small addition in the price of gasoline. If proven economically viable, this technology would represent a major breakthrough, as it would allow the continuation of current life styles while eliminating climate-change concerns. It would allow CO_2 capture by developed countries on behalf of countries that cannot afford to do so and even

would allow consideration of reducing atmospheric CO_2 levels some time in the future. Introduction of this approach requires no changes in the existing infrastructure, and the cost should add less to the price of energy than recent price swings have done.

In short, fossil energy should not be ignored. It is capable of fueling the world for generations to come. Low-cost energy resources can produce hydrogen and electricity that could be used in the markets for which they are best suited. Cars could run on hydrocarbons, which are excellent energy carriers. With increased energy conservation and energy efficiency and the decoupling of the disposal of CO_2 from the consumption of fossil fuels, fossil energy could sustain the economy for several centuries. This would allow more than enough time for alternative forms of energy to become competitive. Predictions of energy technology for the next century should include everything from fusion to photovoltaics, but especially fossil fuels.

Satisfying Energy Demands on a Sustainable Earth

by Klaus S. Lackner and Christopher H. Wendt

Lackner
Wendt

Having observed that human energy usage is fast approaching the levels at which Earth will not be able to provide enough energy to keep up with consumption, Theoretical Division scientists have devised a solution for energy production that can satisfy future energy needs for an expanding global population. They call their plan "recursive automation," a system in which large-scale automated mass production of both goods and equipment is coupled with harnessing of renewable, low-cost energy sources.

Striking a balance between the needs of an increased world population and maintaining a habitable planet for future generations will be one of society's most difficult challenges in the coming decades. Preferably, all peoples of the world should enjoy a standard of living similar to that taken for granted in the industrialized countries. At the same time, however, the human impact on the environment must be greatly reduced from today's levels, leaving a world, a culture, and a technology that can support the generations still to come. The achievement of a high standard of living within the restrictions of a sustainable Earth will require a revolutionary change in the way we produce energy and goods. It will need visionary planning and the will to embrace long-term global solutions.

Striking this balance successfully requires a level of effort and energy that is unaffordable at the current levels of productivity. Novel approaches to automation, which we refer to as recursive automation, could raise productivity to the necessary level by not only automating the manufacture of a product but also the mass production of the associated manufacturing equipment. In a recursive manner, the equipment necessary for the manufacture of all previously introduced equipment is added until the system gradually approaches one that produces everything necessary for its own construction and maintenance. As the input of human labor and capital into the construction and operation of such a system would be reduced, one would gain access to dilute, but inexhaustible resources whose extraction is currently unaffordable. The first such resource would be renewable solar energy. Eventually recursive automation could lead to machine systems that essentially build and maintain themselves and operate on a scale necessary to directly affect the global environment. They could counteract the myriad of deleterious side effects of human activities and allow a world society to live in comfort on a sustainable Earth.

One of the most basic requirements of a modern society is energy. About one sixth of the gross domestic product is spent on it. Fossil fuels, which are the world's most heavily used energy resources, are limited; and their extraction, processing, and ultimate consumption are the major contributors to the world's pollution problem. Since the production of clean, cheap, and copious energy would be an enabling technology for a sustainable global economy, this subject will be the focus of this discussion.

The Need for More Energy

Energy is the driving force of all sectors of a modern economy. Necessary for the production of raw materials, the manufacture of goods, for transportation, for heating, for cooling, and for lighting, energy determines our way of life. Precious as the commodity is, some of it is wasted because its cost may be small compared to other costs in a particular application. In other cases, however, desirable processes are not performed because their energy cost would be prohibitive. Ocean water desalination for agriculture is just one example of such a prohibitive cost.

The shift to a sustainable economy will require less reliance on nonrenewable resources and more effort in minimizing damage to the environment. Both changes will increase energy demand. Dilute resources, including those for energy, will have to replace those currently available. To extract values from these resources is clearly more energy intensive but nonetheless rewarding because supplies are virtually inexhaustible. The dilute emissions of pollutants, which past generations have been able to ignore, have begun to cause serious damage to the environment. Carbon dioxide (CO_2), other greenhouse gases, and ozone-destroying chloro-fluoro carbons could severely affect the global climate. In populated areas, pollutants such as sulfur dioxide (SO_2),

nitrogen oxides, and hydrocarbons, reduce the quality of life. To avoid these emissions will require a technological compromise, the cost of which frequently could be paid for with additional energy. To remove pollutants after the fact, which sometimes may be the only alternative, is labor and energy intensive. In a sustainable world, one must pay in energy, effort, and material, not just for manufacturing a product, but also for producing it from readily available, dilute resources; for removing its environmental impact during production and use; and for its ultimate disposal.

Energy conservation is not a goal in itself: a sustainable world economy is. Such an economy is easier to achieve if energy use per capita can be allowed to rise. For this reason, it is not likely that a sustainable economy will greatly reduce energy consumption. However, energy from fossil fuels is one of the largest contributors to the degradation of the environment and thus cannot be readily used in environmental remediation. Either we find better ways of using fossil energy, or we need a clean and cost-effective energy source.

If per capita energy consumption is not likely to be reduced, total energy consumption is surely going to increase. It would be unrealistic to assume that the world population will not grow to at least 10 billion people over the next 50 yr. Indeed, this already would require a large drop in the population growth rate. A stable or gradually decreasing population is a necessity for a sustainable world economy. Population growth must be stopped. The question is how. Historically, birth rates have dropped dramatically in countries that have reached a certain level of development. Increased wealth, probably in conjunction with better education, has halted the natural growth in all of the industrialized countries. Starvation, low probabilities of survival, poverty, and disease all tend to drive up the birth rate and ultimately lead to population growth. With the possible exception of totalitarian approaches, unacceptable in a free society, the only proven avenue to a stable population is an increase in the standard of living. To outrun population growth with economic growth will require readily available energy.

Alleviating the plight of the third world should also reduce otherwise mounting societal tensions. In today's information age with worldwide mobility, all societies on Earth become interdependent. Therefore, one cannot tolerate the current trend of a world dividing itself into a rich class that controls access to all major resources and a poor class that can be identified not only through its economic status but also through its ethnicity. The outcome could well be a world uncomfortably similar to the South Africa of the apartheid era. Such a world society would not only face similar moral issues but also the same instability and inherent danger of a chaotic and violent overturn. Ideologies attempting to justify this unfortunate state of affairs would arise in order to rationalize the withholding of scarce resources from those who need them the most.

Power production would have to increase by a factor of 10 to provide a world population of 10 billion with the same per capita amount of energy currently available in the U.S. Surprisingly, if the current growth rate in energy production could be sustained, this increase would be achieved in about 50 yr. Unfortunately, with current technology, such growth is not sustainable.

The Energy Options

What are the options for averting this global energy crisis? The first one, which is widely advocated and is based on the assumption that environmentally benign energy is not available, is to use energy much more efficiently and to accept a serious decline in living standards. In a world economy sharing equitably the current world energy output, the U.S. would have to scale back its per capita consumption by a factor of 10. Even this may not be sufficient, because today's emission of greenhouse gases is already considered too high.

The difficulty of achieving such high levels of energy savings is illustrated with a few examples.

(1) The use of private vehicles would have to be severely curtailed.

(2) Superinsulated houses, which could lead to dramatic savings in heating, were advocated before too much was known about indoor pollution, the control of which reduces the potential for energy savings.

(3) Fluorescent lights that dominate commercial buildings and sodium vapor lights in street lighting perform at much better than 10% efficiency. Thus a reduction of energy consumption by a factor of 10 in this sector would require a noticeable reduction in light output.

(4) In industrial processes, energy consumption has been steadily reduced over the last decades. By now the easy savings may well have been achieved. In many cases industrial processes are not far from their theoretical limits. Steel, aluminum, and silicon production are all well within a factor of two of their theoretical minimum, which is set by chemical-binding energies. To achieve an overall energy savings by a factor of 2, without greatly reducing our standard of living, appears to be optimistic.

Large-scale improvement in energy efficiency will only come about through scarcity and increased cost. Thus, if energy could not be obtained in an environmentally benign and cost-effective way, energy conservation could make great progress, but the ultimate goal of a sustainable Earth would be much harder to achieve. The use of energy in improving the quality of the environment would be discouraged just as much as any frivolous use of energy.

Increasing the use of fossil fuels is certainly possible. Oil resources would dwindle rapidly if used at a much higher rate, but coal and probably natural gas are available for the foreseeable future. However, fossil-fuel consumption is one of the most polluting activities of mankind. At a ten-fold increase in consumption, the CO_2 content of the atmosphere would rise rapidly. The current rate of increase has been measured to be about 1.6 ppm/yr. Under the conservative assumption that the rate at which CO_2 accumulates in the atmosphere is proportional to anthropogenic CO_2 emission, at 10 times current emissions the CO_2 level would rise every 17 yr by an amount equal to its natural, preindustrial level. At this rate, serious physiological effects on humans and other life would not take long to develop.

Much has been made about the fact that burning methane (CH_4) generates less CO_2 than burning coal, but these differences are too little to invalidate our conclusions. The 17 yr, based on the assumption of an unchanged mix of all energy sources, would be stretched to 24 yr if all fossil fuels were to be replaced with CH_4.

A revolutionary change would be the zero-emission coal or hydrocarbon plant, which may well become a viable option. In such a plant, all combustion products would be captured and disposed of. In the case of hydrocarbons, in particular CH_4, only the hydrogen component could be used for energy production, leaving carbon behind. Whether the waste product is carbon, CO_2, or calcium carbonate ($CaCO_3$), the amounts would be in the hundreds of billions of tons per year. This is more than could be profitably reused in other industries; and, even in solid form, it would compound the already difficult waste-disposal problems. There is currently no good way of disposing of gaseous CO_2. Pumping it back into abandoned oil wells has been suggested, but such an operation would raise the price of energy. It should be noted here that the binding of CO_2 as relatively harmless calcium carbonate ($CaCO_3$) could become cost-effective through recursive automation.

Nuclear energy is a readily available option. It should not be ruled out, even though it is politically tainted. It generates energy without the emission of CO_2 and thus does not contribute to the greenhouse effect. The drawback of nuclear energy in the context of a sustainable economy is its demonstrated inability to provide energy at significantly reduced cost. Relative to its competitors, nuclear power has become more expensive over time.

Fusion energy is a dark horse in the race. Whether it arrives in time and turns out to be cost effective remains to be seen. Certainly there is no time to wait.

Finally, there is renewable energy. Solar energy, which is the ultimate source of most renewable energy, is plentiful. At 10% conversion efficiency, the area required to satisfy the current U.S. demand for electricity is 120 × 120 km. This land area is comparable to that occupied by military bases in the California and Nevada deserts.

Today, photovoltaic solar energy is an established technology. Energy efficiencies exceeding 10% are commercially available. In the laboratory, 30%–40% efficiency has been reached; but for large-scale applications, the cost of photovoltaics is still too high by about a factor of five. This cost disadvantage is even larger if energy storage is included.

The use of solar energy would also exact an environmental price, mainly through the effects on the large land areas occupied. However, solar energy is much more benign than current forms of energy production, and there are several factors which lower its impact on the environment.

(1) A large fraction of the solar-energy production would occur where it is needed and thus would affect areas that are already modified through human occupation. The rooftop of a single family home is sufficient for providing the energy consumed inside. In sunny areas like the southwest, parking lots could provide energy for cars. The roof of a single carport could satisfy the average energy demand of an electric car.

(2) Large-scale solar energy collection can be done in remote deserts or ocean areas with relatively little impact on adjacent areas. The impact could be further reduced through active controls. For example, the effective albedo of a solar-energy collection area could be controlled so that to a first approximation the area's solar-energy balance is not affected by the installation of solar collectors.

(3) Even though the affected areas would be large, they would still be small compared to the surface of the Earth and even small compared to the area of land severely modified by agriculture.

Renewable energy is the best long-term choice and its use should be aggressively promoted. Renewable energy comes in many different forms, some of which will still take time to mature as, for example, the direct photolysis of water with sunlight. In the case of photovoltaics, however, is it is mainly a matter of improving manufacturing methods so that existing designs become cost competitive. Based on the experience in electronics, one may expect that the semiconductor industry will be able to greatly reduce the cost of photovoltaic panels as demand increases.

In the following section, we will discuss an approach to automation applicable to large-scale mass production. Not only could this approach greatly accelerate the cost reduction of solar energy, it would also have a much broader effect on a sustainable economy by raising productivity to levels at which it becomes affordable to control the side effects of human activities on the environment.

Recursive Automation

Automation in the manufacture of industrial and agricultural goods has progressed during the 20th Century to a remarkable degree. While at the beginning of the century it took a large portion of the population simply to grow food, today this portion is a few percent. Automated factories are becoming a reality. Fully automated production lines for solar cells already exist. Indeed, it is hard to find a manufacturing process, which could not be automated—at least in principle—using today's technology base. Therefore, one may consider the possibility of not only automating the manufacture of a product, but also the construction and maintenance of the manufacturing plant.

Recursive automation defines a new paradigm for mass production. Cost reductions are achieved not by maximizing the efficiency of the final automated production step but by automating the manufacture of all associated production equipment. In order to produce the manufacturing equipment, new equipment is introduced and its automated production will require more equipment. By recursively adding the automated production of all previously introduced equipment, one will eventually reach the point where there are no more new types of machinery needed and the manufacturing system can perform all processes necessary for its own construction and maintenance. Of course, the generation of some items may be so difficult or occur so rarely that automation cannot be done cost effectively. Such items would have to be provided externally.

The major advantage of this approach is that cost savings, which have been achieved in each level, are multiplicative. Hence, very large cost-savings could be realized provided that the potential demand for the product is large enough to justify the mass production of the manufacturing equipment. A large factor in cost savings implies though that the inputs to such systems are of low cost. Thus, for a very large system, the inputs must be readily available raw materials.

Renewable energy is an ideal place to apply these ideas. Energy demand is very large and thus justifies the automated mass-production of the manufacturing equipment. Already, the production of solar cells can be automated; and one can begin to reduce their cost through recursive automation, which limits the input of human labor and capital to the comparatively small amount of externally provided materials and machines. If one can achieve a cost reduction by a factor of 10 or 20, which does not appear very large compared to factors actually achieved in the computer industry, one would cause a revolution in the power industry. Solar energy would be the cheapest form of energy available.

Small entrepreneurs could literally start farming the deserts for energy at an efficiency that far exceeds that of photosynthesis. For some time, solar energy could be used to add to the existing generation and mainly satisfy the peak-demand loads that in arid areas tend to coincide roughly with the time of maximum solar irradiation. For solar energy to gain a larger market share, eventually energy storage would become necessary. Currently used methods include hydroelectric systems that store energy by pumping water into a reservoir at a higher elevation. Effective storage systems of the future might be based on a fuel-cell technology made cost effective by recursive automation.

Energy production would only be the first application for recursive automation. As a system comes close to building and maintaining itself, the scale to which it could expand would increase, and the tasks it could perform would become accordingly larger. Whereas initially these systems would only reduce the price of solar-cell production, in later stages, they can perform most of the assembly necessary to collect energy from a large solar-cell farm. A system that builds and maintains itself would have the potential for exponential growth and could attain a very large size. However, in order to remain profitable, this growth would have to be accomplished without the external supply of equally large provisions. Therefore, all significant requirements would have to be procured by the system itself from its natural surroundings. An example would be the solar energy, which would fuel such a system. Other requirements would have to be satisfied from locally available water, air, and minerals. The minerals would have to be obtained in the form of raw dirt or rock rather than in the traditional enriched forms that come from prospecting or mining operations.

Components of Such a System

It is impossible to predict in detail how a system with such extreme levels of recursive automation would look like. However, any such system would be constrained by the environment it operates within, by the current state of the art of automation, and by the requirement of rapid growth. Taken together, these constraints allow one to obtain a reasonable image of such a processing system.

One component would be a fixed infrastructure, which would include almost complete coverage of the involved area by solar-energy collectors. The infrastructure would also include a simple grid of tracks for transportation and some form of energy distribution. The tracks would provide a controlled environment for the other major component, which would consist of a large variety of machines that we call auxons. The infrastructure would be the interface between these auxons and the natural terrain. The collective purpose of the auxons would be to extend the infrastructure onto the surrounding empty land and to make more machines like themselves. Upon reaching the design size, nearly the entire power-generating and material-processing capacity could be diverted to perform useful tasks.

Thus we start with a picture of machines sitting or moving on a network of tracks, some of which dig up a thin layer of dirt, about one centimeter on average, and pass it on to the beginning of a chemical separation, refining, and manufacturing chain. All the machines in this chain are powered by the solar energy, which is collected in the surrounding infrastructure. The final products of the manufacturing chain are new tracks and solar-energy collectors, which are transported to the periphery of the system and are installed to form additional infrastructure and additional auxons, which will occupy the new infrastructure.

The auxons themselves would be specialized machines with the minimum sophistication necessary to accomplish their assigned task. Only together could they form a growing system in the same way that a factory requires many different machines for its operation. Whenever possible, auxons would use physical constraints and simple rules for their operation. In this sense, they would be similar to machines on the factory floor. However, the system would have to be sufficiently open so that the effects of a failure of a single auxon could easily be absorbed by its neighbors.

We have analyzed such systems in some detail and have come to the following conclusions: the fundamental constraints can be satisfied. By starting with today's level of technology, one can use recursive automation to implement such a system step by step. Even though the task of integrating a very large number of processes is formidable, it appears that no "miracle" technologies would be required for success. For each production chain, which starts with the raw inputs and ends with a single product, we have identified necessary and sufficient productivity requirements that establish the minimum level of productivity and the maximum number of steps in a single production chain. The question of whether complete automation is feasible has been reduced to the satisfaction of these requirements for each individual chain. As a result, it is possible to break the large task up into small pieces, the successful resolution of which can be ascertained without detailed knowledge of the ultimate auxon system.

A large portion of an expanding auxon system's production capability would be dedicated to processing its material input of raw dirt and extracting the elements required for the growth of the system. The typical composition of raw dirt is such that all elements necessary for building an auxon system are available and can be used in roughly the proportions in which they occur. This requires, however, an unusual chemical-separation scheme based on high-temperature metallurgical processes, which we designed for this purpose. Conventional separation schemes would make use of large amounts of relatively rare elements, such as fluorine, and are, therefore, ruled out.

Under reasonable assumptions concerning the energy-collection efficiency, the system mass-per-unit area (10 kg/m^2), and the sophistication of automation achievable, one can conclude that such a system could increase its area coverage at an exponential rate with a time constant on the order of 6 months. As a result, an increase in size from a small seed system to a production system 10^8 times larger could be achieved within a decade. At a total coverage of 10^6 km^2, which is comparable to half the cropland in the U.S. or one-tenth of the Sahara, the electric-energy production would be about 25 times today's worldwide electricity production.

Implementation

It is an important feature of recursive automation that it can be implemented step by step. Each step would improve on current technology and would help increase our industrial productivity. Rather than implementing a complete auxon system today, one would start with reducing the price of solar cells through recursive automation. As one proceeds, the cost of inputs would also have to be reduced, and one would eventually have to use cheap raw materials, which the system could collect on its own. Once such a system could provide cost effectively for its own energy needs, it could be extended to a much larger size and it could be used in other applications. For example, such a system could be designed to desalinate seawater and provide a limitless supply of water for agriculture. The extraction of aluminum, titanium, and possibly other values from low-grade ores would be further applications. Auxon systems could process ores, which would be considered useless at today's extraction costs. As the state of the art in auxon technology progresses, ore composition would differ less and less from ordinary crustal abundances. This would open up a virtually limitless and sustainable resource.

Ultimately, such a system would have a profound effect on the human economy. Since nearly complete auxon systems can grow exponentially, they can reach any desired size in a short time. As a result the investment in human labor into generating an output is no longer proportional to the size of the output or the number of steps in producing it. Instead, the human input is only related to the difficulty of devising a new production chain to be incorporated into an auxon system. Thus, auxon systems cannot only provide a practically unlimited supply of energy, raw materials, and basic manufactured goods, they can also be made large enough to deal with environmental problems such as air pollution or climate change on a global scale. Indeed, the global warming caused by CO_2 emissions from human activities over the last centuries could be arrested with auxon technology. An auxon system covering 10% of the Sahara could remove the excess CO_2 from the atmosphere in about a decade. Because the productivity of an auxon system would be so high, we could afford to spend a large effort on counteracting the deleterious side effects of any processes occurring within an auxon system and those of other human activities.

It is because of these additional benefits, that we believe recursive automation to be the best path to a sustainable world economy. Such a system will make it possible to rely on dilute but very common resources and will lead to plentiful and cost-effective energy, which can be applied to maintaining the environment. Clearly, a strong push to such a high level of automation will cause dislocations and difficulties, as societies adapt to their changing circumstances. However, whereas recursive automation offers a path to a sustainable Earth, business as usual (as presently conducted) will most surely lead to extreme dislocations that could manifest themselves in severe famines, environmental catastrophes, and worldwide societal unrest that could end our civilization.

Air Pollution Transport in Street Canyons

by T. Daniel Butler, Larry Cook, Frank H. Harlow, C. W. Hirt, and Robert S. Hotchkiss

Butler
Cook
Harlow
Hirt
Hotchkiss

In the early 1970s, questions arose concerning the application of work being done at the Laboratory. Specifically, some were asking, "How do the calculations relate to everyday situations or problems?" An answer to this concern came from Theoretical Division's researchers who began investigating air pollution transport in street canyons created by buildings of various sizes.

The transport of pollution in the air can be significantly affected by the presence of terrain sculpture of buildings. The dispersal properties of a noxious plume depend quite strongly on the position of the source relative to the surrounding structures; for example, from a source in a valley surrounded by hills. Likewise the distribution of carbon monoxide from automobile exhaust depends in complicated fashion on the modifying effects that adjacent buildings have on both the mean wind pattern and the turbulence.

The primary source of information on these effects is derived from experimental observations, both in the field and from laboratory models. To correlate these observations requires a combination of both theoretical and empirical reasoning. The value of such rules of correlation is maximal if they can be applied to a wide range of circumstances, but at the same time are simple and convenient to use. Confidence in their wide applicability requires extensive testing with as much data as possible, while simplicity and convenience imply formulation in terms of short analytical or tabular expressions.

Street canyon problems involve very complex phenomena. For a complete understanding of air pollution transport, much more experimental data will have to be obtained and analyzed. These data will need to be extensive in content, not only measuring pollution distributions, but velocity, temperature, and turbulence intensity distributions as well, with significant spatial resolution to account for the complex interactions that occur between buildings and the air. Even with this large amount of data, the fact still remains that no two street canyons are exactly alike, thus implying that the categorization of street canyon results will require more than a single extensive experiment. Not only are data such as these difficult to obtain, they are also very expensive. [1]

The use of three-dimensional (3-D) computer programs to model such flows can greatly alleviate the need for many of the costly experiments and reinforce the experiments that are needed, by indicating the basic structures of these flows and their related implications to pollution dispersal. Data generated in this way not only familiarize the experimentalist with the flow patterns for which he must be concerned, but allow him to concentrate his efforts on those portions of the street canyon for which there is interest, with cognizance of the complications of the nearby flow. Data generated in this way also give the analyst information with which to compare and extend his analytic models of these complicated phenomena at significantly reduced costs. [2]

A computing technique for low-speed fluid dynamics has been developed for the calculation of 3-D flows in the vicinity of one or more block-type structures. The full time-dependent Navier-Stokes equations are solved with a finite difference scheme based on the marker-and-cell method, see the article “The Birth of Computer Fluid Dynamics” in the chapter “Numerical Analysis and Algorithms” in this book. Effects of thermal buoyancy are included in a Boussinesq approximation. Marker particles that convect with the flow can be used to generate streak lines for flow visualization, or they can diffuse while convecting to represent the dispersion by turbulence of particulate matter. The vast amount of data resulting from these calculations has been rendered more intelligible by perspective view and plots of selected velocity and marker particle distributions. (See Fig. 1.)

These early calculations have formed the basis of much larger investigations of pollution transport across large cities, serving as a basis for various processes by which to reduce the adverse effects of noxious materials through the careful placement of potential sources or the identification of existing sources that used mitigation control of the effluents.

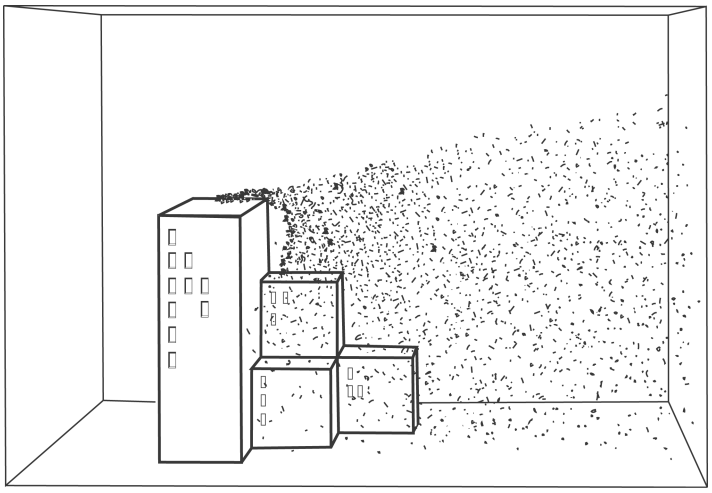


Fig. 1. Computer simulation of the dispersal of pollutants emanating from the top of a building. This and similar calculations show the foundations of pollution transport investigations developed in T-Division.

It [a consulting trip to the wartime Oak Ridge, Tennessee, plant] was great. The lieutenant takes me to the colonel. . .The colonel says, . . . "Okay, Mr. Feynman, go ahead." So, I sat down and I told them all about neutrons, how they worked, da da, ta ta ta, there are too many neutrons together, you've got to keep the material apart, cadmium absorbs, and slow neutrons are more effective than fast neutrons, and yak yak—all of which was elementary stuff at Los Alamos, but they had never heard of any of it, so I turned out to be a tremendous genius to them.

—Richard P. Feynman [Nobel Laureate and former T-Division scientist], "Los Alamos from Below," in *Reminiscences of Los Alamos: 1943–1945*, p.122.


n
p
n
p
p
n
p
n
p
n
p
n
p
n
n
n
n

Introduction to Fluid Dynamics

by Francis H. Harlow

The deforming motion of materials is an integral part of many scientific and engineering processes. Scientific examples range from the slow transport of magma in the Earth's mantle, to the faster dynamics of the turbulent ocean and on to the ultra-high-speed behavior of plasmas in an exploding supernova. Engineering examples abound: the fuel-air mixture in an internal-combustion engine and the flow of air around an airplane are only two of the numerous circumstances in which the contortion of materials is a central issue in the analysis of performance.

The deformations of gases and liquids are among the most readily observed processes involving the evolution of material distortions. Rubber bands and muscles exemplify materials that are not strictly either gases or liquids, but nevertheless exhibit commonly observed deformations. Even solids can respond to strong forces by deforming in remarkably complex fashion. A crash between automobiles is an obvious example; the response of a chunk of metal adjacent to the detonation of a high explosive certainly can result in large distortions of the materials (both in the metal and in the detonating explosive material). Some of these more exotic material-deformation processes are discussed in the "Material Properties" chapter in this book.

In this chapter we discuss some of the fluid-dynamics problems that have been studied in Theoetical Division. Numerical and analytical techniques for solving fluid-dynamics problems are discussed in the "Numerical Analysis and Algorithms" chapter in this book.

From a theoretical viewpoint the most extensively studied fluid flows are those of water or air moving at speeds much less than the sound speed in the material. The material is often described as "incomprehensible," and in many circumstances this is an accurate characterization. More relevant is the classification based on Mach number, which is the ratio of material speed to sound speed. Typical sound speeds vary from about 25 meters per second (mps) in a churning mixture of equal parts water and air, to 330 mps in ordinary air alone, to as high as about 5000 mps in a typical metal.

Many gas or liquid flows in everyday engineering experience take place at very low Mach numbers. Some of them, like pumping air into a tire, are compressible, but nevertheless at very low Mach numbers. Even the flow of air around an airplane behaves mathematically like a low-Mach-number deformation for speeds up to about 300 miles per hour.

Low-Mach-number deformations, like water dynamics in a hydraulic pumping system, have a large body of literature describing theoretically and experimentally the behavior of the liquid or gas as it moves through pipes and valves or around obstacles. The mathematical techniques are especially tractable if the flow is laminar, i.e., free of turbulence. Even turbulent flows, as discussed in this chapter, have been subjected to intensive study because of their relevance to almost any circumstance in which the pressure forces are large compared with the viscous forces.

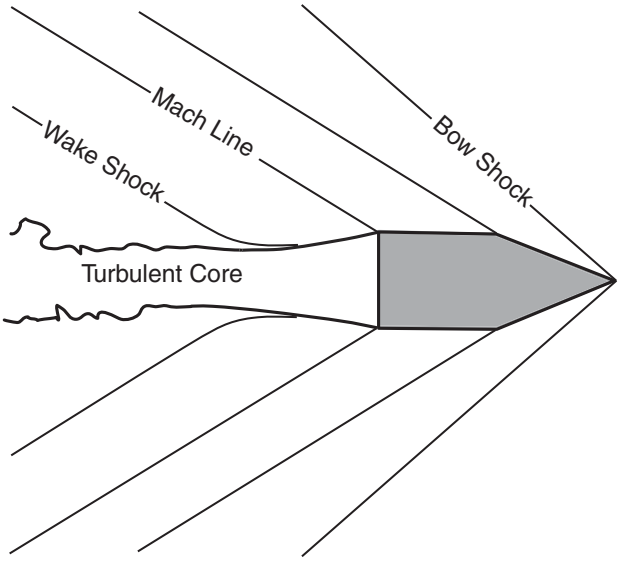


Fig. 1. The illustration shows shocks folding back from the tip of a speeding bullet or supersonic airplane, for example.

Flows at intermediate Mach numbers have many properties in common with those at low Mach numbers. In particular the theoretical analysis in both cases is based completely on the fundamental laws of mass, momentum, and energy conservation, together with "constitutive" relations that characterize each material's pressure variations as a function of density and temperature. When viscosity is important, then that process also requires a characterization that can be dependent on density and temperature.

High-Mach-number flows, however, can also exhibit behavior that is very different from the things that are observed at low Mach numbers. In particular, the presence of propagating shock waves and rarefactions is virtually ubiquitous. A speeding bullet or supersonic airplane has shocks folding back from its tip (Fig. 1); likewise, an explosion can emit a diverging shock wave. In both cases, the approach of the shock wave to an observational instrument is undetectable until the shock wave arrives. (In contrast, the approach of a subsonic airplane can be heard by an observer standing well ahead of the flight path).

The T-Division has been much concerned with these topics of fluid and gas flow. Two other types of fluid dynamics have also received major attention in T-Division. One is the behavior of two or more materials interpenetrating through each other, like bubbles in a liquid or sand grains in a gas. The other is the development of techniques for understanding and modeling fluid turbulence, which has been characterized as one of the most difficult problems in classical physics.

Understanding Fluid Instability and Turbulence

by Francis H. Harlow

Besnard
T. Clark
Daly
Fisher
Harlow
Hirt
Kashiwa
Linn
Meshkov
Nakayama
Plesset
Rauenzahn
Richtmyer
Romero
Spitz
Steinkamp
Turner
P. Wilson
Zemach

Instability means the tendency for a smoothly flowing fluid to develop fluctuations in velocity, which introduces an irregularity, or churning, into the flow. Whenever a region of fluid develops these fluctuations, we say that the flow is turbulent. Examples are abundant in nature: gusty winds in the atmosphere, roiling currents in a stream, or restless motions of the gases in the sun. In the home, the laboratory, and numerous industrial processing plants, the presence of turbulence is frequently manifested by either beneficial or detrimental effects. Mixing of materials, for example, can be dramatically sped up, but at the same time air turbulence can significantly decrease the fuel efficiency of a vehicle moving through air or water.

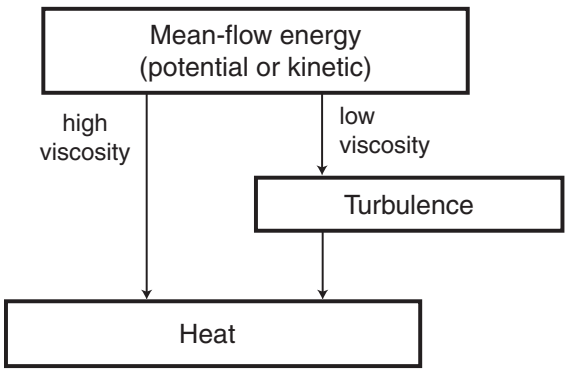
To describe the processes taking place in a turbulent flow, the use of computer calculations is almost inevitably required. But today's supercomputers are not large enough or fast enough for the calculations to describe every turbulent fluctuation. Thus, mathematical models are required to catch the overall essence of the complex flow structure. Much work on developing these models has taken place at major scientific and engineering research institutions. When the Theoretical Division first realized the importance for modeling turbulence and its effects, however, much work still needed to be done. Within a few years, the developments of useful models put T-Division at the forefront of this activity. The use of T-Division modeling techniques is now found in virtually every scientific and engineering laboratory in the world.

Background

Most circumstances of fluid flow are unstable. The basic principle behind this inevitable trend lies in an extended version of the second law of thermodynamics: all processes in nature tend to become progressively more chaotic; i.e., the generalized (informational) entropy increases. In fluid flows, two processes compete. (See Fig. 1.)

If the viscosity of the material is sufficiently high (i.e., the Reynolds number is less than about 50), energy transforms directly into heat. With less viscosity, instability can divert the energy into small-scale fluctuations, called turbulence, which only then can dissipate their energy into heat.

Fig. 1. The illustration shows transformation of energy.



As long as the flow is being driven, the largest scales of turbulent motion are continuously breaking into smaller scales, which in turn can coalesce back into scales that are intermediate between the driving scale and the finest scales. A classic example occurs when two parts of a fluid are shearing against each other beyond the tip of a plate. Small vortices are produced at the end of the splitter plate; but, through the process of vortex pairing, the dominant turbulence scale increases downstream. (See Fig. 2.)

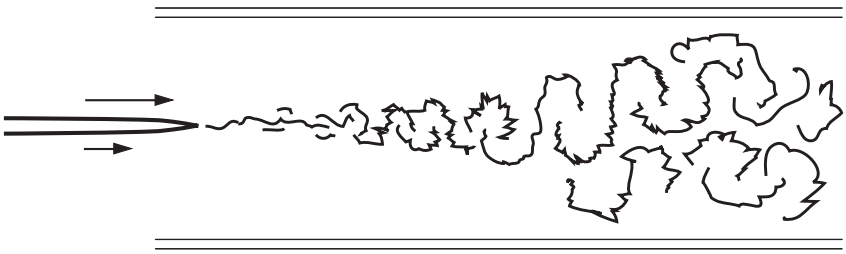


Fig. 2. The illustration shows flow past a splitter plate.

For many scientific and engineering analyses, understanding the behavior of every vortex, from small to large, is not necessary. The relevant question, from a practical problem-solving point of view, is, "How can a mathematical model of these processes be constructed and implemented into computer codes so as to capture the essence of what is going on without a detailed description of every part of the motion?" Direct numerical solutions of the flow equations can give important information about the fine-scale flow structures, but such solutions are too expensive to use for most large-scale scientific and engineering investigations.

The situation resembles that of molecular dynamics (See the "Condensed Matter" chapter.) in which the computer follows the motion of every molecule. The essence of what is happening in a large-scale material flow, however, does not require knowledge of all the details. The effects can be described by bulk quantities, such as heat, pressure, density, viscosity, and by the equations that describe the transport of mass, momentum, and energy.

In an incompressible fluid with turbulence, the dominant consequences of the fine-scale motions can be described by a tensor called the Reynolds stress. To describe the properties of Reynolds stress, we imagine that the fluid velocity and pressure are decomposed into average values plus the instantaneous departures from average. For example, the pressure becomes

$$p = \bar{p} + p' \quad ,$$

while the i^{th} component of velocity is

$$u_i = \bar{u}_i + u' \quad .$$

then the Reynolds number is

$$R_{ij} = \overline{u_i' u_j'} \quad .$$

The beauty of this definition of Reynolds stress is that R_{ij} describes both the energy carried by the turbulence and the resisting force when a turbulent fluid is being sheared or moved along a wall.

The challenge is to derive a Reynolds-stress transport equation that resembles a heat-flow equation. This task presented investigators in T-Division with a significant challenge. The Reynolds-stress transport equation was known, in principle, long before T-Division involvement. In practice, however, its form was incomplete, in the sense that the derivation for Reynolds-stress transport introduces additional unknown quantities. In a never-ending process, the transport equations for these additional quantities can be derived; but even more unknowns are introduced. By some process or other, the system of equations has to be closed.

To address this closure challenge, a subgroup of T-Division investigators was formed, which subsequently grew to include many contributors. The initial people working with me during the latter half of the 1960s included Bart Daly, Tony Hirt, Paul Nakayama, and Nick Romero. Over the following years, the turbulence work in T-Division led to several Ph.D. dissertations and international collaborations. The later T-Division participants working with me included Didier Besnard, Tim Clark, Bart Daly, Bucky Kashiwa, Rod Linn, Rick Rauenzahn, Patrick Spitz, Mike Steinkamp, Leaf Turner, Peter Wilson, and Chuck Zemach.

The K-ε Model for a Single Incompressible Fluid

Three new variables are introduced in the derivation of the Reynolds-stress transport equation. These are the rates of (1) turbulence self-diffusion, (2) the continuous trend to isotropy of the tensor, and (3) the dissipation of Reynolds stress. The last of these was a principal focus for the team during the latter half of the 1960s.

Half of the contraction of the Reynolds stress is called K, which has the meaning and dimensions of energy per unit mass. The time rate of K decay is called ε, which therefore has the dimensions of energy per unit mass per unit time. Accordingly, ε/K is a relaxation rate (reciprocal time) and can be used to describe the rate of return to isotropy and the rate of turbulence self-diffusion. Thus, an understanding of ε was a major issue to be resolved for the successful representation of all three unknowns in the Reynolds-stress transport equation.

The realization that $K^{3/2}/\varepsilon$ has the dimension of length led us to associate that quantity with the dominant scales of the flow structure. But this association seemed paradoxical because we believed that the decay rate for K is a purely viscous process, taking place at the smallest scales rather than the dominant scales. (This paradox was resolved some years later, as discussed below.)

We chose the approach for the evolution of ε to be a transport equation for that quantity. At that time, we had two ways to accomplish the derivation: (1) to develop a transport equation for the viscous dissipation rate directly from the basic equations for fluid dynamics or (2) to develop, through experimentation, a transport equation closely resembling the one for K itself, with care being taken to ensure the proper dimensions of the terms.

We tried both approaches; and our results could be made to resemble each other, if the paradoxical association of energy decay with dominant scales could somehow be resolved. This paradox came from the presence of molecular viscosity in the decay term. However, it is well known that at high Reynolds numbers (small viscosity) the rate of turbulence decay is independent of the molecular viscosity. The arguments we used to resolve the paradox missed the point that the decay of turbulence at high Reynolds numbers arises from a cascade of energy from large vortices to small ones. Because the large scales of fluctuation are the dominant source of bulk effects, the result of this cascade is loss of strength for bulk momentum and energy transport, i.e., a decay of Reynolds stress. Spectral techniques that are described below show the basis for these concepts.

Thus, the idea for two transport equations, one for K and one for ε, was developed in T-Division and subsequently has seen extensive use at laboratories, universities, and industrial organizations throughout the world. This work, together with the pioneering T-Division developments of numerical fluid dynamics, described in this book in the "Numerical Methods and Algorithms" chapter, received the Japan Society of Mechanical Engineers Computational Mechanics Award for 2001.

T-Division researchers soon followed this work with a generalization to the transport of the full Reynolds-stress tensor and an associated extension of ε transport to that of a full decay tensor.

The Besnard-Harlow-Rauenzahn Formulation

If density varies significantly throughout the fluid, then differential acceleration from pressure gradients can drive interpenetration between the high- and low-density fluids. The T-Division team investigated two approaches to the description of this additional turbulence source. Both approaches required the introduction of additional variables.

In the multifield approach, volume fractions and velocities for each of the fluids represent the interpenetrating materials. (See the article, "Multifield Flow," in this "Fluid Dynamics" chapter.)

In the Besnard-Harlow-Rauenzahn (BHR) approach, the additional variables for two-material circumstances are measures of the density fluctuations and of the net mass flux relative to a coordinate system in which no material-volume flux occurs.

The procedure for deriving transport equations is an extension of the procedure described for a single material. In addition to transport equations for a generalized form of the Reynolds stress and its decay term, the density fluctuations and the mass flux also have transport equations for which closures are required.

A notable example of an application for this BHR description is to the unstable mixing of a heavy fluid on top of a light fluid, the Rayleigh-Taylor configuration carried to the limits of late times. Comparison of the BHR results to those of experiments shows the occurrence of a late-time self-similar behavior that is the same as that measured in the laboratory.

We calculated numerous other examples, including the motion of particles and bubbles in a fluid, of droplets or dust in the air, and of the high-speed instability and mixing that occur when a high-explosive shock encounters an interface between two metals. The BHR technique has been used for various engineering-analysis purposes in the United States, France, and Russia, where a team at Arzamas-16, the Russian nuclear weapons laboratory at Sarov, independently developed a similar set of ideas.

The Steinkamp-Clark-Harlow Formulation

Despite the success of the BHR formulation for variable-density turbulence, we soon recognized deficiencies in rapid transients of drive. French scientists had also observed these difficulties and had developed a spectral model for analyzing the behavior of turbulence in a constant-density fluid.

At Los Alamos, we developed the Steinkamp-Clark-Harlow formulation for two interpenetrating fluids, an extension to the French procedure. The idea was to start from the basic equations for fluid dynamics and to develop equations that simultaneously describe the dynamics at two points in space. A Fourier decomposition of the equations relative to the separation leads to a spectral formulation, for which we quickly realized that an equation for the transport of ε was no longer required. We then showed that the single-point transport equation for ε could be derived for turbulence circumstances in which the spectral structure had settled into a self-similar form. We thereby finally resolved the paradox related to the independence of turbulent decay on the fluid viscosity. For more general problems in which rapid drive transients preclude self-similarity, the turbulence could be described very well by using the full spectral transport equations.

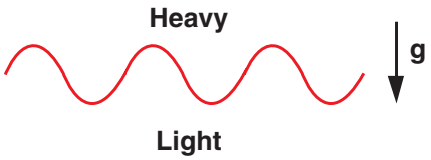


Fig. 3. The illustration shows a heavy fluid over a light fluid.

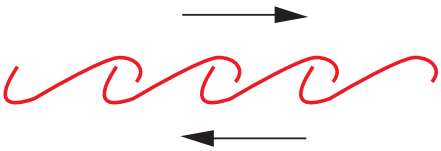


Fig. 4. The illustration shows shear instability.

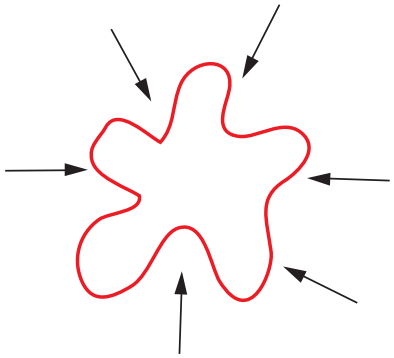


Fig. 5. The illustration shows converging instability.

Four Classic Instabilities

In addition to the development of turbulence-transport descriptions in T-Division, two classic fluid-instability circumstances (Richtmyer-Meshkov and Bell-Plesset) bear the names of T-Division scientists. Four of the most widely known instabilities bear double names: (1) Rayleigh-Taylor instability (example: a heavy fluid over a light one, Fig. 3), (2) Kelvin-Helmholtz instability (example: two fluids shearing against each other, Fig. 4), (3) Richtmyer-Meshkov instability (example: a shock hitting a perturbed interface between two fluids), and (4) Bell-Plesset instability (example: the interface between two fluids converging toward the center of a sphere, Fig. 5).

Robert Richtmyer of T-Division first analyzed the third of these instabilities; he described his findings in a famous report dated July 1954. Experimental work by Evgueni Meshkov at Arzamas-16 confirmed the nature of this instability, which has subsequently been the subject of numerous experimental, analytical, and numerical investigations.

The fourth instability received its first attention from George Bell of T-Division, and his analysis was published in a Los Alamos report in November 1951. The ideas were subsequently extended by Milton Plesset and were published 3 years later. Henry Fisher generalized the theory to include the effects of compressibility, as described in a report of September 1982. In recent years, researchers have demonstrated significant experimental confirmation of this unstable interface growth, with convergences that are driven by explosives, by pulsed power, and by lasers. The late, highly nonlinear stages of the instability growth have been called crenulation, which is exhibited most dramatically by the explosive-driven convergence of a polycrystalline metal, for which the experiment was designed in T-Division.

In T-Division, the Rayleigh-Taylor and Kelvin-Helmholtz instabilities were calculated for the first time using fully resolved direct numerical simulations. In the 1960s, the publication of these results went a long way toward establishing computer studies as a viable way to produce meaningful descriptions of complex processes.

Additional Instabilities

The physics associated with instability growth is always the same: large-scale energy is transformed into small-scale energy. The basic driving mechanism is associated with the creation of localized pressure gradients in the mean flow. The pressure gradient produces differential acceleration between the elements of a nonhomogeneous material. The gradients in shear instability arise through Bernoulli's law and accelerate the perturbed interface.

In each case, the result is turbulence, defined as the condition in which energy is contained in flow scales that are much smaller than that of the mean flow. In addition to the four classic examples described above, T-Division has made major contributions to understanding the instability of flow past a cylinder (Fig. 6), which results in a von Karman vortex street, to the behavior of two-material interpenetration (Fig. 7), and to the effects of shock passage over a polycrystalline metal (Fig. 8).

Current explorations include the study of radiation interactions with unstable interfaces and the effects of non-Newtonian constitutive characterization for the materials.

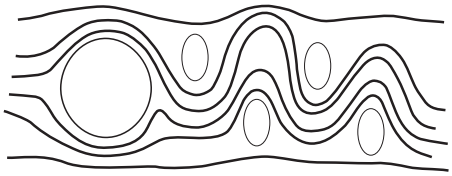


Fig. 6. The illustration shows von Karman vortex street.

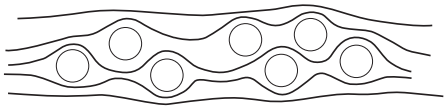


Fig. 7. The illustration shows multifield turbulence.

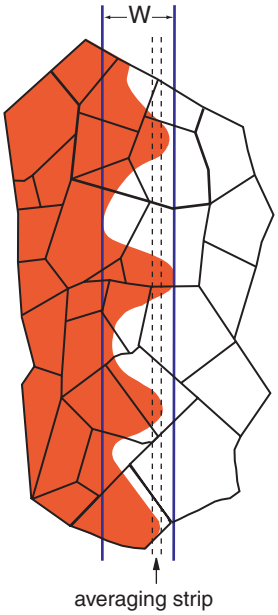


Fig. 8. The illustration shows shock over a polycrystalline metal.

Navier-Stokes-alpha Model of Fluid Turbulence

Chen
Foias
Geurts
Holm
Margolin
Marsden
Olson
Ratiu
Titi
Wynne
Zhang

by Darryl D. Holm

Theoretical Division collaborators at Johns Hopkins University, University of Indiana, California Institute of Technology; the University of California, Davis; Twente University; and Weizmann Institute use a new turbulence closure model for fluids called the Navier-Stokes-alpha model, or alpha model. Theoretical Division developed the model to play a fundamental role as the turbulence component in large-scale numerical simulations of global, coupled ocean-atmosphere dynamics for predicting changes in the Earth's climate in the Laboratory's Climate, Ocean, and Sea Ice Simulation (COSIM) program.

The recent authentication stages of the new turbulence closure model for fluids are summarized. The alpha model's promise for future programmatic applications was verified and validated by analyzing its mathematical properties, testing its numerical performance, and comparing its predictions with experimental data from turbulent flows. For a review of this model and the history of its development, see Foias, Holm, and Titi. [1]

Rather than attacking the full turbulence closure problem, we attacked the much more restricted Computational Turbulence Closure Problem (CTCP). The problem statement for CTCP is to "derive a computable turbulence model from first principles that gives accurate results (as compared to experiments) at the resolvable length scales (for a given computational, or experimental capability)."

We attacked this problem by exploiting a new analytical approach based on the work of Holm, Marsden, and Ratiu [2] that applies both to numerical algorithm development and to physical modeling of turbulence in fluids. This approach resulted in a mathematical regularization of the Navier-Stokes equations that was derived from a physical parameterization of the average effects of turbulence on the resolvable length scale. This parameterization (also a regularization) was obtained by using Lagrangian averaging, i.e., averaging that follows the fluid parcels, rather than averaging at a fixed Eulerian position. The Lagrangian averaging approach resulted in a nonlinear system of flow equations called the "Navier-Stokes alpha model," or "alpha model" for short. As it has developed over the past few years, the alpha model has showed increasing promise as a viable partial solution to the CTCP.

The CTCP requires the development of a computable, reliable model of turbulence in fluids. There are several strengths to the alpha model's approach:

(1) It derives from first principles, namely the generalized Lagrangian mean (GLM) theory of Andrews and McIntyre [3] and from the Taylor hypothesis [4] that turbulent fluctuations are frozen into the mean flow. Separately, these are classical methods and ideas in turbulence theory. However, they had not been previously combined into a closure model until they were used to derive the alpha model.

(2) It allows the integration of theory, modeling, and numerics into the same approach, based essentially on Hamilton's variational principle for fluid dynamics as developed in Holm et al. [2]

(3) It is general enough to add more physics and thereby may be applied to physically realistic situations.

Our target for the alpha model is its application as the turbulence component in large-scale numerical simulations of global, coupled ocean-atmosphere dynamics for predicting changes in the Earth's climate. Future targets for its application are in compressible turbulence, particularly in the problem of turbulent mixing. Some early indications that this model may eventually be successful in this application have recently been found at low Mach number.

"Two-thirds" Scaling and Computational Speedup

The promise of the alpha model approach for future applications is its "two-thirds" scaling. For example, if direct numerical simulation (DNS) of the Navier-Stokes equations requires three decades in lengthscale to resolve the main features of a given turbulent flow, then the alpha model will only need two decades. This two-thirds scaling is borne out in practice and is a mathematical theorem for the alpha model in the case of homogeneous turbulence. [1] In terms of Reynolds number (*Re*, a measure of turbulence intensity), the Navier-Stokes equations possess $Re^{9/4}$ active degrees of freedom.*

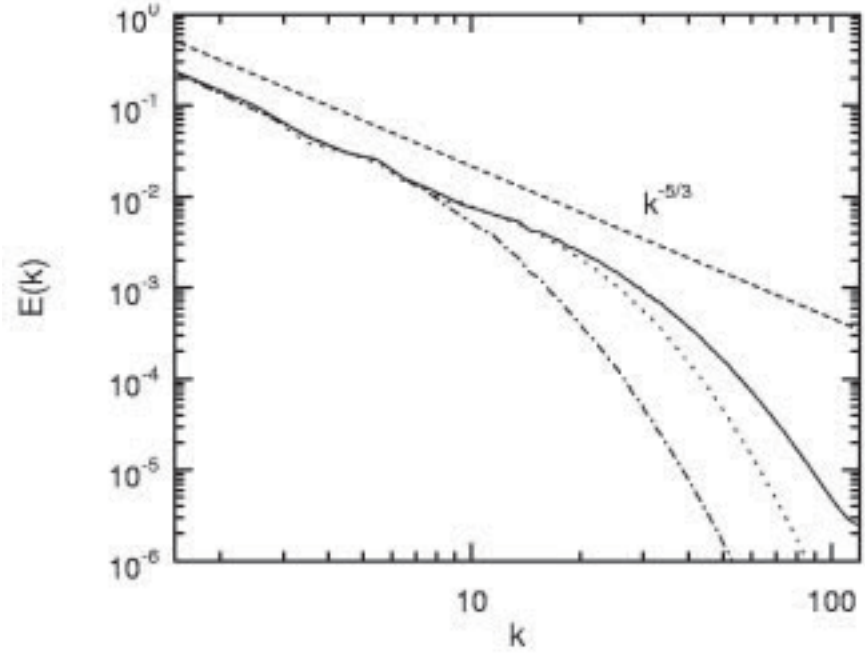


Fig. 1. The direct numerical simulation (DNS) energy spectrum, E(k), versus the wave number k for three cases with the same viscosities, same forcings, and mesh sizes of 256^3 for a = 0 (solid line), 1/32 (dotted line) and 1/8 (dotted-dash line). In the inertial range k < 20, a power spectrum with $k^{-5/3}$ can be identified. For finite alpha, this behavior is seen to roll off to a steeper spectrum predicted to be k^{-3} for ka > 1. This k^{-3} roll-off in the energy spectrum of the alpha model is responsible for its two-thirds scaling properties. The roll off starts at k = 1/alpha.

*This is Landau's famous scaling estimate of the number of degrees of freedom in a turbulent flow, based on $(Re^{3/4})^3 = Re^{9/4}$ using Kolmogorov's dissipation length scale ($\approx Re^{3/4}$) for homogeneous turbulence.

In comparison, the alpha model will excite only the largest [$(Re^{9/4})^{2/3} = Re^{3/2}$] active degrees of freedom. This reduction in dimension occurs because the alpha model alters the nonlinearity of the Navier-Stokes equations so that the balance between transport and dissipation occurs at a dissipation length scale proportional to $Re^{-1/2}$, which is larger by a factor of $Re^{1/4}$ than Kolmogorov's dissipation length scale ($\sim Re^{-3/4}$), where this balance occurs for the Navier-Stokes equations. The reduction in degrees of freedom results in a computational speed-up for the alpha model compared with direct numerical simulations (DNSs) of the Navier-Stokes equations that may be estimated by applying the factor of $Re^{1/4}$ in all three dimensions, as well as in time. This estimated computational speed up by the factor of Re for the alpha model compared with DNS is borne out in practice, without loss of accuracy at the resolved scales larger than alpha. For a comparison of the alpha model's performance relative to DNS, see Chen et al., [5] from which Fig. 1 is taken.

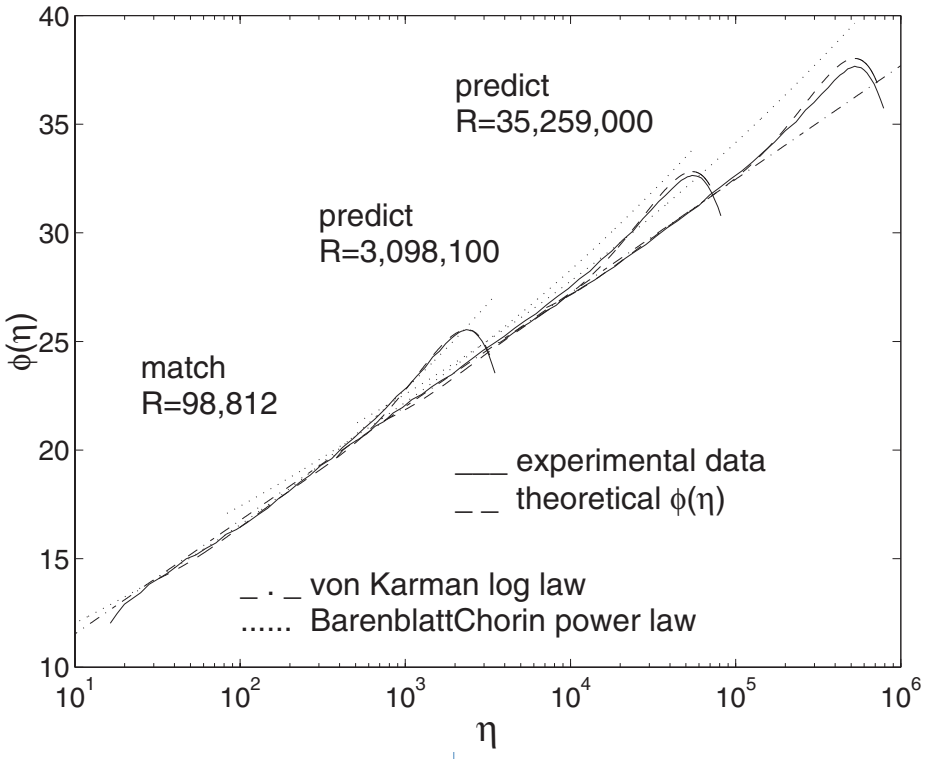


Fig. 2. Steady solutions of the Navier-Stokes-alpha model agree well with experimental pipe flow data at high Reynolds numbers. [6]

We accomplished several things in our research.

(1) We verified the accuracy of the model by comparing steady flows with experimental data for mean velocity in wall-bounded flows. These included flows in pipes, channels, boundary layers, and at stagnation points. Figure 2 shows the mean velocity profile for turbulent flow in a pipe for the Navier-Stokes alpha model compared with the composite experimental data of Zagarola [6] over a range of scales covering nearly six decades as measured in natural "wall units" (distance away from the wall). See Chen et al., [7] from which Fig. 2 is taken.

(2) We performed numerical simulations for turbulence in periodic domains. These included forced homogeneous turbulence that showed the alpha model properly predicts turbulence statistics at length scales larger than alpha, [5] and initial value problems investigating how the turbulence energy cascade is established under alpha model dynamics.

(3) We studied geophysical flows to evaluate the effects of rotation on the predictions of the alpha model. This tested the alpha model's viability as the potential basis for a climate model.

(4) We also studied free unbounded flows with the alpha model. This included numerical computation of turbulent mixing layers in unstable shear, and analysis and comparison with experiments of similarity solutions of the alpha model for turbulent wakes, jets, and plumes. [8]

Our major theoretical achievements and results showed the following:

- Strong solutions of the Navier-Stokes-alpha model exist globally; they are unique; they converge to solutions of the Navier-Stokes equations as alpha tends to zero; and they possess a global attractor with finite fractal dimension that is consistent with the $Re^{3/2}$ scaling for the number of active degrees of freedom in the alpha model [1].

- The finite-dimensionality of the global attractor was found to be consistent with the Karman-Howarth (KH) theorem for the velocity autocorrelations of the alpha model. This KH theorem implies the energy spectrum for the alpha model will scale with wave number as the – 5/3 power for separations or length scales larger than alpha and that this wave number scaling will follow the – 3 power for separations or length scales smaller than alpha. This change in the energy spectrum proved in this KH-alpha theorem reflects the two-thirds scaling property of the alpha model.This two-thirds scaling property enables calculations with the alpha model to obtain accurate turbulence results (at sizes greater than alpha), while requiring only two decades of resolution, for the same turbulence problems that would require three decades of resolution in direct numerical simulations using the exact Navier-Stokes equations.

Multifield Flow

Amsden
Besnard
Butler
Cloutman
Cook
Daly
Demuth
Dukowicz
Harlow
Hirt
Kashiwa
Mjolsness
O'Rourke
Pracht
Ramshaw
Rivard
Romero
Ruppel
Torrey
Travis

by Francis H. Harlow and Bryan A. Kashiwa

Raindrops falling through the air, bubbles rising in a liquid, and sand setting in either a liquid or a gas are three examples of multifield fluid flow. Detonation of a heterogenous high explosive and the burning of a fuel spray in an internal combustion engine are also examples. In fact, many of the problems in industry, academia, and defense involve multifield flows of some sort. The main feature distinguishing multifield flows from single-field flows is the presence of a mean relative motion: raindrops, bubbles, and sediments can all move with a local velocity that is quite different from that of the surrounding fluid. The physical basis for this velocity difference is the inertial response to a force—materials of differing mass density undergo dissimilar accelerations when subjected to the same force field, as from a pressure gradient. The common occurrence of multifield flows, and the attendant technical challenges, produce a fertile subject area for both basic and applied studies. Consequently, the literature is very broad in the area of multifield flow, also called two-phase flow, multiphase flow, two-field and sometimes multimaterial flow, in addition to many material-specific terms such as gas-solid flow.

The first significant multifield investigations in the Theoretical Division came during the energy crisis days of the early 1970s as a result of urgent safety questions regarding pressurized water reactors (PWR). The Nuclear Regulatory Commission (NRC) needed to understand the potential consequences of a hypothetical accidental rupture of the highly pressurized water lines. The consequences of such an accident come from the flashing of some of the water to steam, thus creating a two-field flow with the tremendous capability to cause major damage.

The NRC contract managers had looked at several organizations for technical guidance, but they found that progress in significant problem solving was impeded by concerns for the "ill-posedness" of the two-field flow equations. Resolution of this issue was a major challenge; in T-Division the challenge was resolved, however, by the simple introduction of a dissipative mechanism in the equations, which works very well for solving meaningful problems. For T-Division, this introduction to multifield flow opened the door to numerous applications in many areas besides that of reactor-safety analysis. Examples include blood cells moving through an artery, catalytic particles in a chemical-reaction container, rupture and coalescence of rain drops, volcanic eruptions with solid chunks in a hot dusty gas, sprays of fuel droplets in an internal combustion engine cylinder, the propagation of a wildfire through grasslands or trees, and coal-slurry transport through a pipe.

Our concern in multifield flows is for circumstances in which it is essentially impossible to calculate the dynamics of every entity and its surrounding fluid. Thus modeling is required, leading to the concept of local "averages" for the behavior of entities and fluid. It is for these averages that we use the word "field."

The description of physical processes in multifield flow includes, first of all, the physics of mass, momentum, and energy conservation for the combined motions, and interactions of all the entities (grains, drops, bubbles, etc.) with the surrounding fluid and with each other. Exchanges of mass, for example, arise from the melting of the entities, vaporization, burning, or condensation. These processes are intimately associated with exchanges of energy through convective heat exchange, conduction, and radiation, and with the release or absorption of latent heat or chemical energy.

Velocity differences between the fields can be created through differential acceleration arising from pressure gradients in the fluid. A grain of sand in water, for example, receives less acceleration because of its greater inertial resistance to moving. A closely related case occurs with gas bubbles in a liquid; their small inertia would result in tremendous acceleration by a pressure gradient, if not for the necessity to also move the surrounding liquid to make way for the passage of the bubble. This process is described by including the so-called "virtual mass" of the bubble, which is modeled as if the mass of the bubble includes some of the mass of the surrounding fluid. In addition, there is an exchange of momentum between the fields produced by drag forces that arise from both pressure gradient variations and the action of fluid viscosity.

Another type of physical process takes place within the entity field itself. Collisions can occur between entities, resulting in exchanges of momentum, in fragmentation or coalescence, and in the creation of an effective viscosity (diffusivity of momentum) within the entity field.

All of these processes have to be addressed as the need arises for each new application. Much interaction with other laboratories has resulted in significant sharing of ideas. The multifield flow studies in T-Division, taking place over many years, have also profited from the opportunity to validate our calculations by comparing them with experimental data generated by many investigations.

Numerous difficulties have been examined in T-Division for developing the formulation of descriptive equations for multifield fluid flow. The resolution of some of them has been described in reports and papers. Questions that have been partially or completely answered in T-Division include:

(1) How should the turbulence be described? Suggestions have been advanced and tested, but much remains to be done. (See also the article on turbulence in this chapter.)

(2) What happens during a transition from one field being the fluid and the other field being the entities to a complete reversal of these identities? For example, an expanding flow of air bubbles in water can become a flow of water droplets in air.

(3) How should the effects of a distribution of entity sizes be described? Instead of a single velocity field to describe the dynamics, there is a distribution of velocities, for which one needs to consider the evolution of a probability distribution function to describe what is happening.

(4) How should the dynamics be described as the entity field approaches and reaches a close-packed state?

(5) What technique should be used to represent the effects of entity distortions from spherical to multiply connected networks of filaments?

(6) Can multifield descriptions be extended to circumstances in which the surrounding "fluid" behaves in a visco-elastic fashion as a result of stress application at both low-and high-strain rates?

(7) When one or more of the fields is compressible, what is the relation between the pressure in that field to those quantities in the nearby regions of the other field?

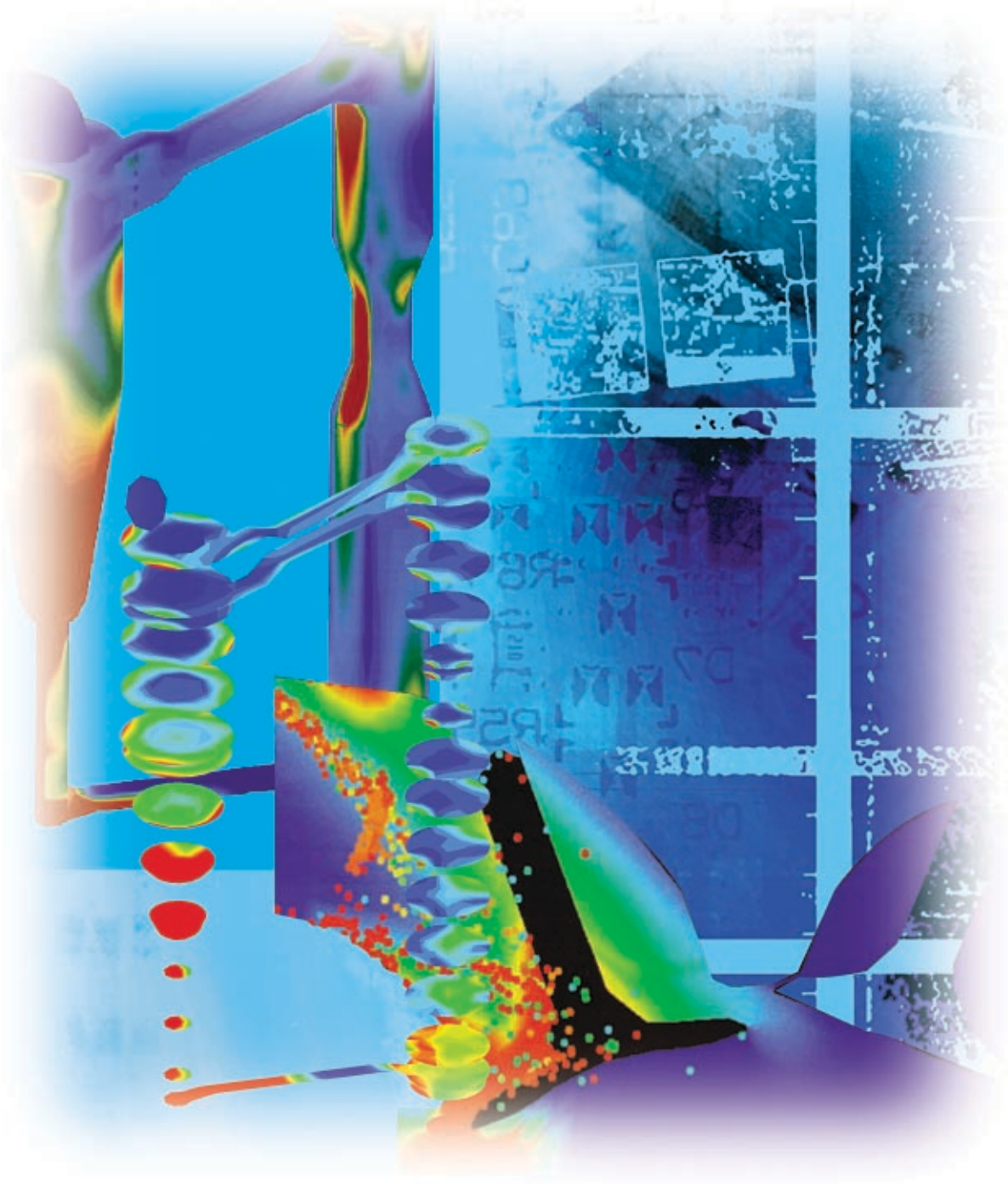
The first contribution from T-Division was to "cure" the ill-posedness by adding a small amount of viscosity to the equations in such a way as to bound the growth rate of perturbations for vanishingly small scales, but leave the physically meaningful instabilities essentially unaltered for scales of the order of entity size and larger. These meaningful instabilities represent the tendency for interpenetration to result in turbulence, which later was modeled by the incorporation of turbulence transport equations. The second contribution was to write a computer code specific to the needs of the NRC, and to start solving real problems with significant relevance to PWR safety issues.

The PWR safety analysis activities ultimately led to a full-system computer program that included all of the components of the full configuration of the reactor, including pumps, heat exchangers, valves, and especially the reactor itself, including downcomer, core, and upper plenum. Interactions with some of the solid components were also necessary to be included, because the violence of some of the postulated scenarios could tear apart the solid structures.

The reactor safety work remained principally in T-Division for several years, but eventually spread to other Laboratory groups, and indeed formed the basis for creation of another Laboratory division, Q-Division.

In subsequent years many major activities in T-Division revolved around the numerical-analysis power that was enabled by the continuously evolving numerical techniques. Several of these applications are discussed in other sections of this book, for example the internal combustion engine work described in the "Industrial and Governmental Collaborations" chapter, and the wildfire propagation studies in the "Environmental Studies" chapter.

Many people are associated with the extensive work on multifield theory and applications in T-Division over the years after 1972. These include the following: Tony Amsden, Didier Besnard, Dan Butler, Larry Cloutman, Tom Cook, Bart Daly, Ruth Demuth, John Dukowicz, Tony Hirt, Bucky Kashiwa, Ray Mjolsness, Peter O'Rourke, Bill Pracht, John Ramshaw, Bill Rivard, Nick Romero, Hans Ruppel, Martin Torrey, Jack Travis, and me.



Industrial and Governmental Collaborations

Much of the scientific work in the Theoretical Division involves interactions with universities around the world. Equally significant are the frequent collaborations with American and even foreign industries and with various agencies of the United States government.

Helping to fulfill the Laboratory's mission to apply scientific excellence to problems of national concern, T-Division has engaged in significant collaborations with numerous outside organizations. The division has used several methods to accomplish these interactions: e.g., exchanges of personnel, development of new and better techniques, trouble-shooting, building of new ideas for the future, and in general, making useful contributions to a great diversity of scientific and engineering problems.

In this chapter, we describe a few significant examples of this work and the resulting consequences for problem-solving in numerous disciplines, as well as for broadening the scope of our capabilities and the experience of our staff. For example, we describe the internal-combustion-engine activities that have had a major impact on the industrial goals of increasing fuel efficiency and reducing the release of noxious pollutants into the atmosphere.

Understanding Complex Flows in Real-World Applications

by C. W. Hirt

Looking for funding sources outside the shrinking budget of Los Alamos National Laboratory in the 1970s, the Theoretical Division decided to seek important, real-world applications to assist government agencies with problems involving complex fluid-flow phenomena. The division took projects with agencies such as the Office of Naval Research and lent technical expertise, computers, software, and numerical modeling of fluid-dynamic processes. Today, T-Division continues to work on projects for outside agencies. The reward for this work is the discovery that real-world applications offer a sense of satisfaction not always found in academically oriented studies.

The first serious reduction-in-force at Los Alamos occurred in 1968. What a shock to employees to learn that their continued employment was no longer ensured. In the Theoretical Division the word went out for all group leaders to be prepared to reduce the size of their staff. As a little ray of hope, there was a suggestion that maybe sources of outside funding could be sought to offset the shrinking Lab budget. At the time, however, the concept of "technology transfer" did not exist and outside funding could only be obtained from government agencies and not from private industry.

Seizing this possible route of salvation, the fluid dynamics group in T-Division convinced itself that there must be many government agencies faced with problems involving complex fluid-flow phenomena. After all, the flows of liquids and gases are ubiquitous. Surely project managers in Washington would gratefully snap up the cutting edge technology developed by T-Division.

Presentations were made to agencies where it was easy to imagine important applications of the group's technology. For example, the group made presentations to the Environmental Protection Agency, the Atomic Energy Commission, the Defense Nuclear Agency, the Office of Naval Research, and the Department of Transportation.

At first it was a shock to discover that the people in charge of important government projects had little or no idea about computers, software, or especially the numerical modeling of fluid-dynamic processes. Fortunately, persistence eventually paid off, and a number of contracts were secured.

The scope of projects undertaken was quite broad. For instance, the Office of Naval Research supported the development of better ways to model wave forces on ships. The prediction of pollution levels in street canyons formed by high-rise buildings was a focus for the Environmental Protection Agency. The Air Force Office of Scientific Research wanted computer models to better understand the operation of chemical lasers.

One of the more interesting projects, one that used a portion of the group's technical expertise and put a new spin on an old problem, came from the National Science Foundation. The task was to try to understand why a small percentage of tornadoes are able to cause considerably more damage than the average tornado. The Lubbock, Texas, tornado in the early 1970s was the prototype chosen for study because a large effort had been made to record the nature and extent of the damage.

Before one can understand differences in tornadoes, it is first necessary to have some idea of the basic structure of a tornado. It turned out that not much was really known about the inner workings of these awesome, and sometimes deadly, swirling flows. Building on the group's experience for modeling many kinds of complex flows, a model of a tornado was generated that revealed a completely unexpected phenomenon. Under the proper circumstances, airflow in the core of the tornado may be moving downwards, rather than upwards, as would be expected. The existence of downward core flow has been verified by observations.

Projects for outside agencies have continued to provide interesting challenges for staff members throughout the Laboratory. For T-Division the reward was the discovery that real-world applications offer a sense of satisfaction not always found in academically oriented studies.

Modeling and Simulation with the U.S. Semiconductor Industry

by David C. Cartwright

A 5-year program involving the Los Alamos, Sandia, and Livermore National Labs, SEMATECH, the Semiconductor Research Corporation (SRC), and SRC-supported universities was initiated in early 1995 to help develop advanced modeling and simulation tools for the U.S. semiconductor industry. The 5-year goal of this partnership was to develop a suite of predictive simulators for the sub-0.1 micron technology generation, which the Semiconductor Industry Association (SIA) Roadmap (called the National Technology Roadmap for Semiconductors [NTRS]) forecasted to be introduced into manufacturing in 2007. After visiting all three U.S. Department of Energy laboratories, the SRC chose the Theoretical Division at Los Alamos as the Center for Semiconductor Modeling & Simulation (CSMS) for this combined effort. The program was initiated with Los Alamos scheduled to receive $3 million per year as the lead laboratory, and Sandia and Livermore to each receive $1.5 million per year.

This collaboration had four major technical thrust areas and each thrust area involved one or more DOE laboratories, SRC-sponsored universities, and SRC member company:

(1) 3-D grid and computer science,

(2) topography,

(3) ion implantation and bulk processes, and

(4) materials reliability.

The operational approach in this partnership was for the three national laboratories to work closely with various SRC-supported universities developing advanced models and simulators and the member companies provide the all-important experimental validation (See Fig. 1.). This three-way collaboration resulted in substantially improved abilities to model semiconductor materials and devices and to simulate certain systems and manufacturing processes that were transferred, via SEMATECH and the SRC, to the participating member companies. The technical success of this partnership resulted in Motorola moving 10 people to Los Alamos, and they are now one of the anchor tenants at the Los Alamos Research Park.

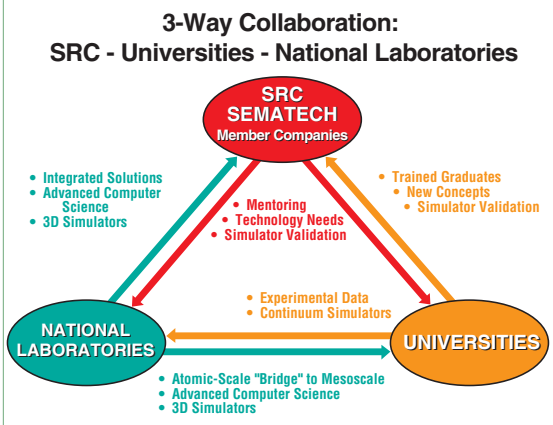


Fig. 1. Identification of the major technical contributions to the other members of the partnership by each member of the three-way collaboration between the SRC, SRC-funded universities, and the DOE national laboratories. This diagram summarizes why this collaboration was successful in that each of the three organizations brought one or more kinds of expertise not available in the other two members but essential to the success of the three-way partnership.

A summary of the major technical accomplishments from this 6-year program, and the individuals at the Laboratory responsible for these accomplishments, are as follows.

3-D Grid and Computer Science

The 1994 NTRS identified 3-D grid generation as the top priority in modeling and simulation. Starting with a raw 3-D unstructured grid generation code developed initially by Harold Trease (Applied Physics Division), Denise George and Andrew Kuprat (Theoretical Division) developed a fully 3-D, adaptive, unstructured grid generation code which was transferred in 1999 to IBM, Bell Labs (Lucent) (See Fig. 2.), Intel, and Motorola for use in their proprietary simulators. This same grid capability, as good or better than anything currently available commercially, has been incorporated into new 3-D simulators for

- material deposition by Robert Walker (T-Division);
- grain growth and mobility by Galen Straub, Denise George, and Andrew Kuprat (T-Division); and
- a general 3-D Poisson solver by David Cartwright, Denise George, Andrew Kuprat, George Csanak, and Robert Walker (all in T-Division).

Topography

Material deposition has been the dominant process in the fabrication of semiconductor devices, but the industry did not have reliable 3-D simulators for predicting the shapes of the deposited materials.

Working with Sandia, Robert Walker developed a 3-D topography simulator which, when combined with molecular dynamics capabilities from David Hansen, Joel Kress, and Art Voter (all T-Division), made predictions of material deposition subsequently confirmed by Motorola.

Ion Implantation and Bulk Processes

Ion implantation, the process of converting an insulator into a semiconductor, is a key process in the semiconductor industry. Charles Snell (Applied Physics Division) spent 3 years working with the University of Texas at Austin developing an advanced simulator for predicting the density profiles of ion implants. This simulator is presently the "standard" used by most of the U.S. semiconductor companies as a design tool for fabricating semiconductors.

In an extension of this simulator to lower ion energies, Keith Beardmore and Neils Grönbech-Jensen (T-Division) developed a molecular-dynamics simulator for predicting 3-D ion implantation at the very low incident energies to be used by the industry for small devices. This simulator won a 1999 R&D 100 award.

Working with Intel, and at the request of Intel, workers at Sandia and Shao Ping Chen, and John Wills in T-Division developed improved models for a phenomenon called "transient enhanced diffusion," which is a process by which the implanted ions move in the solid when it is heated.

Materials Reliability

The electrical reliability of semiconductors can be traced to grain mobility within the electrical interconnects. Galen Straub, Denise George, and Andrew Kuprat worked with MIT and Dartmouth for 4 years to extend the 2-D models developed there to fully 3-D models. This effort has resulted in an extended research program on material properties with professors and graduate students from Carnegie-Mellon University.

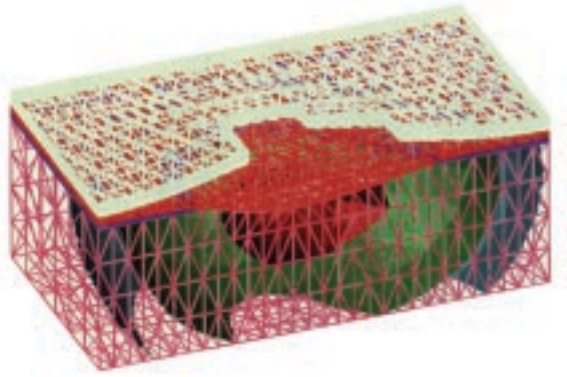


Fig. 2. Illustration of the test problem provided by Bell Labs in 1995 to determine if the grid capability available in the DOE labs was powerful enough to enable the solution of important unsolved 3-D problems at Bell Labs associated with semiconductor device manufacture. The capability desired by Bell Labs was the ability to provide a quantitative, 3-D, time-dependent solution of the ion transport and growth of the SiO_2 layer during the thermal annealing process (following ion implantation). Isosurfaces of the ion density are shown by the dark and green regions near the bottom of the figure, the SiO_2 layer is shown in red, and the light top layer is a mask through which the ions are implanted, all for a particular time during the annealing process. The difficult part of solving this problem was having a grid toolbox that could accurately follow the material interfaces as they moved during the thermal annealing process. The LANL grid code (LaGriT) was the only code in the three participating DOE labs that could provide that capability (See Fig. 3.). LaGriT was transferred to Bell labs and they are now using this grid toolbox with their proprietary semiconductor device design codes to address 3-D device design issues.

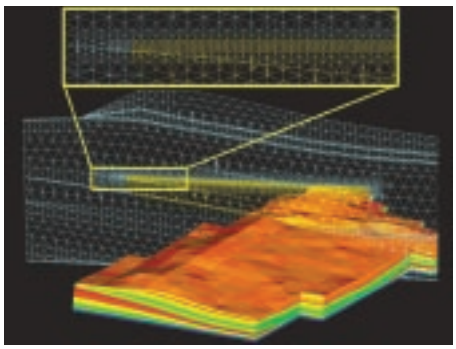


Fig. 3. Illustration of the use of the LANL grid toolbox, LaGriT, to simulate the flow of fluids through multi-layer, high-aspect ratio, geological strata. The insert at the top of this figure is an enlargement of the mesh used in one of the strata. This example came from the work done by Earth and Environmental Sciences Division in the mid-1990s in the study of the practicality of using Yucca Mountain as a storage area for radioactive waste material.

An Overview of Significant T-Division Collaborations

Harlow
N. Johnson
L. Wood

by Linda K. Wood, Norman L. Johnson, and Francis H. Harlow

Los Alamos Theoretical Division researchers have made and continue to make enormous impacts on many aspects of human life through collaborations with industries and government. T-Division expertise has contributed to solving problems on local, national, and international levels. We discuss several of these problem-solving technologies in other chapters of this book. Space and time do not permit a detailed explanation of all T-Division collaborations. Here, we list and briefly describe some examples.

Industrial Collaborations

Aerojet. Aerojet is one of several U.S. industries and government agencies that have requested T-Division VNAP codes—computer programs for calculating turbulent, steady, and unsteady flow. Some of these collaborations have been ongoing. Most aerospace companies used the codes to analyze standard aircraft components, such as propulsion nozzles, engine inlets, and airfoils. In all collaborations, T-Division personnel supplied the codes to others, helped the recipients use the codes, and worked with user problems and comments to improve the codes.

Allied Signal, Inc. See "Aerojet."

American Oil Company (Amoco). Amoco entered into a cooperative research and development agreement (CRADA) with T-Division (funded by the U.S. Department of Energy Office of Defense Programs [DOE/DP]). The object of this research was to better understand and enhance processes used in oil refining. This collaboration resulted in major expansion of the Los Alamos code library CFDLib.

Ashland Specialty Chemical Company.

Atlantic Research Corporation. See "Aerojet."

Bell Laboratories, Inc. (Lucent Technologies). Researchers in T-Division developed a three-dimensional (3-D), adaptive, unstructured grid-generation code that was transferred in 1999 to Bell Laboratories, Inc. (Lucent Technologies) and other industries for use in their proprietary simulators.

The Boeing Company. See "Aerojet."

Caterpillar, Inc. T-Division researchers extended the KIVA code (See also "Cray Research, Inc.") to model the details of an internal combustion engine for studying such issues as efficiency and nitrogen oxide (NO_X) emissions.

ChevronTexaco Corporation.

Cray Research, Inc. (CRI). Another T-Division CRADA, funded by DOE/DP and CRI, created an unstructured-grid version of CFDLib. Also, because KIVA and KIVA-II software were originally written to run on a Cray computer, they contained coding that enhanced their performance on Cray platforms. CRI personnel realized the marketing potential for both the software package and Cray computers. In 1992, Los Alamos granted a license to CRI for KIVA-II that was limited to Cray platforms. CRI workers improved the user interface and graphics, giving the program a more user-oriented focus for design engineers, and marketed their version as CRI/TurboKIVA.

Cummins Engine Company. This collaboration involved searching for a more efficient fuel-consumption engine and reduced air pollution.

DaimlerChrysler.

Detroit Diesel Corporation. T-Division partnering on the KIVA computer codes that aided fuel efficiency also benefited Detroit Diesel. See "Caterpillar, Inc." and "Cray Research, Inc."

The Dow Chemical Company. T-Division scientists derived the constitutive properties of a polymeric foam for application to seat cushions and other configurations.

Eastman Kodak Company. T-Division research with Kodak involved applying a VNAP code to gas flowing through a supersonic nozzle and into a large vacuum tank. See "Aerojet" for additional information.

E. I. DuPont de Nemours and Company. See "Multiphase Fluid Dynamics Research Consortium."

Electric Power Research Institute. T-Division investigators performed studies relevant to nuclear reactors for energy production.

ExxonMobil Corporation. T-Division researchers teamed with personnel at ExxonMobil in yet another CRADA to study the injection zone of a chemical reactor used for cracking crude oil in the refining process for making gasoline feedstock. Also, see "Multiphase Fluid Dynamics Research Consortium."

FluiDyne Engineering Corporation. See "Aerojet."

Ford Motor Company. T-Division researchers collaborated with Ford personnel on the KIVA project to achieve more efficient fuel consumption and cleaner air.

Garrett Air Research. See "Aerojet."

General Atomics. T-Division investigators partnered with General Atomics personnel to provide computational support in search of improved energy and power production.

General Dynamics Corporation. See "Aerojet."

General Electric Company. T-Division scientists studied the behavior of composite materials for application to strength and potential damage when strong impacts occur to machine components.

General Motors Corporation (GMC). T-Division researchers collaborated with GMC personnel on the KIVA computer code to produce a more fuel-efficient engine that would lead to reduced fuel consumption and decreased air pollution.

Grumman Aircraft Engineering Corporation. See "Aerojet."

Hercules Aerospace. See "Aerojet." A T-Division scientist provided Hercules personnel with copies of the VNAP series of codes and helped them use the codes to analyze propulsion nozzles on their rocket engines.

Intel Corporation. See "Bell Laboratories, Inc."

International Business Machines Corporation (IBM). See "Bell Laboratories, Inc."

International SEMATECH. See "Semiconductor Research Corporation."

McDonnell/Douglas Corporation. See "Aerojet."

Millenium Chemical Corporation. See "Multiphase Fluid Dynamics Research Consortium."

Motorola, Inc. See "Bell Laboratories, Inc." and "Semiconductor Research Corporation" in this "Industrial Collaborations" listing and the article, "Modeling and Simulation with the U.S. Semiconductor Industry," in this "Industrial and Governmental Collaborations" chapter.

Multiphase Fluid Dynamics Research Consortium (MFDRC). This 6-yr project has been funded by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy, Office of Industrial Technologies (DOE/EERE/OIT). Purpose of the project is to improve U.S. industrial energy efficiency by developing and validating an engineering-level model for multiphase turbulence. Applications span a wide range, representing the separate interests of the participating organizations. Generally, oil refining and chemical manufacturing industries have gained the most significant benefits.

Physical Sciences Inc. See "Aerojet."

Physics International Company. See "Aerojet." Physics International used the VNAP codes to calculate the effects of nuclear weapon blast waves.

Pratt & Whitney. See "Aerojet."

Primex Technologies, Inc. (formerly Rocket Research Company and Olin Corporation, Ordnance Division). A funds-in agreement between Primex and T-Division staffs resulted in a spin-off code that analyzes arc-jet thrusters. These thrusters are small jets used for satellite control in space. The code was a derivative work based on the code library CFDLib.

Procter & Gamble. A T-Division partnership with Procter & Gamble and other Los Alamos National Laboratory divisions yielded an improved process for producing a better-quality baby diaper. Procter & Gamble personnel wanted a multiphase modeling capability for their diaper manufacturing process. Specifically, they wanted to reduce costs by making quicker fault analyses of existing processes and by eliminating expensive hardware design steps. (See the article, "An Example of a Successful Industrial and Governmental Collaboration" in this "Industrial and Govermental Collaborations" chapter.)

Scientific Applications International Corporation (SAIC). In a 20-yr ongoing project, T-Division researchers have provided 3-D nonlinear magnetohydrodynamic simulations for the SAIC. These simulations aid in energy research.

Semiconductor Industry Association. In early 1995, a 5-yr program was initiated to develop advanced modeling and simulation tools for the U.S. semiconductor industry. This T-Division collaboration resulted in substantially improved abilities to model semiconductor materials and devices and to simulate certain systems and manufacturing processes that International SEMATECH and the Semiconductor Research Corporation transferred to participating member companies.

Semiconductor Research Corporation (SRC). See "Semiconductor Industry Association."

Silicon Graphics, Inc. A CRADA with Silicon Graphics produced the basic groundwork for a new version of the code library CFDLib. Because the same physics and the same numerical method are used in both versions, this expansion furnishes the ability to test efficiencies of processing on new computer architectures. Depending on the particular computer design, the difference in data structure can substantially affect computational speed.

3M. Drivers can more easily read street and highway signs because of a T-Division/3M CRADA called "Properties of Polymers and Organic Dye Molecules." The T-Division Theoretical Chemistry and Molecular Physics Group developed models that led to more durable dyes on traffic signs, better performance of sensors, and more effective photographic sensitizing dyes.

United States Steel Corporation. See "Multiphase Fluid Dynamics Research Consortium."

Government Collaborations

Argonne National Laboratory (ANL). The sea ice model (CICE) is software developed by T-Division for use in climate models. This unique software is available free on the World Wide Web and can be adapted for local or global use. Each user alters the model for his/her organization's specific needs. In some cases, users compare CICE against their own models. In other cases, they take only the part of the code that they want to use, such as thermodynamics, dynamics, or advection, and adapt CICE for their specific user needs. ANL derived an adjoint model from CICE to test the sensitivity of some critical CICE parameters that are not well known from physical measurements.

Atomic Energy Commission (AEC). The AEC was the principal source of nuclear weapons funding and oversight for T-Division and many other groups of the Los Alamos Laboratory. The AEC later became the Energy Research and Development Administration.

Center for Disease Control and Prevention (CDC). One T-Division collaboration with the CDC has resulted in collected and organized epidemiological data about deaths from influenza-like illness in every county in the U.S. Los Alamos is using this data to develop new epidemiological models that track the spread of influenza in the U.S.
A second T-Division collaboration with the CDC has produced an influenza sequence database that provides genetic information about influenza evolution and the variety of influenza strains circulating throughout the world. Researchers worldwide rely on this database.

Defense Advanced Research Projects Agency (DARPA). T-Division research for DARPA focused on questions of shock and interactions and damage to engineering equipment.

Defense Nuclear Agency. This agency funded some T-Division research on fluid-flow phenomena.

Energy Research and Development Administration (ERDA). As a predecessor of the Department of Energy, ERDA oversaw nuclear weapons research in T-Division and in many other groups of the Los Alamos National Laboratory.

Environmental Protection Agency (EPA). The EPA funded some T-Division research on the dispersal of pollutants in the atmosphere.

Federal Energy Technology Center. See "Multiphase Fluid Dynamics Research Consortium" under "Industrial Collaborations."

Geophysical Fluid Dynamics Laboratory. See "Argonne National Laboratory."

Lawrence Livermore National Laboratory. See "Semiconductor Research Corporation" under "Industrial Collaborations" for an example of numerous collaborations with Los Alamos's "sister" laboratory. Many such collaborations are reported throughout this book.

National Aeronautics and Space Administration (NASA). See "Aerojet."

NASA's Jet Propulsion Laboratory. See "Argonne National Laboratory."

NASA's John C. Stennis Space Center. See "Argonne National Laboratory." Researchers at Stennis are configuring and testing the T-Division CICE model for use in the U.S. Navy's new model for predicting the ocean state.

National Center for Atmospheric Research. Researchers at the National Center for Atmospheric Research in Boulder, Colorado, use the T-Division CICE model to look at global climate trends. The control climate run has now reached 1000 years, an unprecedented length of time for such a complex model on a grid that resolves ocean properties reasonably well.

National Center for Biotechnology Information (NCBI). In the early 1980s, T-Division researchers began work on the Genbank to produce a technology that is now handled by the NCBI.

National Center for Genomic Research. Several T-Division personnel who worked in genomic research have formed the National Center for Genomic Research and continue studies that have applications to both agriculture and humanity.

National Institute of Allergy and Infectious Diseases (NIAID). In the interest of conquering infectious diseases, T-Division staff have partnered with NIAID personnel to research, create, and maintain accurate databases on influenza, infectious diseases, and human immunodeficiency virus (HIV). The NIAID funds this ongoing research.

National Institutes of Health. NIH furnished major funding for T-Division's work on GenBank and HIV databases, including immunology and drug-resistent mutations. In addition, T-Division worked with NIH on genetic networks, modeling of immune systems, and other major areas of medicine and biology.

National Science Foundation (NSF). With an NSF grant, T-Division investigators performed a detailed analysis of the Lubbock, Texas, tornado in the early 1970s.

Nuclear Regulatory Commission (NRC). With many years of funding from the NRC, T-Division personnel performed major calculations to study the safety of a pressurized-water reactor that was subjected to various hypothetical accidents.

Oak Ridge National Laboratory. See "Multiphase Fluid Dynamics Research Consortium" under "Industrial Collaborations."

Office of Naval Research. The Office of Naval Research funded some T-Division investigation of the interaction of naval structures with sea-surface and undersea disturbances.

Pacific Northwest National Laboratory. See "Multiphase Fluid Dynamics Research Consortium" under "Industrial Collaborations."

Sandia National Laboratories (SNL). See "Multiphase Fluid Dynamics Research Consortium" and "Semiconductor Research Corporation" under "Industrial Collaborations" for examples of the many collaborations that Los Alamos has conducted with SNL.

United States Air Force. Personnel at the Air Force Office of Scientific Research collaborated with T-Division researchers to produce computer models for better understanding the operation of chemical lasers.

United States Army. In the 1940s, T-Division, along with the rest of the Laboratory, was part of the Manhattan District of the U.S. Army Corps of Engineers; and researchers performed atomic weapons research. The Army's Cold Regions Research and Engineering Laboratory scientists have collaborated with T-Division researchers on the CICE model. See also "Argonne National Laboratory" in "Government Collaborations" and "Aerojet" in "Industrial Collaborations."

United States Department of Transportation (DOT). The DOT funded some T-Division research on the drag of air on a moving vehicle.

United States Department of Commerce, National Oceanic and Atmospheric Administration. See the article, "Climate, Ocean, and Sea Ice Modeling," in the "Environmental Studies" chapter of this book.

United States Energy Research and Development Agency. Under the auspices of this agency, T-Division researchers studied direct fuel injection, stratified charge gasoline (DISC) and diesel engines, and more conventional lean-burn homogenous charge gasoline engines. The goal of this study was cleaner-burning, more efficient engines.

United States Forest Service. This ongoing T-Division collaboration focuses on the behavior of wildfires in rugged, forested terrain.

United States Navy. See "Aerojet" under "Industrial Collaborations" and "Office of Naval Research" in this "Government Collaborations" section.

International Collaborations

All-Russian Research Institute of Physical-Technical and Radio-Technical Measurement (VNIIFTRI) Multicharged Ion Spectral Data Center (MCISDC). T-Division scientists performed spectral modeling calculations that provide density and temperature diagnostics for MCISDC high-resolution plasma spectroscopy experiments. The work has applications in pulse power, biological imaging, and nuclear fusion.

Russian Federal Nuclear Center (VNIIEF). This T-Division collaboration produced a major report on the theoretical and experimental processes in the turbulent mixing of materials.

Asia Pacific Center for Theoretical Physics, South Korea.

Atomic Weapons Establishment, United Kingdom (UK). T-Division researchers partnered with their UK counterparts on major technical interactions concerning problems of scientific and engineering interest to both laboratories.

Bedford Institute of Oceanography, Canada. See "Argonne National Laboratory" under "Government Collaborations." Bedford is using the T-Division CICE model to study the ice and ocean conditions in the nearby region, particularly the Labrador Sea and the eastern part of the North Atlantic Ocean.

British Petroleum (BP), United Kingdom.

Centre National de la Recherche Scientifique, France. Researchers in T-Division are working to understand the properties and behavior of complex fluids, such as foams. This work is being done in partnership with the Centre National de la Recherche Scientifique (the French equivalent of the U.S. National Science Foundation) and the Chinese Academy of Sciences.

Chinese Academy of Sciences. See "Centre National de la Recherche Scientifique."

Commissariat a L'Energie Atomique (CEA). This extensive T-Division collaboration resulted in the development of the basic theories of turbulent fluid flow.

Dutch National Influenza Center, The Netherlands. On this project, the Dutch provided data on which to test algorithms developed in T-Division and collaborated on the development of algorithms. The Dutch National Influenza Center participates in the World Health Organization influenza surveillance program.

European Centre for Theoretical Studies in Nuclear Physics and Related Areas (ETC*), Italy.

GlaxoSmithKline, United Kingdom.

Hadley Centre for Climate Prediction and Research, United Kingdom. See "Argonne National Laboratory" under "Government Collaborations." Hadley predicts climate, using the dynamics and thermodynamics portions of the T-Division CICE model.

Institut fuer Kernphysik, Germany.

International Atomic Energy Agency (IAEA). T-Division nuclear physicists have led and participated in many IAEA collaborative projects. One such effort is an international database for describing photonuclear cross sections that are important for medical applications in external-beam cancer therapy. The same T-Division/IAEA database is also important in hindering the spread of nuclear weapons technology. Working with the IAEA, T-Division personnel have recommended a set of nuclear input parameters for nuclear reaction model calculations.

Japan Atomic Energy Research Institute (JAERI)-High Energy Accelerator Research Organization (KEK), Japan.

Nansen Environmental and Remote Sensing Center, Norway. See "Argonne National Laboratory" in "Government Collaborations."

Nuclear Energy Agency. The Nuclear Energy Agency (NEA) in Paris has adopted NJOY, the T-Division applied nuclear physics computational tool for nuclear data processing, as the NEA's standard resource in reactor, accelerator, and waste transmutation technologies.

Public Health Laboratory Service, United Kingdom.

Swedish Meteorological and Hydrological Institute. Swedish Meteorological and Hydrological Institute researchers use part of the T-Division CICE model to do localized climate studies. See "Argonne National Laboratory" in "Government Collaborations."

World Health Organization.

KIVA: Working for Clean Air and Fuel Efficiency

by Anthony A. Amsden, Peter J. O'Rourke, and T. Daniel Butler

Amsden
Butler
Cline
Cloutman
Dukowicz
Farmer
Harlow
Hubbard
Nakayama
O'Rourke
Ramshaw
Rivard

In the mid-1970s, Theoretical Division's Fluid Dynamics Group began developing a series of computer simulation programs to aid in producing cleaner-burning, more fuel-efficient automotive engines. These programs evolved into the KIVA family of three-dimensional modeling tools. Today, KIVA calculates reactive flows in complex geometries that may include moving pistons, multiple moving valves, cylinder wall ports, and entire intake and exhaust manifolds.

KIVA is in everyday use in the design of internal combustion engines, gas turbines, industrial furnaces, heaters, catalytic converters, evaporative cooling towers, smokestacks, waste incinerators, and electric-power generating systems. From the beginning, Los Alamos and industry had a close working relationship that continues to the present.

KIVA saves designers expense and time in product development by suggesting optimum configurations, reducing expensive experimentation, and providing insights into the understanding of complex flows. Being in the public domain gives KIVA a broad user community. Major automotive manufacturers, universities, and laboratories worldwide are using KIVA programs.

Introduction

Combustion is a major process affecting our lives because it provides over 90% of the energy we use at home or when traveling from place to place. Unfortunately, the combustion of fossil fuel is also the main source of environmental pollution. It is estimated that in the United States, automobiles and trucks account for 50% of the petroleum consumption, 20% of the total energy consumption, and 50% of the pollutants entering the atmosphere. Improving combustion processes is, therefore, of paramount importance for reducing both fuel usage and its emissions. When we consider the tens of millions of vehicles in continuous use worldwide, even relatively minor improvements that result in cleaner, more-fuel-efficient engines yield significant benefits.

The physics and chemistry involved in combustion systems are extraordinarily complex and the controlling parameters numerous. Modern power-generating systems, such as internal combustion engines, are indeed sophisticated, with the simultaneous interplay of numerous physical processes. These may include the turbulent flow of incoming fresh air and recirculation of exhaust gases through a tortuous geometry and their interaction with the rapid motion of valves

and pistons. Accompanying this activity is the spray of a carefully formulated liquid fuel, precisely metered, focused, and injected at considerable velocity, with subsequent vaporization and mixing of the fuel and air. Ignition follows, initiating combustion and the turbulent propagation of an expanding high-temperature flame front, generation of useful energy, transfer of heat to the surrounding metal walls, flame extinction, and finally, exhaust of the hopefully small amount of pollutants that result from the process.

Designers of combustion systems have come to use all means at their disposal to optimize these systems to meet both efficiency and pollution standards at low production costs. Design guidance and understanding, which previously was achieved almost exclusively using experimental techniques, are now routinely supplemented with sophisticated numerical tools that model the detailed fluid and combustion dynamics. In this latter area, T-Division has pioneered developments that have led to the general acceptance of this additional approach.

Over time, increased computing power, coupled with improved numerical algorithms, has brought about this acceptance. These factors and experimental diagnostic techniques have made it possible to simulate complex combustion processes and to verify the quality of the simulation experimentally in circumstances where data are available. This paper details the background and developments that led to the numerical approach and its worldwide acceptance by industry and academia.

Over many years, the U.S. Department of Energy has funded a comprehensive combustion research program. One aspect of this program has been the development of KIVA, a computer program that simulates the dynamics in practical combustion devices. Since the mid-1970s, Theoretical Division has developed a sequence of first two-dimensional and then three-dimensional simulation codes. These codes draw on the computational fluid dynamics expertise at Los Alamos developed over the years for modeling fluid flow regimes ranging from incompressible flows to compressible low- and high-Mach-number flows. The challenges for simulating such devices include the ability to numerically solve the time-dependent, three-dimensional equations of motion for turbulent, low-Mach-number, multicomponent two-phase flows, mixing, embedded-density gradients, chemical reactions, and heat transfer.

KIVA calculates reactive flow in complex geometries that may optionally include moving pistons, multiple moving valves, ports in the cylinder wall, and entire intake and exhaust manifolds. Processes modeled may include fuel-spray dynamics, formation of liquid wall films, evaporation, turbulent mixing of fuel and air, and combustion

with resultant heat release, wall heat transfer, and exhaust-product formation. Because of its broad range of features, KIVA has been applied to many devices besides internal combustion engines, including gas turbines, industrial furnaces, heaters, catalytic converters, evaporative cooling towers, smokestacks, and waste incinerators.

In this article, we explore how T-Division became involved with the automotive industry, describe the origins and continuing evolution of KIVA, and discuss the successful process that has occurred in transferring KIVA technology to a worldwide user community.

Origins of KIVA

The origins of KIVA are in computational methods still in common use for nuclear weapons design. In the early 1970s, a team (Dan Butler, Otis Farmer, Peter O'Rourke, John Ramshaw, and William Rivard) in the T-Division Fluid Dynamics Group developed a reactive fluid dynamics program to study hydrogen-fluorine (HF) continuous-wave chemical laser systems. They did this work under contract to the U.S. Air Force.

Several years later, the U.S. found itself in its first energy crisis. In 1976 the National Science Foundation sponsored a meeting in which the participants were asked to propose ways to make more fuel-efficient and cleaner-burning automotive engines. The team realized that the program for modeling chemical lasers could be adapted to simulate reactive flows in an internal combustion engine. They prepared a motion picture of the modified HF program and showed it at the meeting. Participants thought the program had much promise. Thus began the affiliation between Los Alamos National Laboratory and the combustion research community.

Under the auspices of the U.S. Energy Research and Development Agency and its successor, the DOE, representatives from industry, universities, and several DOE laboratories formed collaborative working groups for advanced engine concepts. Engine concepts to be studied included direct fuel injection, stratified charge gasoline (DISC) and diesel engines, and more conventional lean-burn homogenous charge gasoline engines.

The groups met quarterly at first and then semiannually. The working group format and the diversity of its members provided a good venue for cultural sharing, which enabled the universities and national laboratories to learn the needs of industry, and industry to gain an appreciation of numerical modeling as an adjunct to experimentation.

The role of the Los Alamos Fluid Dynamics Group was to develop a major combustion simulation program, initially to be used exclusively by the working group participants. At the time, industry had neither the computational fluid dynamics expertise nor the computing power necessary for undertaking such a development, but Los Alamos could meet both of these requirements. With input from the automotive industry, the research program evolved in several well-defined increments.

The approach that T-Division took for the modeling effort was to solve the governing partial differential equations that represent the complex processes described previously by employing multidimensional finite-difference techniques. These techniques provide detailed descriptions in space and time of these processes. The spatial region of interest is subdivided into computational cells, and the computed flow variables in these cells (e.g., velocities and pressures) are advanced in tiny time increments. Thus, the solution of the equations evolves much like the frames of a motion picture, in which each frame differs only slightly from its predecessor.

Over time, we developed increasingly complex computer programs for the combustion project, incorporating the piston as a moving boundary. The new approach allowed the description of curved and/or moving boundary surfaces, an important feature for internal combustion engines. The basic method was extended to include the transport of chemical species, their mixing, and chemical reactions.

The dynamics of a liquid fuel are crucial to the operation of fuel-injected gasoline and diesel engines. John Dukowicz devised and Peter O'Rourke further refined, a statistical fuel-spray model to represent the interplay of air with an evaporating liquid moving through it. [1,2] This model represents the spray as a finite number of computational particles that exchange mass, momentum, and energy with the gas. This approach has the advantage of allowing the convenient representation of a spectrum of particle sizes, and it eliminates fictitious numerical diffusion in the calculation of liquid transport. The model was introduced in a two-dimensional computer program written by Anthony Amsden, who joined the team along with Lawrence Cloutman, who also made contributions to KIVA.

The next stage of the evolution added a full three-dimensional capability; this program became the original KIVA. [3,4]

KIVA did not spring into existence full-grown. It evolved from a series of pilot programs that were each considered innovative for their time. However, the initial version of KIVA, written in 1981–82, was too slow even on a Cray-1 computer—the supercomputer of its day—to be of practical use for complex applications. Accordingly, we revised the numerical solution algorithm and replaced the implicit solution technique in use at the time with a computationally faster explicit subcycling method.

At the urging of the industrial participants in the collaborative groups, we expended considerable effort to tailor KIVA to work in a more optimal fashion on the Cray computer by taking advantage of the vector capabilities of the machine. This vectorized version of KIVA ran nearly five times faster than before, a speed much more acceptable to potential users.

We released KIVA for collaborator-testing to General Motors Research Laboratories in 1983. Shortly thereafter, we released the program to Cummins Engine Company, Princeton University, Purdue University, and Sandia National Laboratories/California. We needed feedback from this group of "friendly users" in order to improve the program to a level where it could be considered for public release. The challenge facing the small KIVA team was twofold: to demonstrate the usefulness of three-dimensional combustion modeling on a computer to a skeptical audience and to continue improving the program.

We made the first public release of KIVA in 1985 through the National Energy Software Center at Argonne National Laboratory, which served at the time as the official distribution center for DOE-sponsored software.

KIVA-II: Moving into a Wider World

In 1987, the T-Division team presented a paper at the Society of Automotive Engineers (SAE) International Congress. The paper presented a KIVA calculation of a DISC engine with a complex three-dimensional geometry. The calculation modeled the compression of air after intake valve closure, the fuel injection process, the spark ignition, and the burning of the air-fuel mixture. Calculations were made under three different engine load conditions; the reported results included comparisons with experimental data of cylinder pressure histories and analysis of exhaust products. Some of these compared well with experiments, others not so well.

More important, perhaps, was that KIVA revealed flow details that were inaccessible to the experimentalists. Of primary importance were graphics that illustrated the position of the burning fuel cloud as a function of time. These graphics provided a possible explanation of why the engine, although performing quite well, had a higher level of emissions than was acceptable to meet the ever-tightening federal emissions standards. This was one of the first times that such a detailed study was reported. The paper received a 1988 SAE Arch T. Colwell Merit Award for making an outstanding contribution to the automotive literature.

One study alone hardly constitutes comprehensive benchmark testing. The Los Alamos team and other users worldwide soon began testing KIVA in a broad variety of applications. Over time, users presented a significant number of papers, each focusing on some aspect of the

model and often offering extensions and improvements. The model itself gradually became more efficient and realistic, resulting in the public release in 1989 of an improved version called KIVA-II. [5,6]

In contrast to the original KIVA, the temporal difference scheme in KIVA-II is largely implicit. The scheme employs coupled, implicit differencing of the diffusion terms and terms associated with pressure-wave propagation. What enabled the use of implicit methods, which had been abandoned in the original version of KIVA, was the development of a new class of implicit equation solvers called "conjugate gradient methods" and a new method for solving systems of coupled implicit equations. Because of these developments, we calculated the time steps used in KIVA-II based on accuracy, not stability conditions. The combined use of both implicit and explicit difference approximations in KIVA-II resulted in considerable savings of computational time in most combustion problems.

Another major difference between KIVA-II and KIVA is the inclusion of a two-equation K-ε turbulence model. T-Division's Frank Harlow and Paul Nakayama first developed the K-ε model; and researchers at Imperial College of Science, Technology, and Medicine in London further developed the K-ε model for numerical predictions of turbulent flows. [5,6] Because flows in practical combustors are turbulent, accounting for turbulence transport and production in combustion simulations is necessary. Use of the K-ε model allowed the inclusion of these effects in a computationally economical and robust fashion.

Use and acceptance of the program grew rapidly after the introduction of KIVA-II; and it came into use by major U.S. auto and diesel manufacturers, many federal laboratories, and mechanical engineering departments at numerous universities. In 1990, a patent was issued to General Motors for a high-turbulence piston design that specifically identified three-dimensional computer simulation for making the invention possible. KIVA-II played a major role in this development.

Another area in which KIVA-II saw heavy use was in modeling gas turbine combustors. Under NASA sponsorship, Michael Cline of T-Division and university researchers conducted a combined computational-experimental program to study a variety of combustor designs. This effort contributed to improved combustors with reduced oxides of nitrogen (NO_x) production for the next generation of civilian jet aircraft engines. Figure 1 depicts a KIVA calculation of the flow field near the nozzle in the combustor section of a gas turbine engine. In addition to predicting and suggesting ways to reduce NO_x formation in gas turbine engines, KIVA calculations were used to suggest ways to reduce thermal loading on combustor walls.

Because KIVA and KIVA-II were originally written to run on a Cray computer, they contained coding that enhanced their performance on Cray platforms. Cray Research Inc. (CRI) realized the marketing potential for both the software package and Cray computers. In 1992,

Los Alamos granted a license for KIVA-II to CRI that was limited to Cray platforms. CRI improved the user interface and graphics, giving the program much more of a user-oriented focus for design engineers, and marketed their version as CRI/TurboKIVA.

KIVA-3: The Next Stage of Development

The first two versions of KIVA lend themselves well to confined in-cylinder flows and to a variety of open combustion systems, but they can become quite inefficient when applied to complex geometries that contain such features as intake and exhaust ports. KIVA and KIVA-II employ a single, logically rectangular block of cells that must encompass the entire region of interest. Communication among cells is through a set of three indexes, one for each logical direction; and each index has a range that remains unchanged throughout the calculation. Thus, if adjoining computational regions have differing sizes, such as ports whose cross-sectional dimension is smaller than that of the cylinder to which they attach, a significant number of cells are not part of the flow region. This condition causes inefficient use of computer memory and time. The situation is analogous to packing an odd-shaped object in a cardboard box, where large amounts of filling material must be used to fill in the empty spaces.

KIVA-3, [7] introduced in 1993, circumvented this inefficiency by employing a block-structured mesh composed of an arbitrary number of large, small, long, or short blocks of cells that are patched together to form the entire computational region. In KIVA-3 each cell is identified by a single index, rather than by the previous three; and neighbor identification is maintained in arrays that specify the indexes of the six adjacent cells.

In addition, KIVA-3 allows motion of the piston past any number of intake and exhaust orifices in the cylinder wall. With the piston moving down and up, the port volumes are connected or disconnected from the cylinder volume as the piston moves past them. This condition necessitates changes in grid topology. Considerable effort is expended in KIVA-3 to ensure flexibility in handling such circumstances while maintaining computational efficiency. Also, special care is taken to facilitate the different types of boundary conditions.

The first applications of the block-structured grid capability were to two-stroke engines. KIVA-3 also introduced the capability to model two opposed pistons in the same cylinder, and, under contract to the Advanced Research Projects Agency, was applied to study combustion in an advanced high-power-density two-stroke diesel engine. Figure 2 is from a KIVA-3 calculation of the opposed piston engine.

Central issues for two-stroke engines are how much of the residual combustion gases are replaced by fresh air, where the fresh air and residual gases reside, and to what extent these are mixed together

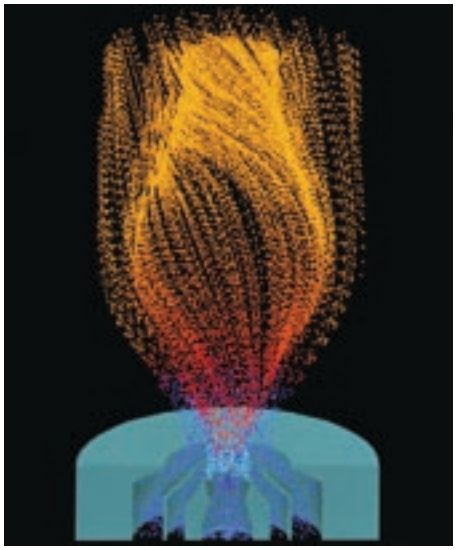


Fig. 1. The swirling air/fuel mixture is produced by a three-passage air-blast fuel nozzle used in the combustor section of a gas turbine engine. The nozzle is at the bottom and the flow is from bottom to top. The small and large particles represent the gas flow and liquid phase, respectively. The color of the gas particles represents temperature (blue, low; red, medium; yellow, high) and droplet size for the liquid phase (blue, small; green, medium; red, large). (Margaret Hubbard of T-Division produced the plot.)

before the next combustion event. KIVA-3V calculations can provide detailed information in response to these and similar questions that cannot be answered by experimental measurements.

By late 1993, the computing power and sophistication of FORTRAN compilers on high-end workstations had advanced to the point where these workstations became a viable alternative to the supercomputers that preceded them. Although many major manufacturers and universities by then had supercomputer power available for computational fluid dynamics modeling, the momentum was gradually shifting toward workstations. Accordingly, T-Division made a transition from Cray-specific coding in KIVA-3 to a "generic" approach. The generic approach gained universal favor in user choice and has been the norm ever since. The increased portability of the generic KIVA-3 attracted an even broader user community.

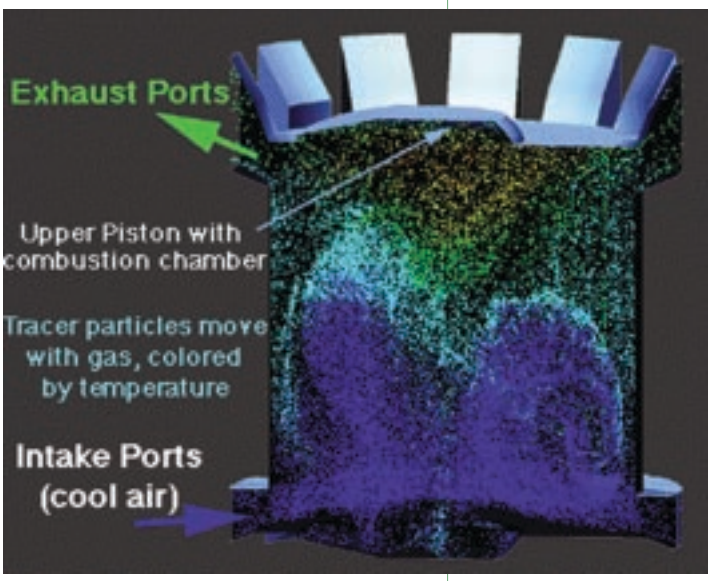


Fig. 2. This plot is from a KIVA-3 simulation of a two-stroke diesel engine with two opposed pistons. The KIVA-3 simulation includes the full physics of the problem: the flow of gases, injection-penetration-evaporation-mixing of the fuel, heat transfer, combustion, and creation of pollutants. The computational region includes the regions where gas exchange and combustion occur: the intake and exhaust passages (the box shapes at the bottom and the top of the figure, respectively) and the cylindrical chamber between the two pistons. In the plot, the flow of the gases is represented by particles that move with the local gas velocity. The color of the particles changes according to the local temperature: from blue at 400 K through cyan, green, and yellow, to red at 1900 K. The colors were chosen so that blue indicates the location of the cool intake air, and red, the product gases of combustion. (Margaret Hubbard of T-Division produced the plot.)

KIVA-3V: The Addition of a Valve Model

With the introduction of KIVA-3V in 1997, an effective approach to model valve movement became available. KIVA-3V is an extension of a model originally introduced by the University of Wisconsin-Madison (see references in [8]), in which valves are treated as solid objects that move through the mesh. Any number of vertical or angled valves in the cylinder head of an internal combustion engine can be represented. Because the valve motion is modeled exactly and the valve shapes are as exact as the grid resolution will allow, the accuracy of the valve model is commensurate with that of the rest of the program. With the addition of intake and exhaust valves, KIVA-3V is now routinely applied to full- and multiple-engine cycle studies with fuel injection and combustion in modern gasoline and diesel engine designs.

Figure 3 depicts events in a KIVA-3V calculation of a modern high-performance four-valve gasoline engine with a pent-roof combustion chamber. We used KIVA-3V calculations to help explain the changes in pollutant emissions that occurred when the timing of the spray injection was varied during engine cold-start events.

Another feature introduced in KIVA-3V is an effective particle-based liquid wall film model. Liquid films are an important consideration in the design and operation of both gasoline and diesel fuel-injected engines. Fuel film deposition and vaporization play a strong role in engine emissions, particularly under cold-start conditions. This is as

true in the port-injected gasoline engines in common use today as it is in the direct-injection (DI) or gasoline-direct-injection (GDI) designs that are coming into vogue. The problem is no less severe in diesel engines, and it is aggravated in small-bore designs in which severe wall-wetting is encountered.

The T-Division team was able to take advantage of KIVA-3V's boundary-condition generality in successfully modeling an engine that had both an inlet port in the cylinder wall and an exhaust valve in the head. This engine was for use in a proposed electrical power generation system. This represents a combination of geometrical capabilities that are likely unattainable by any other such comprehensive computer program in the world.

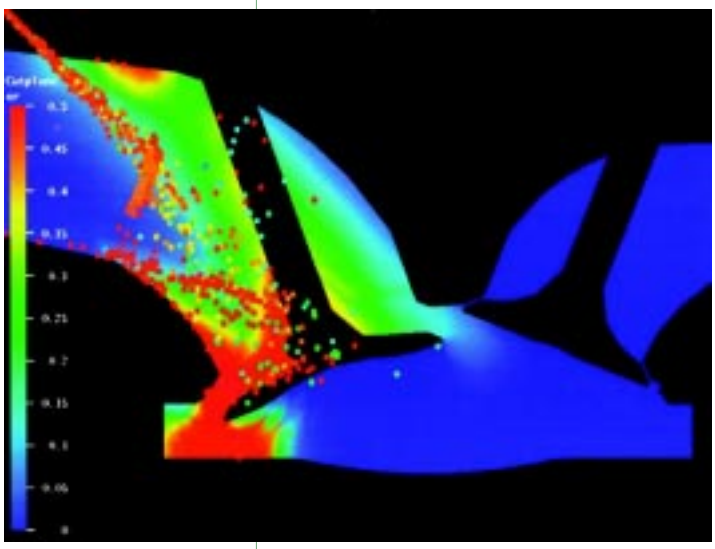


Fig. 3. This plot is from a KIVA-3V calculation of fuel injection and charge intake in a modern four-valve, port-fuel-injected gasoline engine. The cross-sectional view is through one intake valve (left side of picture) and one exhaust valve (right side of picture). The plot is during the intake stroke, at which time spray injection is occurring, the intake valve is opening, and the piston (bottom boundary) is moving downward and drawing the vaporized fuel (red region) into the engine cylinder. The colored particles represent the spray and the wall films that are formed when the spray impinges on the back of the intake valve and on the intake port surfaces. The particles are color coded according to their temperature. The spray droplets are initially at 300 K (red particles) and cool as they vaporize (yellow, green, cyan, and blue particles). The background is colored according to the local equivalence ratio, which is the fuel-to-air concentration ratio divided by the stoichiometric value of fuel-to-air concentration. Blue regions are pure air. At the time of this plot, the fuel-rich red regions are being drawn into the cylinder by the downward motion of the piston (bottom boundary of picture).

The Technology Transfer Process and the KIVA User Community

The close working relationship between Los Alamos and industry and with other collaborative users of KIVA was established at the outset and continues to the present. The importance of personal interaction is indispensable, because the true technology transfer of KIVA has taken place at the grass-roots level. Members of working groups communicate regularly with the KIVA team on an informal basis, and the interactions often lead to formal published papers and presentations.

It has become increasingly common for KIVA users to come to Los Alamos for days, weeks, or even months to work directly with the KIVA team. Cummins Engine Company assigned a mechanical engineer to work at Los Alamos for a year. Other industrial collaborators have come from General Motors, Ford Motor Company, Caterpillar, and Detroit Diesel Corporation.

KIVA does not provide precise answers, because inaccuracies occur in the numerical approximations and physical models. The program is also limited in accuracy by the resolution of the domain of interest, this limitation being imposed by the computing resources and time available. However, KIVA has demonstrated that it can save designers enormous expense and time in developing a product because it can suggest optimum configurations, eliminate a significant amount of expensive experimentation, and can give designers better insights as to the controlling processes that cannot be gained by experimentation alone.

Because it is in the public domain, the availability of the source code has created a broad KIVA user community, particularly in the university sector. In fact, its use in academia is arguably one of its greatest contributions to the engine modeling community. With Masters and

Ph.D. engineering students using the program for their thesis work, graduates with practical computational fluid dynamics modeling experience are immediately useful to industry. A major benefit to those working with the KIVA source code is that one person, Anthony Amsden, designed and wrote all the program versions of KIVA. The result is continuity of style and ease of readability.

KIVA programs are also in use by the major automotive manufacturers and many universities and laboratories in Canada, France, Italy, England, Germany, Sweden, Norway, Switzerland, the Czech Republic, Spain, Israel, Australia, India, China, Taiwan, South Korea, and Japan. KIVA user groups meet, publish newsletters, and maintain Web sites worldwide.

Conclusion

In May 1990, the then three-member KIVA team (Anthony Amsden, T. Dan Butler, and Peter O'Rourke) received a Los Alamos National Laboratory Distinguished Performance Award. This honor was followed in April 1993 by a Federal Laboratory Consortium (FLC) Award for Excellence in Technology Transfer. FLC awards recognize federal laboratory employees who have done an outstanding job of transferring technology developed in the laboratory to outside users, such as other government agencies or the private sector. We regard this award as a significant honor, considering the hundreds of federal laboratories across the U.S. that are members of the FLC and that only 28 submissions received awards that year. In 2001, the DOE included KIVA in its list of the top 100 innovations that have emerged from DOE facilities over its 23 years of existence.

Improving the submodels in KIVA-3V is an ongoing process, particularly in the areas of spray modeling, representation of complex fuel mixtures, chemical kinetics, and turbulence. Current applications focus on GDI and diesel engines. Although the automotive industry has been the driving force behind KIVA development over the years, new applications continue to arise, such as the analysis of stationary gas turbines for electric power generation.

Hindsight shows what has made KIVA such a success. Instead of working in isolation and then presenting the world with a finished combustion simulation program, we have worked steadily on a daily basis with the user community throughout the development process, responding to industry needs. The various versions of KIVA have progressed in a logical fashion, with significant increments in capability being added at each stage in the process. Finally, the availability of the KIVA source code has made it widely disseminated throughout the world, particularly in universities, so that today students who graduate with KIVA and other computational fluid dynamics skills are ready to apply their knowledge in industry.

KIVA-3V is distributed by the Energy Science and Technology Software Center at Oak Ridge, Tennessee.

Multiphase Flow in Fast-Fluidized-Bed Reactors Used for Petroleum Refining

by Bryan A. Kashiwa, Nely T. Padial-Collins, W. Brian Vanderheyden, and T. Daniel Butler

Butler
Kashiwa
Padial-Collins
Vanderheyden

Used in oil refining and manufacturing chemicals, the recirculating fast-fluidized bed (FFB) reactor is a common device to keep reactants well mixed for a controlled period of time. The Los Alamos computer code library, CFDLib, was used to analyze the flow and temperature distribution on FFB units. The large-scale simulation is a means to test the validity of the theories and to provide important guidance to the design and operation of modern equipment. For example, one industrial user has created a state-of-the-art performance analysis program for satellite thrusters using CFDLib.

Like grains of sand pulled along a streambed by the drag of flowing water, multiphase flow is the interpenetrating motion of dissimilar materials. Applications of multiphase flow span the range from astrophysics to zeolites and many interesting technologies in between. Typical examples are sediment transport in geology, pneumatic transport in engineering, corpuscle motion in the blood plasma of mammals, and space shuttle solid-rocket fuel combustion. In each of these examples, one material moves relative to another, resulting in motion. Like many physical effects in nature, this interpenetration has features useful in modern technology, as well as features that are destructive. One useful feature is the promotion of mixing in chemical reactors. A destructive feature is the erosion of pipe walls in pneumatic conveying equipment.

Recirculating Fast-Fluidized Bed Reactor

The recirculating fast-fluidized bed (FFB) reactor is a common device used in oil refining and manufacturing chemicals. The goal is to use the interpenetrating motion of gases and liquids, relative to a field of solid catalyst grains, as a method of keeping reactants well mixed for a controlled period of time. The wire-frame mesh (see Fig. 1) shows a generic configuration, having many of the features of the device as it is used in gasoline production.

In the generic configuration depicted in Fig. 2, solid catalyst grains circulate in a flow loop consisting of a cyclone separator with a gas exit at the top, a vertical standpipe that leads to an inclined feed pipe at the lower end of the vertical riser, and a horizontal transfer pipe that connects the top of the riser to the cyclone, completing the loop. The catalyst is heated and metered at an idealized regeneration point residing at the intersection of the standpipe and the inclined feeder. A fluidizing

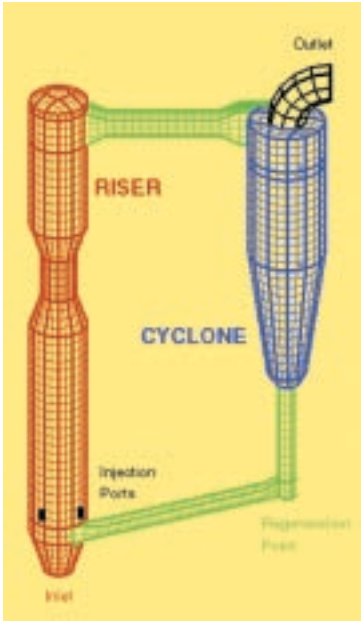


Fig. 1. The figure shows the wire-frame mesh of a recirculating fast-fluidized bed unit.

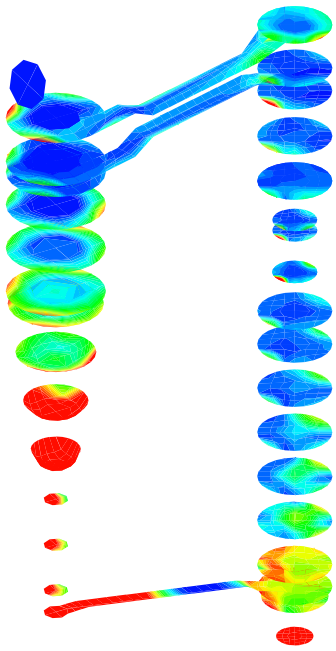


Fig. 2. The figure shows the contour plot of catalyst volume fraction. Red indicates high-catalyst concentrations. Blue indicates low concentrations. Note the high-catalyst concentration below injection zone in riser. Also note high concentrations of catalyst at the bottom of the cyclone, indicating good recovery.

gas is pumped into the bottom of the vertical riser in order to lift the close-packed bed of catalyst to an injection region just above the entry of the inclined feeder. At the injection level, oil is pumped into the riser. When the oil makes contact with the hot catalyst, it rapidly evaporates and accelerates up the riser carrying the solid grains with it in a highly turbulent fashion. The turbulence promotes contact between the solid grains and the oil vapor, which is "cracked" just enough to produce gasoline in the vapor phase. The length of time that the vapor and catalyst remain in contact is a crucial aspect of the FFB design. The time must be just long enough to produce the product weight of interest. Therefore, the cyclone separator is positioned in such a way that the vapor and solid phases are disengaged rapidly, sending the vapor to downstream condensing equipment and sending the solid grains downward toward the point of regeneration.

We used the Los Alamos computer code library, CFDLib, to analyze the flow and temperature distribution on FFB units. CFDLib is capable of solving a wide range of computational fluid dynamics (CFD) problems in two-and three-space dimensions. All flow-speed regimes are accessible in the library, ranging from fully incompressible to hypersonic; and code volumes exist that enable multifluid and multiphase computations with an arbitrary number of fluid fields, each with their own set of conservation equations.

Items of Interest

The main item of interest is the conversion efficiency of the liquid or gas that is injected at the bottom of the riser. Other items include the potential for fouling of surfaces, erosion, and energetics of the operation. All such performance items can be affected by changes in vessel geometry. In this case (see Fig. 3), a restriction in the vertical riser has been added to demonstrate the capability of CFDLib for handling complex geometries.

Physics of Interest

A realistic simulation of the FFB reactor is dependent upon the physical models used to represent the effects of chemical species conversion, physical kinetics of phase change, granular flow, and multiphase fluid turbulence. The equations that embody these physical models are developed using a combination of detailed mathematics, definitive laboratory experiments, and physical intuition. The large-scale simulation is a means of bringing together these diverse sets of information in order to test the validity of the theories and to provide important guidance to the design and operation of modern equipment.

More Complex Applications

There exist a large number of problems that are even more complex than the FFB reactor, such as the burning of pulverized coal in a furnace. These more complex problems involve even more interactive flow processes, like radiative heat transport. CFDLib is designed to serve as a vehicle for the exploration of problems in complex material dynamics and has served to spawn the creation of very specialized analysis tools. For example, one industrial user has created a state-of-the-art performance analysis program for satellite thrusters using CFDLib.

There also exist many applications that are a subset of the FFB application, such as the two-phase flow in human cardiovascular systems, or the dynamics associated with a lifeboat dropped onto the sea from a search-and-rescue aircraft. These and many other contemporary applications in modern technology can be addressed by using CFDLib. Some current applications include the smelting of iron ore, alumina precipitation, combined granular and fiber flow in manufacturing, and the effects of a near miss in the performance of defensive missiles.

Fig. 3. In the contour plot of gas temperature shown, the red indicates high temperatures; blue indicates low temperatures. Note the temperature maldistribution in the riser.

An Example of a Successful Industrial and Government Collaboration

by Norman L. Johnson

N. Johnson

Because the Theoretical Division was the birthplace of many successful initiatives at Los Alamos (e.g., Computing Division, Nuclear Safety Division, Human Genome Project), it may not be surprising that the division was often a leader in its outreach to industry, academia, other nations, and other government agencies (besides the Atomic Energy Commission and then the nuclear weapons component of the U.S. Department of Energy). Much of the vitality of T-Division's research and applications came from these collaborations. Sometimes these collaborations built upon the dual use of the Laboratory's core mission technologies, and sometimes they developed new capabilities for the core mission.

Combustion-engine modeling is one example of our collaboration with industry and government. Early in the 1960s, the fluid modeling within the core mission programs had matured to the point of expanding to new applications. One of the first undertakings by T-Division was in reactive flows, and these codes were some of the first computer codes to be available outside the Laboratory. By the 1970s, the identification of combustion modeling with T-Division attracted interest by DOE's energy programs and by industry. The synergy of industrial interest and government sponsorship developed a unique technical expertise in T-Division. The resulting combustion simulation codes, exemplified by the KIVA family of codes, became one of the most well known and most broadly used codes developed by the national laboratories.

A very successful collaboration between the Laboratory and industry was, and continues to be, with Procter & Gamble (P&G). In August 1994, P&G approached T-Division to assist them in developing a multiphase modeling capability of their diaper-manufacturing process with the goal of reducing costs by quicker fault analysis of existing processes and by eliminating costly hardware design steps in developing new manufacturing processes.

Initially, both parties agreed that the problem was much too challenging for current capabilities. But from P&G's insistence that there were no other resources available in the world to take on such a problem, T-Division assembled a team to collaborate with P&G to achieve the unachievable. By 1999, all previous expectations had been exceeded, culminating in a simulation of the entire geometry of interest and all of its components and materials, representing one of the most advanced

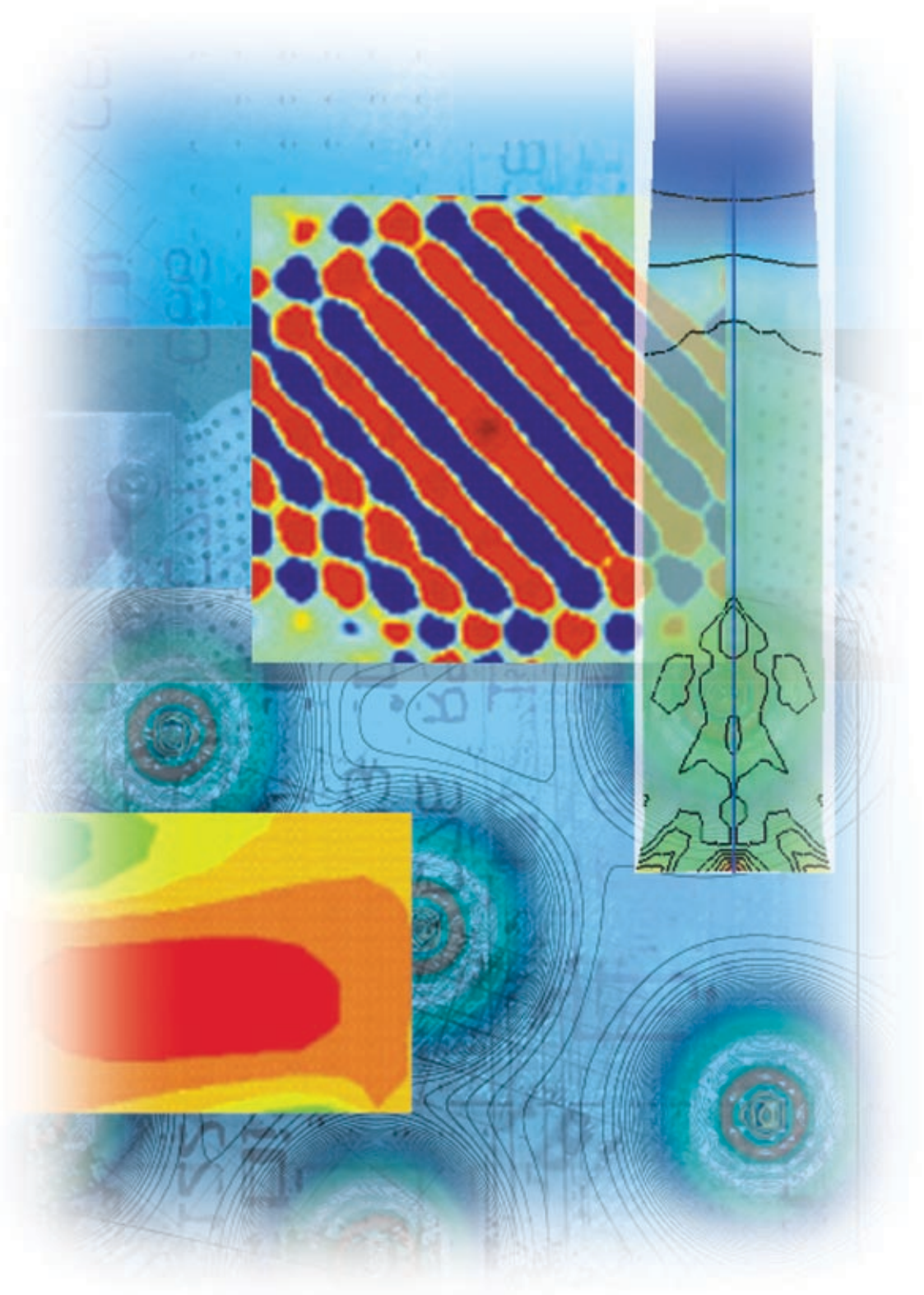
and complex three-dimensional multiphase simulation capabilities developed in the world to date. The capability included a gas phase, two solid phases, and moving internal parts within a complex geometry, all described by a newly developed multiphase stochastic transport model.

Although the multiphase flow project was wholly funded by P&G, the success extended to the scientific community at large, to other industrial interests, and to the Laboratory's programmatic efforts. The T-Division project has become a model example of how the national laboratories can be funded directly from industry, yet all constituencies benefit. The T-Division team won a Los Alamos Distinguished Performance Award for the project's success.

The project focused on maturing and tailoring the capabilities of an existing code, CFDLib, for the application of interest. The developed simulation capability significantly reduces the costs of trouble-shooting problems with existing processes and in the design of new processes. The project required coordination with the computational and experimental staff at P&G and an experimental effort at Purdue University.

The agreement with the industrial partner for the project was that the general development of CFDLib (excluding proprietary extensions) would be made available to the scientific community at large and to other industrial programs. The multiphase methodology developed in the project and the unique capability of CFDLib has been successful. The CFDLib team supports this outreach, both to direct collaborators and to the broad user community. As an indication of the interest in the code, a CFDLib email mailing list is available to the community. There are 163 people who have voluntarily subscribed to the list, about half in government and industry and the other half at universities.

In all, the CFDLib team has provided exemplary fulfillment of the needs of an important industrial sponsor, as demonstrated by the increased level of support by the sponsor at Los Alamos. Furthermore, the team has been able to leverage progress in this project to both programmatic efforts at Los Alamos and to the industrial and scientific community at large. Team efforts have increased both the scientific and public presence of the Laboratory in the national and international community. Overall, the project is a model example of how mankind can benefit from collaboration between a national laboratory and industry.



Introduction to Material Properties

Materials that concern modern researchers—including researchers working in the United States' nuclear weapons program—are an almost endless set of metal alloys, composite materials, and chemical compounds in solid, liquid, gaseous, or plasma phase. In the Los Alamos Theoretical Division, we have focused particularly on the response of these materials to mechanical, thermal, and radioactive insults at strain rates that vary from extremely slow (in the earth's mantle) to extremely fast (in a nuclear explosion). Often scientists need descriptions of internal forces that result from the application of external influences.

The challenge is to relate the effects of fine-scale structure of the material (individual crystals in a metal, grains, and binding material in a composite) to the overall bulk stress of a macroscopic chunk of material. Even in an isolated homogeneous material, the large-scale response is often complicated by the effects of atomic structure, which can change (e.g., melt) during a dynamical process.

An especially famous and important contribution of T-Division has been the development of "equations of state" (EOS). An EOS describes, in its simplest terms, the pressure that develops in a material because of variations in its density and temperature. The EOS can sometimes be written as an algebraic relation, but most often it is expressed by tables of numbers. We describe these tables in this "Material Properties" chapter in relation to the Los Alamos SESAME tables that are used internationally for a remarkably broad set of applications.

There is, however, another type of material deformation that is equally important, occurring within the interior of so-called solids. Rubber, for example, is composed of tangled molecular chains whose links constantly vibrate and rub against adjacent chains. Glass is the extreme ultimate of a viscous fluid, sometimes requiring centuries of high pressure to accomplish even slight deformations (except of course at high temperatures). Metals, however, are perhaps the most interesting of the complex solids, and at the same time are of great importance to Laboratory projects so that their analysis occupies a significant position in T-Division activities. Virtually any chunk of solid metal consists of crystals ranging in size from a micron (a millionth of a meter) to as large as several centimeters. Each crystal can be mildly to extremely nonisotropic

in its structure, so that for example the sound speed may vary by as much as a factor of three in the various directions within a crystal. When subjected to a mechanical insult, the atoms in each crystal may respond in any of a variety of ways. These include simple elasticity, plastic flow, microfracturing, growth of pores, twinning, complete tearing apart (spall), and the transition from one crystal structure to another (i.e., a solid-solid phase transition). These important and intriguing processes are described in several of the sections of this chapter. All of these processes become especially challenging when the metal is a binary alloy or a heterogeneous mixture of tiny chunks or fibers of two or more different metals or of a metal in some organic matrix.

The Mechanical Response of Cellular Materials

by Mark W. Schraad, Francis H. Harlow, and Cheng Liu

Harlow
Liu
Schraad

Theoretical Division has developed a modeling approach that can be used for describing the mechanical response of cellular materials to highly dynamic, large-strain and high-strain-rate loading conditions. This modeling approach is being used for Laboratory programs simulating the mechanical response of stress pads and cushioning materials.The constitutive model has been implemented into a small-scale test code and a large-scale performance code.

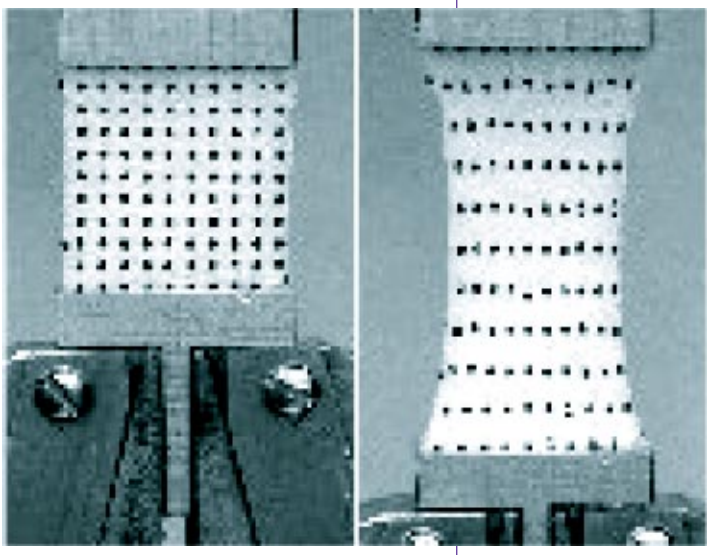


Fig. 1. A block of foam has been marked with a regular array of ink spots and subjected to a tensile stress, with results that show the evolution of strain.

The purpose of this project is to develop a modeling approach that can be used for describing the mechanical response of cellular materials to highly dynamic, large-strain and high-strain-rate loading conditions (see Figs. 1–3). The programmatic relevance of this project derives from the implementation and use of this modeling approach in computer codes for simulating the mechanical response of stress pads and cushioning materials.

Cellular materials are viewed as consisting of two separate but interacting material fields—the cellular solid and the permeating fluid (e.g., usually air)—each with its own distinct velocity field. The modeling effort, therefore, begins with a governing set of continuum-level conservation equations derived in two-field form. For example, the linear momentum equations for the cellular solid and the fluid describe for each material the response to fluid-pressure gradients and to the drag impedance arising from the relative flow between the materials. In addition, the evolution of momentum for the cellular solid includes the effects of stresses within the material.

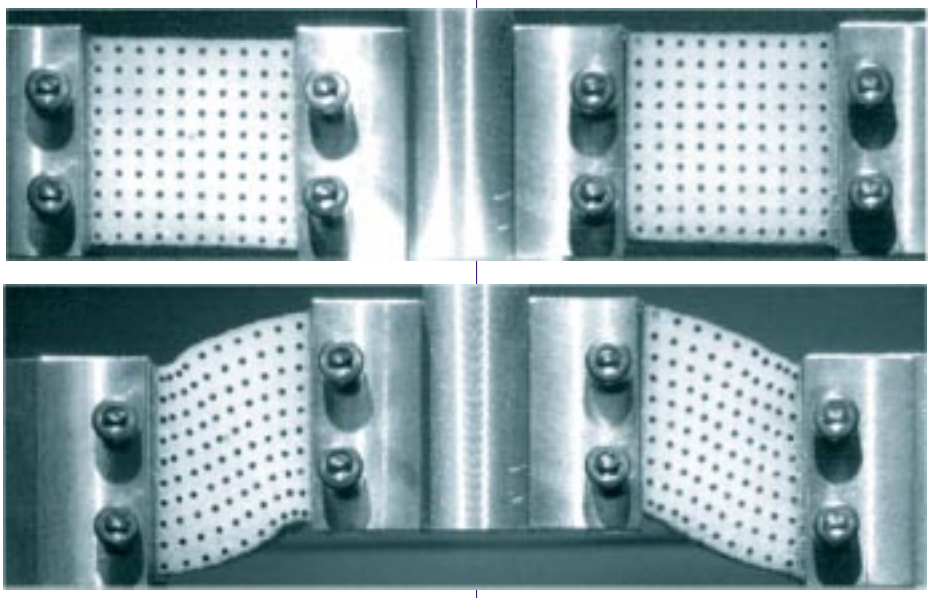


Fig. 2. The block of foam has also been subjected to various combinations of compression and shear.

The form of the governing equations is chosen to allow for consideration of evolving closed-cell fluid pressure and open-cell fluid flow in the overall description of the cellular material response. Solutions to the linear momentum equations and the corresponding conservation of mass equations provide for each material the changing velocities, densities, and volume fractions. In addition, it is necessary to determine the local pressure gradients, the changing drag-impedance coefficient, and the current state of stress in the cellular solid.

A general, three-dimensional, finite-deformation constitutive law has been developed to describe the mechanical response of a single foam cell. A probability distribution function, which describes the initial cell-size distribution and the corresponding cellular-scale structural evolution, has been introduced, and its variations have been formulated to be coupled with the single-cell model to form the global-scale constitutive law, which relates the overall stress in the bulk material to the bulk strain, strain rate, and stress-evolution rate. This analysis technique is different from any other that has been developed for foam characterization.

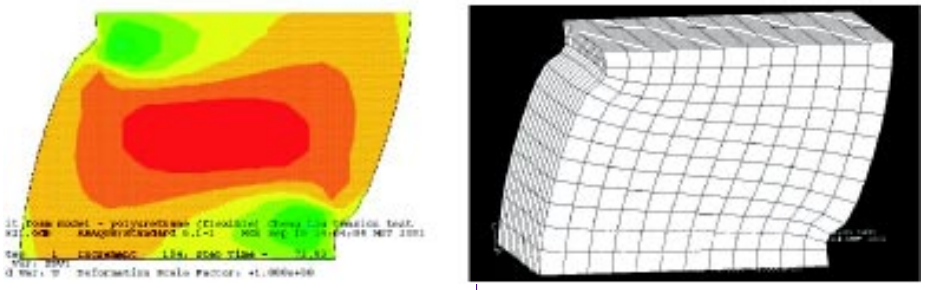


Fig. 3. Numerical calculations show the predictions of deformation.

As the cellular structure evolves, so too does the probability distribution function, thereby bridging the cellular and global length scales by representing the cellular-scale structural evolution within the global-scale constitutive response.

The general, 3-D, rate-dependent, global-scale constitutive model has been implemented within a small-scale test code and a large-scale performance code. The material model parameters have been calibrated against both single-cell numerical simulations and global-scale experimental data, and small-scale tests have been performed for parameter refinement. The model also has been verified and validated against experimental data for low-rate but large-deformation compressive, tensile, shear, and combined loading of a commercial open-cell polyurethane foam. In the near future, attention will be focused on further characterization of the rate-dependent behavior of cellular materials, implementation of models for materials with closed-cell structures, and testing the models for applications involving foam materials of relevance to both Laboratory and industrial applications.

Albers
Boring
Ericksson
Hill
Johansson
Kmetko
Skriver
Weinberger
Wilkins
Wills
J. Wood

The Electronic Structure of the Actinides

by John M. Wills

The electronic structure of a material, that is, the energy levels available to the electrons, determines all the properties of that material. An understanding of the electronic material provides the basic underpinning for understanding the structural properties and the equation of state of a material. Although the equations governing electronic structure are known, their exact solution—involving approximately 1022 electrons and ions—is beyond reach. For this reason electronic structure theory has evolved within a context of approximate systems and many-body perturbation theory.

Los Alamos, T-Division in particular, has played, and continues to play, an important role in untangling the mystery of actinide electronic structure—the last elements in the periodic table to be successfully described. Although the success is still partial, we are involved in the study of the most interesting element in the periodic table—plutonium.

An important discovery to come out of Los Alamos was that the ground-state electronic structure of the elemental actinides thorium through plutonium could be quantitatively described using density functional theory in the local density approximation by performing detailed calculations, which included all relativistic effects.

T-Division and collaborators are in the forefront of efforts to understand the transition from itinerant to localized f-electron behavior in plutonium. This effort has been given special emphasis by the Laboratory goal of science-based stockpile stewardship.

The electronic structure of a material—the energy levels available to the electrons—determines all the properties of that material. Electrons shared between atoms bond the atoms together, and the electronic structure determines the crystal structure and the way atoms interact in a crystal, hence underpinning all the structural properties and the equation of state of a material. An understanding of the electronic material thus provides the basic underpinning for understanding the material as a whole. Although the equations governing electronic structure are known, their exact solution—involving approximately 1022 electrons and ions—is beyond reach, so electronic structure theory has evolved within a context of approximate systems and many-body perturbation theory.

Despite the many-electron nature of the electronic structure problem, most of the elements in the periodic table have been successfully described by an independent-electron model. Density functional theory (DFT), which established a relation between the electron density and energy of a material, and the development of computers and computer algorithms capable of solving DFT equations, have provided a theoretical basis for the independent electron description. This theory, as implemented with the local density approximation (LDA), provides an excellent qualitative (and a good quantitative) description of a broad range of materials, treating electron correlation in an average way. It fails when strong electron correlation dominates the properties of a material, as in superconductors. The actinides—the last series in the periodic table, starting with thorium, and including plutonium—were the last elements in the periodic table to be successfully described, and the success is still partial, and involves a study of the most interesting element in the periodic table—plutonium.

The actinides have properties that make them unique in the periodic table. The trend in the volumes and cohesive energies of the light actinides (thorium through plutonium) is indicative of bonding f-electrons, while the post-plutonium actinides have a trend like the rare earths, indicative of nonbonding f-electrons. The crystal structures of the light actinides are low-symmetry structures with multiatom unit cells. The heavy nuclei of the actinide elements make relativistic effects (like spin-orbit coupling) essential in describing the electronic structure. In addition, actinide elements form a wealth of compounds that display interesting correlated electron behavior, such as heavy fermion, mixed valent, and unusual magnetic behavior. Finally, elemental plutonium, which straddles the transition from itinerant to localized behavior, displays a number of anomalous properties, including six structural phase transitions with temperature that span a volume change of 25% over the zero temperature state and negative thermal expansion in the δ phase. All of these unusual properties of the actinides have made their theoretical description difficult and uncertain.

Los Alamos, T-Division in particular, has played, and continues to play, an important role in untangling the mystery of actinide electronic structure. From the first quantitative calculations of Kmetko and Hill,

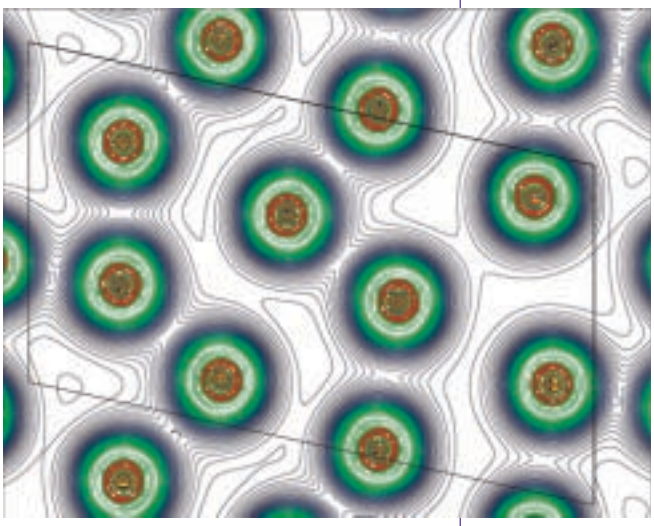
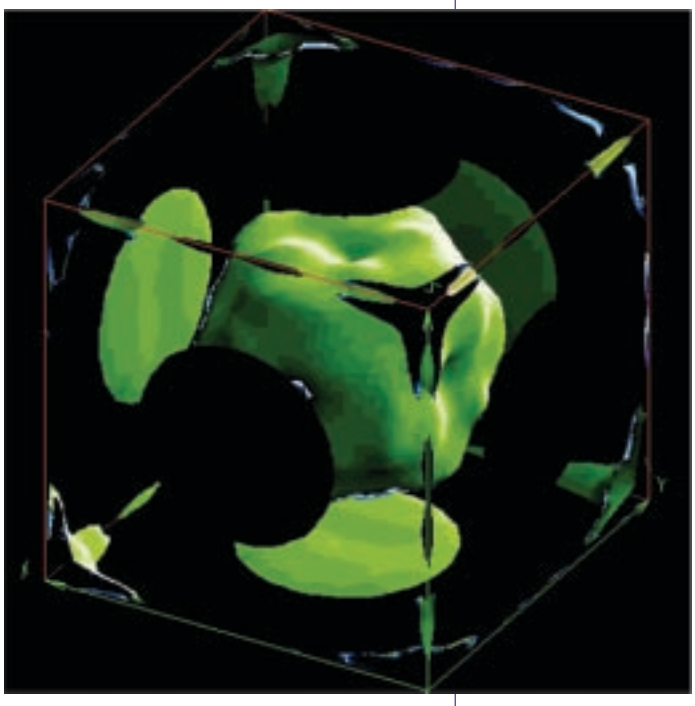


Fig. 1. Charged-density contour plot for a plane perpendicular to B-axis in alpha (grade) plutonium calculated with electronic structure theory.

[1, 2] pioneering calculations by John Wood, Mike Boring, and Bob Albers, and I developed numerical methods and probed the electronic structure of actinide elements and compounds. Through the efforts of these researchers and their university collaborators, such as Peter Weinberger (Vienna), Hans Skriver (Denmark), John Wilkins (Ohio State University), and Borje Johansson and Olle Eriksson (Uppsala Universitet, Sweden), our current state of understanding of the actinides has emerged.

An important discovery to come out of Los Alamos was that the ground-state electronic structure of the elemental actinides thorium through plutonium could be quantitatively described (including the unusual crystal structures and volume trend) using density functional theory in the local density approximation by performing detailed calculations which included all relativistic effects. [3] An important step in this work was the correct description of the ground-state crystal structure (the α-Pu crystal structure) of plutonium [4] (Fig. 1). This established the itinerant nature of the ground state of the light actinides, and made significant the failure of LDA to describe the high-temperature and alloy behavior of plutonium. To highlight this success, T-Division and Uppsala University, using an electronic structure method developed in T-Division, showed that the highest temperature charge density wave transition observed in uranium could be described quantitatively using an itinerant f-electron theory. [5] Finally, it was shown that the low-symmetry crystal structures found in the light actinides are a consequence of a high density of states at the Fermi energy through a Peierls-like distortion rather than being driven by an f-spatial shape. [6]

Fig. 2. Three-dimensional surface calculated for the 115-compound $CeRhIr_5$.



T-Division and collaborators are in the forefront of efforts to understand the transition from itinerant to localized f-electron behavior in plutonium. This effort has been given special emphasis by the Laboratory goal of science-based stockpile stewardship, which ultimately requires an ab initio equation of state for plutonium, including the correlated phases. A recent theory of correlation in actinides [7] has been used to describe δ-Pu quantitatively and is being applied to plutonium compounds in order to develop a complete theory of correlation in plutonium (Fig. 2). Researchers in T-Division are also investigating magnetic theories of correlation. T-Division is also working with Rutgers University and others to investigate the predictions of the dynamical mean field theory of electron correlation in plutonium.

Dynamic Deformation of Materials

by Francis L. Addessio

Addessio
Albers
Germann
Holian
Lomdahl
Lookman
Mason
Maudlin
Saxena
Tome
Williams
Zuo

Los Alamos scientists are developing predictive, physically based material models. These mathematical models are implemented into computer analyses and used to address safety, delivery, and performance issues for the nation's nuclear stockpile. Efforts to develop advanced models have involved many divisions within the Laboratory and collaborations with industry and academe. A number of complex phenomena must be modeled for the many materials that are of interest to nuclear weapons technology programs at the Laboratory, including material phase transitions and crystal reorientation (twinning) phenomena at high rates of deformation. To address the dynamic response of these materials, theoretical investigations of length scales from nanometers (atomistic) to meters (structural) coupled with experiments are being pursued.

Introduction

Materials may be found in many phases. As an example, water exists in a solid (ice), liquid, or gas phase, depending on its pressure and temperature. Similarly, metals exist in many solid phases, depending on the arrangement of their atoms. Plutonium-gallium alloys that are relevant to weapons design, for example, can be found in six solid phases. An understanding of the basic structure of materials begins at the atomistic length scale, which is on the order of nanometers (10^{-9} meters). As the load in a material is increased beyond the elastic limit, it experiences a permanent deformation. This is the result of the generation and motion of defects or dislocations, which are found at the atomic length scale. The next length scale, which is necessary for understanding the deformation characteristics of metals, is that of a grain or crystal. Grains are formed at nucleation sites as metals are cast from a hot liquid state. This process may be compared to the formation of bubbles at nucleation sites at the bottom of a hot pan as water boils. The relative size of grains in most metals is on the order of micrometers (10^{-6} meters). These grains may exhibit a directionality or texture depending on the forming process used to generate an engineering structure. This texture leads to strong and weak directions in the material, similar to a wood beam, which exhibits different strength characteristics depending on whether it is loaded along or transverses to its grains. Furthermore, if the atomic structure (phase) of the metal is in a nonsymmetric arrangement, the substructure of its grains may be found in different orientations or variants. As the metal is loaded, variants aligned with the direction of the force grow at the expense of the remaining variants. This is referred to as a detwinning process. Consequently, an understanding of the deformation characteristics of metals, including elasticity, phase transformations, crystal reorientation, plasticity, and failure, at the

macromechanical length scale (on the order of 10^{-2} meters) begins with a knowledge of the atomistic and granular characteristics of the material.

Objective

Los Alamos scientists are developing predictive, physically based material models. These mathematical models are implemented into computer analyses and used to address safety, delivery, and performance issues for the nation's nuclear stockpile. Efforts to develop advanced models have involved many divisions within the Laboratory and collaborations with industry and academe. A number of complex phenomena must be modeled for the many materials that are of interest to the Laboratory's nuclear weapons technology programs. Among the plethora of phenomena that are being addressed are material phase transitions and crystal reorientation (twinning) phenomena at high rates of deformation. In an effort to address the dynamic response of these materials, theoretical investigations of length scales from nanometers (atomistic) to meters (structural) coupled with experiments are being pursued.

Solid-solid phase transformations (SSPT) occur in many metals, including actinides, steels, and shape-memory alloys. These transformations result in changes to the crystal structure in microscopic regions within a single crystal. The SSPT processes are a complex function of the bulk temperature, bulk pressure, and time. The resulting changes in the single-crystal atomic structure cause complex changes in the overall bulk behavior of the transforming polycrystalline material. In order to model metals that undergo SSPT in structural applications, it is necessary to have appropriate models that account for the effects of SSPT on the overall response of the material.

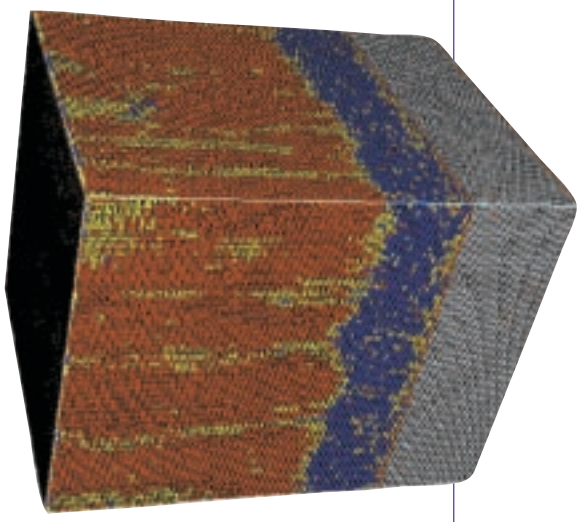


Fig. 1. Molecular dynamics simulation of phase transition in iron.

Atomistic Length Scale

At the smallest length scale (10^{-9}), molecular dynamics (MD) simulations are being conducted to investigate shock-induced phase transitions. Atomistic simulations of shock-induced phase transitions in iron have been used to study the transition of a body-centered cubic (bcc) atomic structure to a hexagonal close-packed (hcp) structure. Sample sizes of 10^8 atoms routinely are considered.

Atomistic simulations of a shock propagating into the α phase (bcc) of iron using an empirical atomic potential for iron are provided in Fig. 1. Improved potentials may be obtained by pursuing quantum mechanical analyses, which are obtained on a smaller (10^4 atoms) length scale. The shock wave is generated by the impact of an infinitely massive piston on the left. The shock propagates from left to right in the figure. A distinct split two-wave structure is observed. The leading wave is perfectly planar within one atomic diameter. This wave uniaxially compresses the unshocked atoms in a reversible (elastic)

manner. The second, slower wave is nonplanar and transforms the compressed bcc structure to a close-packed structure. Incipient grain or twin boundaries are provided in yellow. It is observed from the simulations that the shocked state is in a twinned hcp (ε-phase) structure with several stacking faults giving a local face-centered cubic-like (fcc) structure. A closer inspection indicates that the bcc planes orthogonal to the shock direction transform to hcp planes, rotated by ± 45 degrees so that a stable twin plane develops over time. It also is observed that this transformation is reversible, as demonstrated by the unloading, which occurs after the shock reaches the free end of the sample. This relief wave restores a bcc structure, but at an elevated temperature.

Crystalline Length Scale

Molecular dynamics analyses can be used to guide free-energy formulations, which may be used to investigate phase transitions at the length scale of a single crystal (approximately 10^{-6} meters). The Ginzburg-Landau (GL) theory models the free energy of a transforming crystal structure in terms of an order parameter. The theory uses the group-subgroup relations between the symmetries of the two phases. Besides the free-energy difference between the two phases as a function of temperature and pressure, it also includes interfacial energies and the elastic energy cost of the phase transformation. The fundamental elastic fields are related in a nonlocal fashion through the use of the compatibility relations of geometrically linear elasticity. The GL formalism has been extended in a consistent fashion to incorporate the effects of mechanical loading and thermal dilatation effects on the single crystal behavior. It is envisioned that future work on the GL theory will incorporate the effects of plasticity on the SSPT behavior within single crystals. A typical microstructure predicted by the GL theory in a single crystal subjected to a stress or strain defect is provided in Fig. 2. The defect nucleates the phase transition. Because of the nonlocal effect, phase transitions within the crystal are nucleated at many locations. Above the transition temperature, a long-range elastic field develops around the initial defect and a tweed-like pattern evolves. At a later time, a self-accommodating or twinned-structure is observed.

Polycrystalline Description

In an effort to develop engineering material models, the response of a collection of crystals subjected to large deformations must be obtained. For this purpose, homogenization techniques are employed to assemble many crystals or grains and determine the polycrystalline response (approximately 10^{-4} meters). The Ginzburg-Landau theory is used for the response of the individual crystals. A general homogenization theory has been developed. This homogenization approach uses a simple cubic Representative Volume Element (RVE) to model the polycrystalline behavior. (See Fig. 3.) A fully coupled, thermomechanical

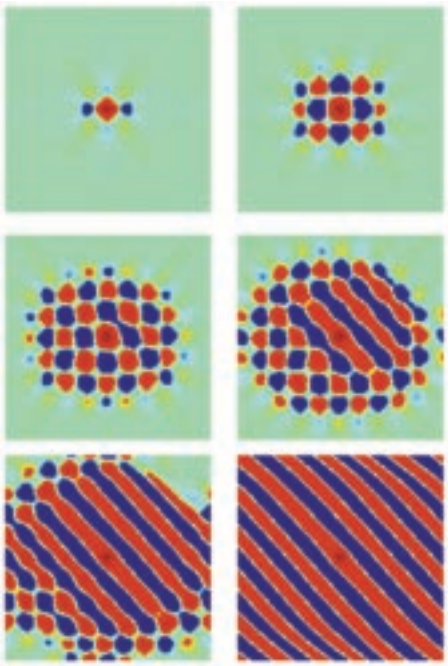


Fig. 2. Simulations of defect-induced transitions in a single crystal using Ginzburg-Landau theory.

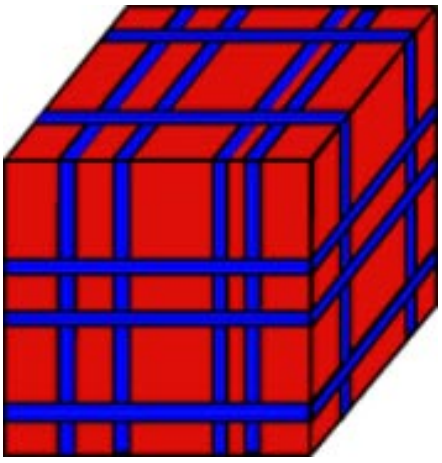


Fig. 3. Representation volume element used to determine polycrystalline response (Ginzburg-Landau theory provides single crystal [subceII] response, upper left).

theory was developed for this purpose. The homogenization model assumes the existence of a deterministic polycrystalline structure and initially models each crystal (subregion) in the polycrystal as a distinct subregion. The homogenization analysis satisfies the governing equations of linear continuum mechanics (equilibrium and energy) in an integral sense within each single crystal in the polycrystalline microstructure. The responses within the different crystals are coupled through the interfacial constraints on displacements and stresses and the assumption of periodicity. The theory is capable of analyzing general microstructures and it can, therefore, be used to study localization effects within a polycrystalline structure.

The coupled GL/homogenization theory will provide the necessary tool to study solid-solid transformation processes. Phenomena within the individual crystals and within a polycrystal can be observed. The interactions of the single crystals and how deformations impact the bulk behavior of the metal as a function of the bulk thermomechanical loading history will be studied.

Macromechanical Length Scale

At the largest length scale (approximately 10^{-2} meters), material models, which are computationally robust and efficient, are being developed. Macromechanical models are implemented into finite-element (FE) and finite-difference (FD) computer codes. These codes are used to analyze the deformations of engineering structures, which are relevant to nuclear weapons applications. The micromechanical theories coupled with experimental data are important to develop the necessary mechanical potentials, phase diagrams, and material parameters, which are required by the macromechanical models. Material models based on the approach of classical plasticity are being developed. These models typically undergo validation and verification by testing their ability to either reproduce the data used to generate the model parameters or to adequately predict the results of independent mechanical tests at different strain-rates or temperatures.

A comparison of the predicted and measured stress versus strain response is provided in Fig. 4 for small (0.1/s) and large (2000/s) strain rates of a uranium-niobium alloy (U-Nb). It may be observed that both reorientation (twinning) and slip plasticity are predicted by this material model. The Taylor cylinder experiment, where a solid cylinder is impacted against a rigid platen, is a validation test. In this

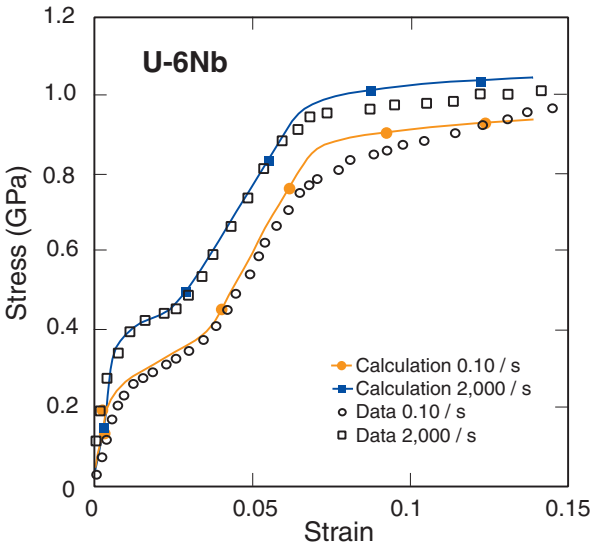


Fig. 4. Comparison of theory and experiment for a uranium-niobium alloy at small and large rates.

single experiment, the cylinder experiences a range of plastic strains, strain rates, and stress states as a function of time and location. (See Fig. 5.) The final shape of a U-Nb cylinder after impact at a velocity of 130 meters per second is shown in Fig. 5a. Finite-element simulations using classical approaches to modeling plasticity are not capable of accurately modeling the response of the U-Nb cylinder (Fig. 5b). The lack of agreement between the FE simulation and experiment results from the inability of the classical model to address reorientation (twinning) and phase-transition phenomena. An improvement to the classical plasticity model, which addresses the reorientation phenomena, is shown to provide improved agreement with the experiment in Fig. 5c. Current material modeling efforts will rely on experiments and the coupled GL/homogenization theory to construct better inelastic potentials in order to capture the shape-memory behavior (reorientation and phase transition) of U-Nb alloys.

As an example of an application of structure reorientation (deformation twinning), the three-dimensional formation of a shape charge jet is shown in Fig. 6. This application is of interest to the Department of Defense community. High-explosive detonation products load the convex side of a metallic shell in this code simulation producing a stretching jet of a very textured material that deforms through crystallographic

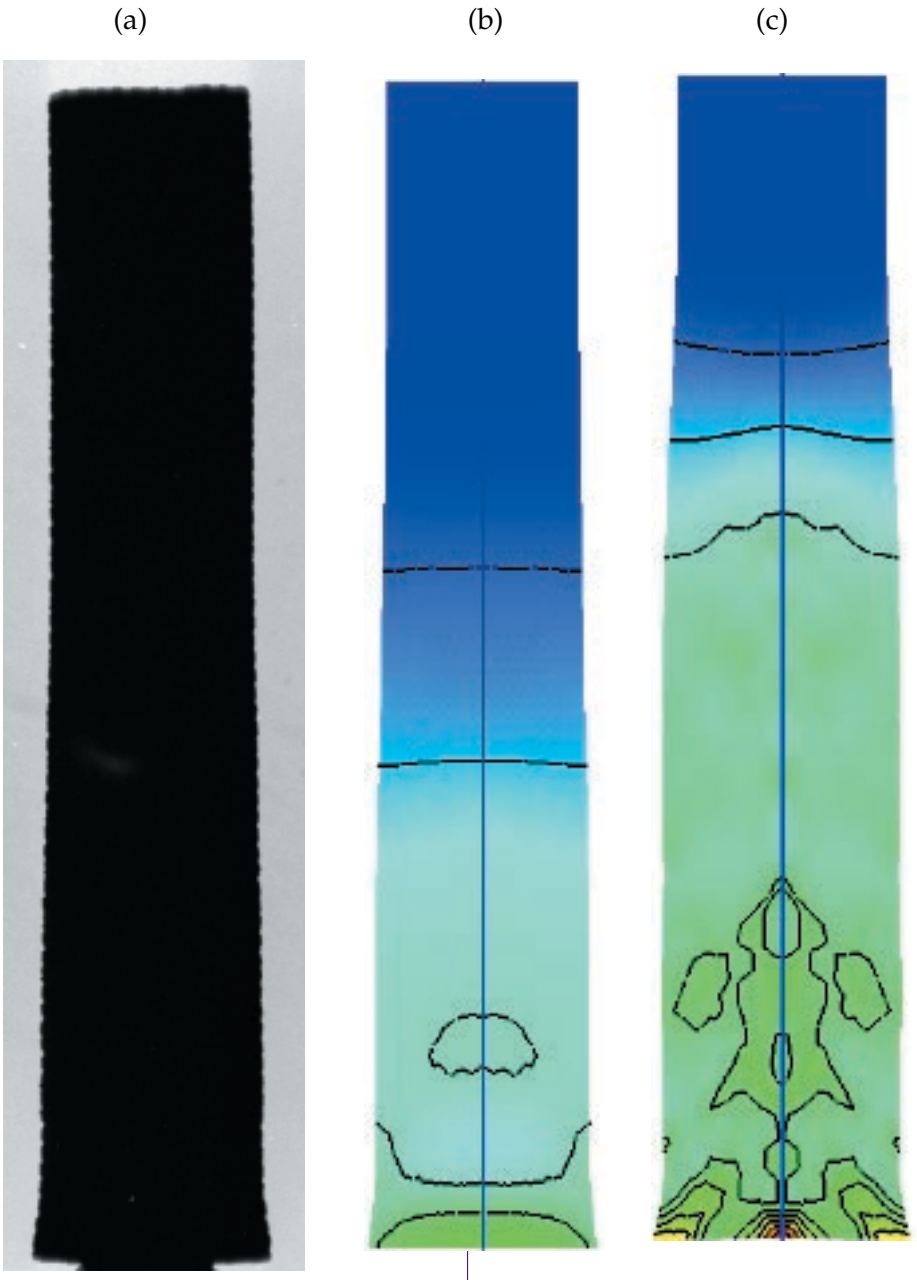


Fig. 5. Validation of macromechanical models using the Taylor impact experiment: (a) experiment, (b) classical model, and (c) improved model.

slip and twinning. Gradients of rigid body rotation in conjunction with the plastic anisotropy (texture) modeled in this simulation produce variations in cross-sectional shape along the longitudinal axis of this jet.

Other contributors to this work include: Lomdahl, Germann, and Holian (atomistic modeling); Albers, Saxena, Lookman, and Tome (crysalline modeling); Williams and Tome (polycrystalline modeling); and Addessio, Maudlin, Mason, and Zuo (macromechanical modeling).

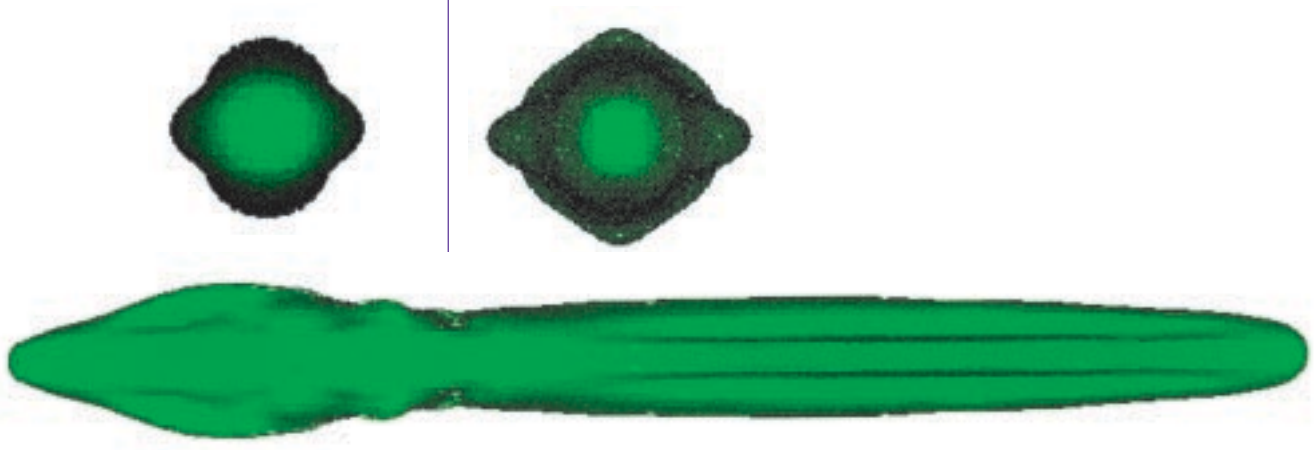


Fig. 6. Shape charge jet formation application.

Material Flow, Failure, and Fragmentation

by John K. Dienes

Blewett
Daly
Dienes
M. Evans
Griffith
Harlow
Hill
Kashiwa
Kershner
Middleditch
Orowan
Polanyi
Prager
Ruppel
Seaman
Solem
Taylor
Zuo

With the advent of computers it became practical to solve the equations that govern material response numerically, and thereby to address nonlinear, inelastic problems such as fracture and plasticity. We have developed a unified tensor theory that accounts for both ductile and brittle behavior, including the response of both open and closed cracks to a general stress tensor. The theory is implemented in a computer algorithm called SCRAM (an acronym for Statistical CRAck Mechanics). The fundamental idea was to add the effects of many microscopic cracks in a statistical way analogous to what is done in physics with statistical mechanics. The SCRAM theory accounts for arbitrarily large strains and rotations by means of the polar decomposition theorem and its consequences. This theory is now in use throughout the world in finite-element calculations of structural integrity.

The most promising vehicle for advanced SCRAM calculations is a mixed method in which the SCRAM stress and fracture fields are tracked along material particles that flow in an Eulerean grid. This approach has been pursued by combining SCRAM with CFDLib, a hydro code developed by T-Division. This combination has promise for computing oblique impact involving brittle materials with failure planes in many directions.

In the early days of Theoretical Division, we were concerned with the behavior of materials at the very high temperatures caused, for example, by shock waves, nuclear reactions, or magnetic fields. The dominant materials problem was to represent pressure through an equation of state that accounted for ions, electrons, and radiation. Scalar equations of state were sufficient to model the flow of materials under these conditions (the properties appeared the same in all directions).

In the 1970s, however, problems arose that concerned the behavior of matter at more nearly standard conditions (lower temperatures), where scalar behavior is not adequate because the stresses and strains are significantly different in different directions. For these problems, tensor analysis is needed.

Would nuclear reactors survive the accelerations imposed by earthquakes, the impact of errant aircraft, or the forces imposed by tornados? How does radiation damage in reactors affect the strength of materials? What causes the accidental ignition of explosives and propellants following blast or impact? How do shaped charges perform when they impact tank armor? Can nuclear tests be contained underground? Can other nations conduct underground nuclear tests surreptitiously?

Can oil shale be economically broken up and converted into fuel for cars and aircraft?

These and many other such questions required some knowledge of the flow and failure properties of solids at modest temperatures, well below melting. The problems cited above concern the strength of materials under conditions much more complex than covered by the standard methods of elasticity, which were the traditional basis for the design of buildings, bridges, and other conventional structures.

Background

Plasticity, which governs the ductile flow of metals, during fabrication for example, became a subject of theoretical study in the late 19th century. The physical basis for plastic flow, the motion of flaws at the atomic level that came to be called dislocations, was first described by Taylor, Orowan, and Polanyi independently in 1934. Interest accelerated during the Second World War with the need for better cannons, armor, and shells. After that war, books by Hill and Prager stimulated further interest in plasticity, which became a standard subject for engineers, physicists, and mathematicians.

The systematic study of a very different kind of behavior, fracture, began around 1920 with the work of Griffith. Fracture and plasticity began to merge in the 1930s under the influence of Orowan, though the theories are entirely different; plasticity is a stable process that can be included in thermodynamics, but fracture is essentially unstable and requires a different approach. These early theories were primarily of research interest, and could not play much of a role in the practical design of engineering structures because the governing equations are strongly nonlinear, causing them to be generally insoluble by classical methods.

Solving Material Response Equations Numerically

With the advent of computers, it became practical to solve the equations that govern material response numerically, and thereby to address the nonlinear, inelastic problems mentioned above. Plasticity theory could be included in hydro codes (computer programs that simulate fluid flow) with only a few lines of programming, but the simple engineering theory in common use assumed a constant strength. This is not a good assumption under extreme conditions such as those involving shock waves. For example, under some conditions a weak wave, the “elastic precursor,” propagates ahead of the main shock, and to explain the observations accurately, it is necessary to account for the increase in material strength at high-strain rates and the decrease with temperature. This topic plays a major role in a variety of high-energy impact problems but, to this day, the motion of dislocations in the elastic precursor is not fully understood.

A quite different problem concerns the spall of materials, which can occur when a shock wave strikes a free surface, forcing a layer of solid to break off from the main structure, a problem that involves fracture rather than plastic flow. This process was represented by a dynamic tensile strength in the early hydro codes, using data from experimentalists. Eventually the coalescence of cracks to form spall layers was modeled by theorists in T-Division. Elasticity theory was included in studies of the response of pressurized nuclear reactors to earthquakes and the subsequent blowdown, and the resulting calculations by Daly and me were in superior agreement with experiments compared with more standard approaches.

The fracture of rocks due to shock waves and the subsequent tension (due to rarefactions and hoop stresses) needed to be understood because of the containment problems for nuclear weapons detonated underground. The need for an adequate fracture theory increased with the energy crisis that arose in the 1970s, when the extraction of fuels from oil shale appeared promising. Not only was it necessary to represent fracture in a computer simulation, but the connection of cracks to allow for flow through networks of cracks needed to be accounted for. Thus, the usual continuum theories would not do.

We attempted to apply the theory of Seaman at Stanford Research Institute, which involved the nucleation and growth of defects during spall, to such problems, but had little success because the stress field due to chemical or nuclear explosions involved simultaneous tension and compression (tension in one direction, compression in another). Consequently, we began to develop a computer algorithm that is called SCRAM (an acronym for Statistical CRAck Mechanics), which took some of Seaman's NAG-FRAG concepts, but cast them into a more systematic mathematical approach (tensor analysis) that addressed the simultaneous tension and compression (in different directions) by accounting for both open and closed cracks.

The fundamental idea was to add the effects of many microscopic cracks in a statistical way analogous to what is done in physics with statistical mechanics. The strength problem is far more complex, however, because the cracks are not necessarily stable, as are atoms, but can become catastrophically unstable. Furthermore, the theory requires the entire stress tensor to describe crack behavior, and not just the mean stress (essentially the pressure, which is adequate in gases). Thus, SCRAM needed to account for the opening, shear, growth, and coalescence of an ensemble of cracks with numerous orientations. Patterns of cracks formed in such a way can be readily seen in almost any rock.

The theory was presented at a rock mechanics symposium at MIT and was acknowledged in the summary of the meeting as a major contribution by Los Alamos. Of special interest to physicists was the role of percolation processes when the cracks become sufficiently connected and the permeability increases dramatically. In a subsequent symposium it was reported that the theory can be used to determine the

permeability of fractured rock, and thereby to determine the flow of gases or liquids through fractured rock, an essential ingredient to recovering fuels from oil shale. The energy crisis diminished and the oil-shale program came to an end, but the theory continues on in other applications. It has been confirmed in studies of sandstone with Gueguen in Strasbourg.

Using SCRAM to Study Violent Reactions

As mentioned above, explosives and propellants are subject to accidental reactions that can be even more violent than the planned detonations (in the case of explosives). This was the subject of "XDT" experiments that were initiated following the explosion of a Trident rocket motor that blew out windows in nearby Salt Lake City. (The X indicated that the violence was not understood, and the DT stood for detonation.) The thesis in this application of SCRAM was that the propellant was essentially brittle, though the brittle grains of explosive were embedded in a soft binder, giving the overall material a rubber-like character.

Standard theories of explosive reactions were not relevant because the pressure and temperature were so low that the overall chemical behavior was negligible. The SCRAM algorithm hypothesizes that small hot spots can form where the faces of cracks rub together and that these could initiate the observed violent reactions. Quantitative comparisons of the theory, embedded in the finite-element code PRONTO, explained the observed reactions and their statistical character (one in five shots exploded violently because of the larger defects in the one). This work continues in a current program concerning explosive safety, the high explosive violent reaction experiments (HEVR), and SCRAM shows promise for explaining the observed sensitivity.

Large Deformation

The SCRAM theory accounts for arbitrarily large strains and rotations by means of the polar decomposition theorem and its consequences. Interest in such theories was stimulated in the 1970s when I published a paper demonstrating that the standard theory of material rotation then current, and based on material vorticity, predicted violent instabilities that were quite unrealistic. By developing an exact theory for the rates of material rotation and strain, it became possible to characterize very large deformations in numerical calculations. This theory is now in use throughout the world in finite-element calculations of structural integrity. To illustrate the theory, recall that in hydrodynamics texts it is claimed that fluids that have no vorticity are "irrotational." The theory shows that this is not true in general, and that a line in an "irrotational" fluid will indeed rotate. As shown in the box, the rate of rotation is actually the sum of the vorticity and another term that is ignored in classical hydrodynamics. Papers in *Acta Mechanica* and an extensive Los Alamos report discuss details and examples.

The vector rate of rotation is given by

$$\omega = \mathbf{w} + (\mathbf{I}\, tr\mathbf{V} - \mathbf{V})^{-1}\mathbf{z}$$

with vorticity w, stretch V, and z the commutator of stretch and stretching.

The Statistical View in SCRAM

Another aspect of SCRAM is the statistical view, in which the overall rate of deformation is considered to be the sum of contributions from a variety of physical processes. Among these, in brittle materials, the theory must account for a large number of cracks of varying orientation, size, and opening. In fact, some cracks, the shear cracks, may have no opening at all. Though in some respects such cracks are difficult to visualize, they can be observed in most rocks and they are responsible for the fact that rocks are on the order of 10 times stronger in compression than in tension, while ductile materials such as metals have about the same tensile and compressive strength. The SCRAM theory shows that it is possible to prove a superposition theorem, which states that the strain rates that are due to various physical mechanisms can be added together, as can the rates of stretching, i.e., $D = \Sigma D^{\alpha}$. A version of this concept has been used in plasticity since the 1930s, often called Prandtl-Reuss theory, but the new result has three advantages: it holds for materials containing microcracks as well as dislocations; it is valid for arbitrarily large deformations; and it is based on mathematical and physical developments, not just put forth in an ad hoc manner.

These results may serve as a basis for more rational theories of material behavior in the future. One unexpected illustration of the superposition concept is in the highly nonlinear behavior of low-density foams at high strains, when the pores are nearly compressed out. It was observed by Solem that when stress in foams is plotted against strain, the curve is almost exactly a hyperbola, and he asked of me why this should be so. As a result of our discussions, the fact that strains due to pore compression and matrix elasticity can be added was demonstrated based on exact results in the theory of elasticity. The theory matched the deformation of three kinds of foams at many initial densities with good precision.

Simplified versions of SCRAM have become popular, but we are still pursuing a very general approach in which cracks having many orientations evolve and coalesce. In extreme cases, regions form where the microcracks connect to form very large cracks and the strength virtually disappears. Embedding SCRAM in an Eulerean hydro code allows for large deformations but has limited promise because materials at various damage levels often become mixed. SCRAM has traditionally been used in Lagrangean codes, but large deformations are not well handled.

Advanced SCRAM Calculations

The most promising vehicle for advanced SCRAM calculations is a mixed method in which the SCRAM stress and fracture fields are tracked along material particles that flow in an Eulerean grid. (See the chapter on numerical methods and algorithms in this book.) This approach has been pursued by combining SCRAM with CFDLib, a

hydro code developed by Kashiwa that is an extension of the early particle-in-cell work of Harlow and Evans in the 1950s. This combination has promise for computing oblique impact involving brittle materials with failure planes in many directions. New computers, such as Cray's MTA with its special architecture, may make these logically complex calculations efficient without extensive parallel programming.

Acknowledgments

I am indebted to a myriad of colleagues within and outside the Laboratory for technical contributions, discussions, and coping with recalcitrant computers. These colleagues are too numerous to cite individually, but you know who you are. Applications of SCRAM would not have been possible without John Middleditch, Jim Kershner, and Hans Ruppel. Pat Blewett helped unravel the mysteries attending large rotations, and Ken Zuo made it possible to understand the stability of penny-shaped cracks in all possible stress states.

Network Theory for Solid Polymers at High Strain Rates

by Duan Z. Zhang, Eric N. Harstad, and Francis H. Harlow

Harlow
Harstad
Liu
Zhang

A plastic-bonded explosive (PBX) is a composite of energetic grains and polymer binder. We have shown that the mechanical properties, especially the combined processes of strain, strain rate, and dissipative heating, have very important effects on the overall constitutive behavior of the PBX. We also demonstrated that it is important to physically understand the binder behavior, especially under high-speed deformations. Many important phenomena associated with high explosives, such as the propagation of shock waves, the development of hot spots, and the achievement of detonation conditions, need these constitutive relations in order to be characterized.

Direct measurements of the polymeric constitutive behavior from experiments at high-strain rates are difficult to conduct. Information about the relation at high-strain rates can be obtained indirectly from experiments such as flyer plate experiments. Experiments have also been carried out at lower-strain rates using a rheometer and then employing a time-temperature superposition principle to infer the behavior of the polymer at a higher strain rate. Although this principle is widely used, it is also known that many polymers do not obey it. For materials that do not obey time-temperature superposition, it is particularly valuable if guidance in modeling the material can be obtained from the basic principles of physics.

A typical polymer behaves in a viscoelastic manner. The stress depends not only on the strain but also on the strain rate throughout the history of the deformation. The history dependency can be represented by a time integral. The kernel of the integral is a decaying function of the time difference. One of the simplest models for a polymer is the well-known Maxwell model in which the kernel decays as an exponential function. While this model qualitatively describes the constitutive behavior of a polymer, it is found that to quantitatively predict the mechanical behavior of the polymer binder used in high explosives, a single relaxation time is not enough, and many Maxwell elements must be used. Our derivation of an extended Maxwell model is based on a statistical theory of molecular interaction. To accomplish this goal, we extended the current version of network theory for application to polymers, considering the effects of friction between polymer chains.

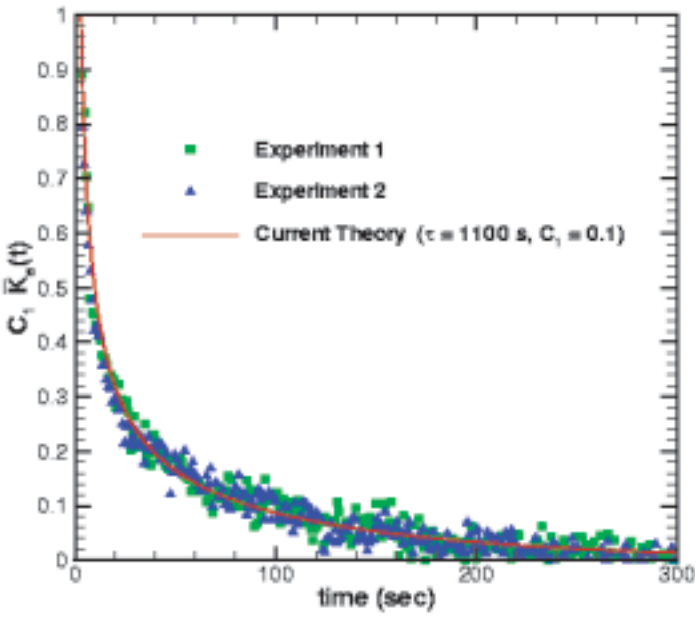


Fig. 1. Comparison of the calculated history integral kernel (experiment 1) with the experimental results of Cheng Liu (experiment 2). See explanation on next page.

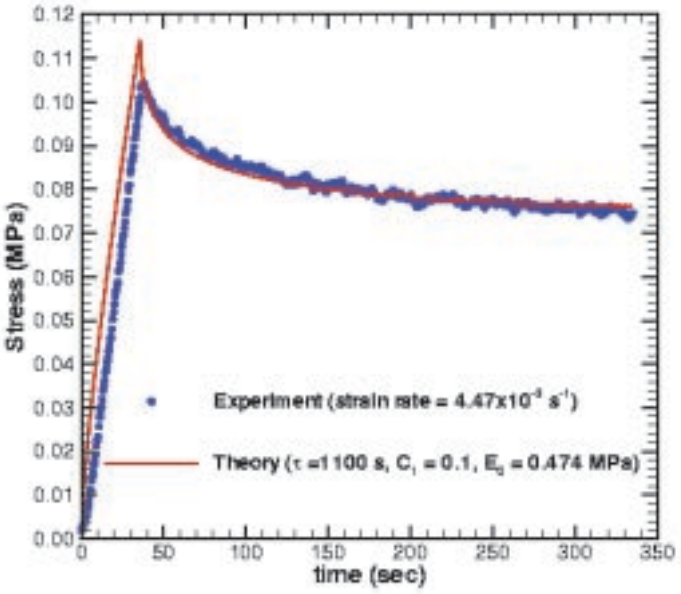


Fig. 2. Application of the theory to a compression-and-hold experiment performed by Cheng Liu.

The polymer system is modeled as a collection of springs imbedded in a viscous fluid. The elasticity of each spring is a result of thermal motion within the molecular chain; the viscous effect comes from the interaction between a polymer chain and its surrounding chains. When the spring is rapidly compressed at both ends, the deformation initially concentrates on the ends of the spring, and then gradually propagates into the center. Because of the concentration of the deformation, the force on both ends of the spring is large initially and then relaxes as the deformation approaches a uniform state. This theory predicts that, for a small time, the kernel of the history integral is proportional to the inverse of the square root of time, a much faster relaxation than in the single-element Maxwell model. For large times, the new kernel approaches an exponential function. The history kernel can be measured experimentally, by rapidly deforming a piece of polymer material and then holding its deformation. The decay of the stress is directly related to the kernel. Such experiments were performed by Cheng Liu of the Materials Science and Technology Division. Figures 1 and 2 show the comparison of our new theory to his results.

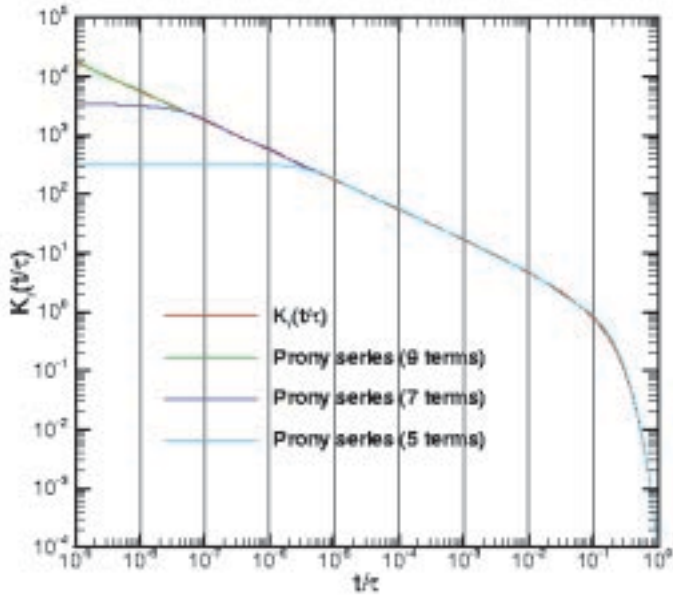


Fig. 3. Prony series compared with full kernel.

In many computer codes, the constitutive relation is represented by a Prony series. To convert our expression to a usable form in those codes, we represent our kernel by an expansion in Prony-kernel form. The result is compared in Fig. 3 with the full kernel expression. The principal significance of this analysis is that it describes for the first time the multiple rates that are empirically determined for the Prony series in terms of an expansion of a history-integral representation.

The results of this work have been to furnish for the first time an analysis technique applicable to the constitutive behavior of polymers subjected to large strains at large-strain rates.

Dynamical Response of a Polymer Chain to Rapid Distortions

by Eric N. Harstad and Francis H. Harlow

Harlow
Harstad

The stochastic response of a polymer chain to longitudinal extension or compression has traditionally been described by a combination of thermodynamics (for the thermal forces) and molecular dynamics (for the interatomic forces). For rapid strains, we have developed a description based on a Langevin equation for the thermal fluctuations, and a Liouville equation to follow the evolution of a probability distribution function.

Polymer chain nodes are in constant motion from thermal fluctuations. These fluctuations cause the chain to move through a region of space, which we refer to as a chain cloud. In quasi-static circumstances, the configuration of the cloud and the force required for deformation can be described on the basis of thermodynamic analysis. A polymer chain exhibits a rate-dependent component of chain-deformation that comes from viscous interactions of elements of the same chain and between an element of the chain with elements of other chains. These interactions impede the chain's motion to an equilibrium configuration.

For intermediate strain rates (up to 10^3/s) it is sufficient to assume incompressibility of the three-dimensional chain cloud and Gaussian statistics for the thermal fluctuations. The viscoelastic response of the chain is then dominated by the viscous drag of chain nodes through the entangled environment of numerous other chains and by the constraint of finite extensibility.

For higher strain rates (10^3/s–10^6/s), the Liouville equation is solved using a Chapman-Enskog procedure to extract the first-order departure from Gaussian. The results display the effects of compressibility, the consequences of thermal dissipation, an inertial, wave-like contribution to signal propagation, and a more significant viscous wave signal.

Moments of the Liouville equation produce transport equations for mean density, velocity, and energy of the chain nodes. With suitable closures, the equations become sufficient to describe chain response to deformations communicated through interchain tie points. In solving these equations, our goal was to derive single-chain constitutive relations suitable for homogenization into multiple-chain response descriptions. See the article, "Network Theory for Solid Polymers at High Strain Rates" in this "Material Properties" chapter.

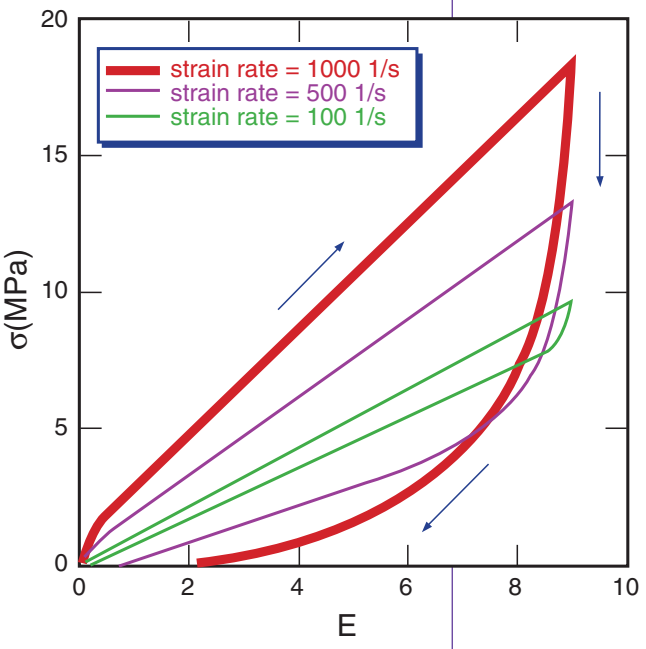


Fig. 1. Single-chain cyclic loading simulations for three strain rates.

We first considered the idealization of an azimuthally symmetric single-chain configuration with tie points on the axis, to which we prescribed velocity or force as a function of time. Because the analysis required complicated numerical solutions, we simplified even further to a "long-wavelength" one-dimensional approximation, from which useful solutions can be derived with a combination of simple numerics and analysis. At low-to-intermediate strain rates, the formulation is like that of a spring in a viscous fluid, even including the effects of finite extensibility (the spring pulled out to a nearly straight configuration). With the use of parameters relevant to the actual polymer-chain circumstances, the very simple model exhibits behavior remarkably like that of the bulk viscoelastic polymers. For example, Fig. 1 shows the expected hysteresis of cyclic loading for several loading rates (with arrows showing the direction of the loading path).

The next phase of the investigation was to find a constitutive model that represents the single-chain behavior. Our model for representation of the constitutive behavior contains a nonequilibrium descriptor variable related to the nonuniform deformation of the chain. To use this stress-strain relation, it was necessary to derive an equation to describe its evolution. The goal has been to describe the dominant polymer behavior through an entire cycle of macroscopic strain evolution that includes various periods of load, hold, and unload—all at strain rates with appreciable viscoelastic nonequilibrium.

The results of this work have been to furnish for the first time an analysis technique applicable to the constitutive behavior of polymers subjected to large strains at large-strain rates.

Equation-of-State Research and the SESAME Library

by Bradford E. Clements, James D. Johnson, and Duane C. Wallace

This brief article contains discussions on some notable highlights in the Los Alamos equation-of-state (EOS) program, with focus placed on the early days of the program. There is also a discussion on the renowned SESAME EOS and Material Properties Library. SESAME contains data for several hundred materials, including simple elements, compounds, metals, ceramics, minerals, polymers, and composites. The EOSs for existing materials in the library are upgraded when appropriate, and new materials are added as needed.

Many researchers, both inside and outside the Laboratory, use the library. Today, approximately 50 external requests are granted per year for use of the openly distributed version of SESAME. This includes requests coming from DOE and DOD laboratories, defense contractors, the National Aeronautics and Space Administration, aerospace industries, and universities and laboratories throughout the world.

Ashkin
Barnes
Clements
R. Cowan
Feynman
J. D. Johnson
Kerley
Liberman
Metropolis
More
E. Teller
Wallace
K. Warren
Wills
Young
Zimmerman

Introduction

An equation of state (EOS) describes the response of a material to applied mechanical and thermal stresses, using thermodynamic variables such as density, temperature, pressure, and energy. At Los Alamos National Laboratory, EOSs are used in modeling physical processes occurring in many Department of Energy (DOE) and Department of Defense (DOD) programs. Their primary application is in work supporting nuclear weapons research, although they have and continue to play a significant role in other programs, such as laser fusion and nuclear reactor safety. They are also of great importance in fields of basic research such as materials science, astrophysics, geophysics, and planetary science.

SESAME EOS tables usually cover a much wider range of densities, pressures, and temperatures (from ambient to astrophysical conditions) than can be studied by experimental methods. To construct a table, it is necessary to employ theoretical models of solids, liquids, vapors, and plasmas, for mixtures and chemical compounds, as well as pure elements. While there is no universal method for constructing an EOS that is valid for all densities, temperatures, pressures, and chemical compositions, generally the effort required involves compilation and analysis of experimental data, application of theoretical methods and interpolations, the accumulated experience of our researchers, and evaluation of the results. The aim of the SESAME library is to have thermodynamically consistent equations of state that are made with the best possible

In Los Alamos we met with the added feature of isolation from the world as enforced by the fence, the Army, mail censorship and so on. . . Even the inhabitants of Santa Fe, where we went of frequent shopping expeditions, knew only that all our shopping was delivered in town, at the office of the United States Army Corps of Engineers, although we lived up on the hill and maybe on the hill there was a camp for pregnant WACs [Women's Army Corps personnel] or a submarine base. (I'm sure that the idea of the submarine base must have come from Edward Teller, who is probably the Hungarian-born scientist with the greatest imagination—Los Alamos is a terribly arid place.

—Laura Fermi [wife of Nobel Laureate, Enrico Fermi and former T-Division scientist], "The Fermi's Path to Los Alamos," in Reminiscences of Los Alamos: 1943–1945, p. 92.

physics and produce the best possible agreement with available experimental data.

Historical Background

The earliest relevant theory developed at Los Alamos, that ultimately led to the present SESAME EOS library, was produced by Richard Feynman, Nick Metropolis, and Edward Teller. [1] In their article "Equations of State of Elements Based on the Generalized Fermi-Thomas Theory," the authors formulated a finite-temperature version of Thomas-Fermi theory for the electronic contribution of the EOS. In doing so, the foundation was laid for generating EOS in extreme regimes. In 1957, Robert Cowan, from Los Alamos, and Julius Ashkin, from the Carnegie Institute of Technology, formulated a theory that extended the Feynman, Metropolis, and Teller theory to include quantum mechanical exchange effects at finite temperatures, thus obtaining a finite-temperature Thomas-Fermi-Dirac (TFD) theory for the electronic contribution to the equation of state. [2] In this work they also determined practical numerical methods for solving TFD equations. Cowan further produced a self-contained form for the ionic EOS contribution in all regimes, including a fluid EOS model. These models have been described in More, Warren, Young, and Zimmerman. [3, 4]

The First EOS Database

The first extensive Los Alamos EOS database came largely from this work by Cowan. The next important development occurred in 1971, when under the direction of Jack Barnes, the SESAME EOS library came into existence. Because Los Alamos National Laboratory (then called Los Alamos Scientific Laboratory) applications required large ranges of densities, pressures, and temperatures, and because of the very diverse kinds of materials and phases encountered, Barnes also used a standardized tabular format for the EOS database. The tabular-formatted equation of state has several clear advantages over other formats: it can represent phase transitions accurately, it can span the necessary wide range of thermodynamic conditions, it is rather easily updated to incorporate new experimental or theoretical results in specific regions of temperature and density with no effort required on the part of the SESAME user, and it allows for complex, numerically intensive models to be used in the EOS.

Two main changes were made on Cowan's early database. First, a tighter standard was established for the various EOSs, and their tables, that would reside in SESAME. For example, the fineness and improved ranges of the thermodynamic points stored in SESAME were specified. Second, based on a 1967 publication, [4] Barnes added an empirical correction used to force the zero-pressure density and bulk modulus to equal the experimental values. The result was an improved EOS at lower temperatures. The essence of the Barnes' theory was to use a modified-Morse model to generate the pressures on the cold curve

(basically zero temperature isotherm), at low densities, and match the parameters in this model so that it agreed with TFD used at higher densities. For isotherms below 10,000 K, the thermal excitation of the electrons was ignored, and it was assumed that the total pressure was given by the sum of the cold curve contribution plus a lattice contribution. The thermal ionic contribution was calculated with a Debye model. Parenthetically, shockwave experiments then and now provide one of the primary means for generating data for SESAME EOS construction. The data from shockwave measurements yield the locus of states that can be reached in a single shock process from a given initial state. This locus of states is called the Hugoniot. The EOS was then compared with experimental Hugoniot data and adjustments were made as needed. For expansion regimes, a virial-like EOS was used. Thermodynamic Maxwell constructions were used to replace the van der Waals loops with the equilibrium values that occur in liquid-gas phase transitions, unless a tension region was desired.

Current EOS Program

Following these developments, responsibility for developing new EOSs shifted to Barnes and Gerry Kerley, the project leader for SESAME. In fact, the EOS format developed by Barnes and Kerley is the one still used today. A major effort was made under the direction of Kerley to improve the physics models used to generate various pieces of the EOS. One example of improved modeling was the CRIS fluid model. [5] Kerley also developed the present well-behaved interpolation for EOS surfaces.

Continuing on, other researchers began to implement more theories and models into the EOS program. These theories, which relied heavily on the previous approach of Barnes, Cowan, and Kerley, produced most of the EOSs in the present SESAME library. They assume that the pressures and energies are the sum of three separate contributions coming from the cold-curve, nuclear thermal, and electronic thermal contributions. The cold curve model consists of diamond anvil data, where available, with extrapolation to TFD theory at high densities, or by using the density dependence of the Gruneisen coefficient and Hugoniot data. Combinations of electronic structure theory and shock data can also be used to determine the cold curve. The interested reader is referred to J. Wills' contribution in this chapter on advances in electronic structure, especially as they pertain to actinide research.

Besides using TFD theory to obtain electronic thermal contributions, some EOSs have been produced using the INFERNO model of David Liberman, [6] which was based on an atom embedded in jellium at finite temperature. Both of these models are most appropriate for metallic elements, but they are also used for nonmetallic elements,

compounds, and mixtures. The current nuclear ionic EOS takes account of solid lattice vibrations, translational nuclear motion for fluids, and, in a crude sense, intramolecular vibration and rotation. The solid lattice vibrational terms can be calculated using Debye or Einstein theories. Liquid-gas phase transitions follow naturally from this modeling. Solid-solid phase transitions are handled crudely through structures added by hand to the cold curve. The location of melting is determined by a Lindemann's criterion and by experimentation, with the heat of fusion being obtained by either rule-of-thumb analysis or experimental data. Accurate interpolation is then done to the ideal gas at high temperatures and low densities.

This present procedure gives fairly reliable EOSs over a broad range of temperatures and densities. Other powerful methods, models, and techniques for EOS construction are under current development. One of these developments being pursued is based on replacing the current nuclear model and interpolation region with lattice and liquid dynamics free energies for each crystal phase and the liquid phase. [7] For each crystal phase and the liquid phase: the static structural potential (which gives the cold curve), the vibrational frequency distribution (which gives nuclear thermal free energy) and the electronic density of states (which gives electronic thermal free energy) is needed.

For most elements in the periodic table, these quantities can be calculated accurately from *ab initio* electronic-structure theory. The term *"ab initio* electronic structure" should not be construed simply to mean local density approximation; rather here it implies the proper inclusion of electron correlations. Liquid dynamics theory currently provides equation-of-state properties for liquid metals and is still under development for higher temperatures. TFD is still used at high temperature and pressure.

Density functional theory (DFT) is invoked to calculate the electronic ground-state energy for all important crystal phases, at all compressions of interest. From DFT the stable crystal structures are determined, as a function of compression, and give the cold curve. To date, this can be done accurately and reliably for most elements in the periodic table with the exception of some of the heavier actinides. For the stable crystal structure, at each compression, DFT is again used to calculate frequencies of selected phonons in the Brillouin zone, and the density of excited electronic states. This can be done reliably and accurately for all the elements for which the electronic ground-state energy can be calculated. A model interatomic potential is then fit to the phonon frequencies using the techniques of lattice dynamics, and the nuclear motion-free energy is calculated up to melting, at any compression, from molecular dynamics. From the electronic density of states, the electronic-excitation free energy can be calculated reliably and accurately, again to the melting temperature at any compression.

The EOS Library Today

The SESAME library is maintained by the Theoretical Division. In this brief summary, we have made no attempt to include the names of all those researchers that have contributed to the SESAME library and the EOS program since its inception, as the list would be very long. The interested reader can consult the chapter references for this information. [8-11] SESAME contains data for several hundred materials, including simple elements, compounds, metals, ceramics, minerals, polymers, and composites. The equations of state for existing materials in the library are upgraded when appropriate, and new materials are added as needed. Most equations of state in the library have been generated at the Laboratory although SESAME does include many that were originally constructed by researchers at other institutions (such as Lawrence Livermore National Laboratory and the National Institute of Standards and Technology) and subsequently tabulated for the library. The thermodynamic data stored in the Library include tables of pressure and energy, and in many cases the Helmholtz free energy, each as a function of density and temperature. Besides these, there are also tables of electronic and ionic contributions. Some materials have vaporization, melt, and shear tables. The typical compressions and temperature ranges are from 10^{-6} to 10^{4} and 0 to 10^{5} eV, respectively. Material properties such as opacities, conductivities, and mean ion charge are also included.

The library is used by many researchers, both inside and outside the Laboratory. The practice of distributing the SESAME library outside the Laboratory appears to have begun in 1972 when the SESAME EOS for deuterium, developed by Kerley, was sent outside the Laboratory for use in laser fusion research. Today, approximately 50 external requests are granted per year for use of the openly distributed version of SESAME. This includes requests coming from DOE and DOD laboratories, defense contractors, the National Aeronautics and Space Administration, aerospace industries, and universities and laboratories throughout the world.

Void Growth and Fracture in Ductile Solids

Addessio
Carroll
Holt
J. N. Johnson

by James N. Johnson

Models of ductile fracture are important in the design of a broad range of military hardware to withstand material separation or to know with some precision where and when it is apt to occur. This includes nuclear weapons design, armor applications, and submarine vulnerability, among a host of related applications.

From this brief description and associated references, a researcher can follow the development of ductile void-growth models leading to what is now known as "TEPLA," standing for tensile plasticity. These models and their offspring provide a direct, unified approach to modeling spallation, necking, and fragmentation in ductile materials.

Elasticity is the property of a deformed solid to return to its original shape once the deforming stress is removed. Plasticity is the property to retain some of the deformation permanently. Rate-dependent plasticity is a description of permanent deformation that depends on how fast that deformation occurs. There are myriad elastic-plastic models of high-rate deformation for intact, fully dense materials. Ductile fracture encompasses the partial or complete separation of material when subjected to large tensile or pulling stresses resulting in localized plastic deformation, usually in the vicinity of small, growing voids.

Fig. 1. Micrograph of fracture in copper. [1] The jagged region in the center of the photo is complete material separation resulting from the nucleation, growth, and coalescence of individual voids (here seen on either side of the fracture). In copper, coalescence occurs when the distention α reaches a value of approximately 1.5.

On the mesoscale (between the atomic scale and the continuum), the "elementary particle" of plasticity is the location, an atomic-scale defect present in all but the most carefully prepared crystalline materials. The corresponding element in the theory of ductile fracture in tension is the void. The micrograph of Fig. 1 shows a region of ductile fracture in copper subjected to high-rate tensile loading. [1] The dark jagged image in the center of the micrograph is a region of complete material separation. This took place by the formation of individual voids, developed in the early stages of the fracture process, followed by their growth and, eventually, coalescence to form a fracture surface. Individual voids can be seen above and below the fully fractured region.

For solids in compression, void volumes tend to decrease, if the compressive pressure is large enough. Carroll and Holt [2] treated the problem of void collapse in porous materials in detail with a clever analytical approach. The same mathematical model can be applied to that of void growth simply by changing the sign of the pressure surrounding the prototypical void in the Carroll-Holt development. This was something I thought I could handle and the result is given in [3].

The additional effect of rate-dependent plasticity is added to the Carroll-Holt formulation in the void-growth model. The growth of

voids is affected by the density of the solid (inertial effects) and the yield strength. Inertial effects simply refer to forces involved in accelerating a mass to some velocity, such as pushing a stalled automobile; the more fully loaded the automobile (i.e., greater density), the more force required. Yield strength is the stress necessary to produce permanent plastic deformation in a material. The rate of void growth is expressed in terms of the time rate of change of the distention. The distention α is defined as the ratio of the volume of material containing voids to that of the solid volume (no voids) at the same pressure and is a number greater than unity.

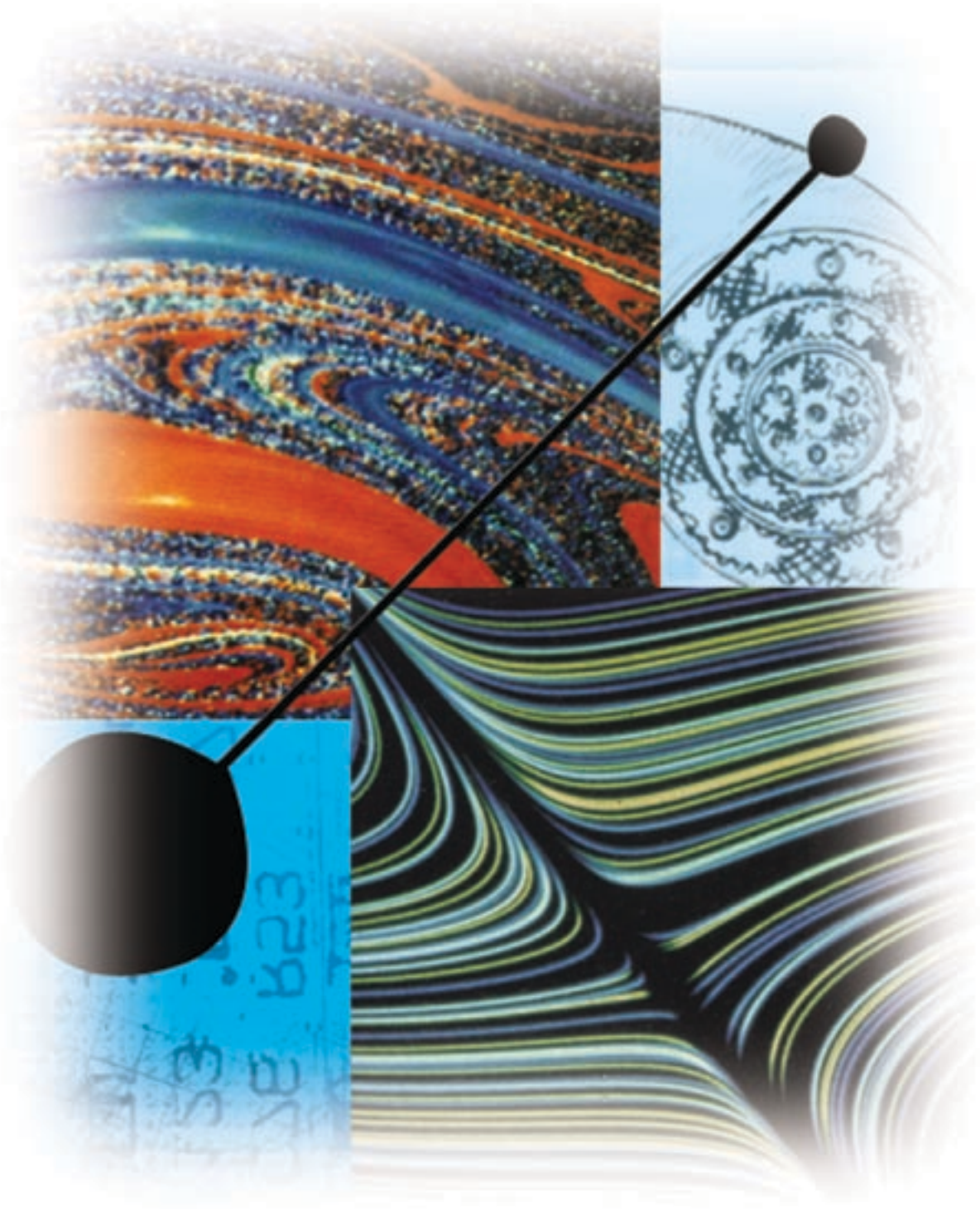
Computational descriptions of these processes require some form of mathematical rate law for their ultimate use. The end result of what might be called the "standard void-growth model" is summarized by the following equation for the time rate of change of a (with inertial effects ignored):

$$\dot{\alpha} = -\frac{(\alpha_0 - 1)^{2/3}}{\eta}\alpha(\alpha - 1)^{1/3}\Delta p, \quad \text{for} \quad \Delta p < 0$$

$$\text{where} \quad \Delta p \equiv \bar{p} + (2Y/3\alpha)\ln\left(\frac{\alpha}{\alpha - 1}\right) \quad .$$

Here a zero subscript indicates an initial value, $\bar{p}$ is the negative pressure, and Y is the yield strength. The quantity η represents rate-dependent plasticity of the matrix material surrounding the void and has the units of stress·time. The second term on the right-hand-side of the lower equation represents a threshold for the growth of voids in tension. Voids grow when $\Delta p < 0$ and void growth is arrested when $\Delta p \geq 0$. That is, the negative pressure must be less than some threshold value in order for voids to grow by means of plastic deformation. The fewer the voids, the more negative the threshold pressure.

The standard void-growth model has had considerable success in representing ductile failure of metals under plate-impact and explosive loading conditions as long as the absolute value of $\bar{p}$ is large compared to Y. The main example is in the case of spallation of metals. "Spallation" is defined as the planar fracture that occurs when a sharp stress pulse, within the material, impinges upon a stress-free, or near-stress-free surface. There are many important physical situations where the condition of $\bar{p}$ being large compared to Y is not satisfied and yet we need to be able to calculate the growth of voids leading to material failure. An example is the tensile fracture of a wire being pulled on each end, which involves "necking," a rapid reduction in the cross-sectional area of the wire in the region of failure. These types of problems are related to fracture that can occur in the vicinity of rib welds in submarines, for example. An exposition of how more-advanced models are developed within the framework of void growth is presented in the work of Johnson and Addessio. [4] These ideas have also been applied to "fragmentation," the fracture of a material leading to a large number of individual pieces.



The Ergodic Hypothesis: A Complicated Problem in Mathematics and Physics

by Adrian Patrascioiu

This article is excerpted with permission from *Los Alamos Science, Special Issue: Stanislaw Ulam 1909–1984* (Los Alamos National Laboratory, 1987), pp. 263–265.

There are a few problems in physics that stir deep emotions every time they are discussed. Since physicists are not generally speaking an emotional group of people, the existence of these sensitive issues must be considered a strong indication that something is amiss. One such issue is the interpretation of quantum mechanics. I will take a moment to discuss that problem because it bears directly on the main topic of this article.

In quantum mechanics, if the question asked is a technical one, say how to compute the energy spectrum of a given atom or molecule, there is universal agreement among physicists even though the problem may be analytically intractable. If on the other hand the question asked pertains to the theory of measurement in quantum mechanics, that is, the interpretation of certain experimental observations performed on a microscopic system, it is virtually impossible to find two physicists who agree. What is even more interesting is that usually these controversies are void of any physical predictions and are entirely of an epistemological character. They reflect our difficulty in bridging the gap between the quantum mechanical treatment of the microscopic system being observed and the classical treatment of the macroscopic apparatus with which the measurement is performed. It is usually argued that we physicists have difficulty comprehending the formalism of quantum mechanics because our intuition is macroscopic, hence classical, in nature. Now if that were the case, we should have as much difficulty with special relativity, since we are hardly used to speeds comparable to that of light. Yet, strange as it seems at first, I have never heard physicists argue about the "twin paradox," the classic example of an unexpected prediction of Einstein's relativity. So there must be something about quantum mechanics that "rubs us" the wrong way. The question is what?

Perhaps the best way in which the strange predictions of quantum mechanics can be quantified is a certain inequality first formulated by Bell. [1] For illustration, consider a positronium atom, with total angular momentum zero, that decays into an electron and a positron. Suppose we let the electron and the positron drift apart and then measure their spin components along two axes by passing them through two magnetic fields. Now in quantum mechanics the state of the positronium

atom is a linear superposition of spin-up and spin-down states: $(|\uparrow\rangle_+|\downarrow\rangle_-|\downarrow\rangle_+|\uparrow\rangle_-)/\sqrt{2}$. We could therefore ask ourselves whether in each passage through the apparatus the electron and the positron have a well-defined spin (up or down), albeit unknown to us. Some elementary probabilistic reasoning shows immediately that if that were the case, the probabilities for observing up or down spins along given axes would have to obey Bell's inequality. The experimentally measured probabilities violate this inequality, in agreement with the predictions of quantum mechanics. So the uncertainties in quantum mechanics are not due to incomplete knowledge of some local hidden variables. What is even stranger is that in a refinement of the experiment in which the axes of the magnetic fields are changed in an apparently random fashion, [2] the violation of Bell's inequality persists, indicating correlations between space-like events (that is, events that could be causally connected only by signals traveling faster than the speed of light). While in this experiment no information is being transmitted by such superluminal signals, and hence no conflict with special relativity exists, the implication of space-like correlations hardly alleviates the physicist's uneasiness about the correct interpretation of quantum mechanics. Of course this uneasiness is not felt by all physicists. Particle physicists, for instance, take the validity of quantum mechanics for granted. To wit, anybody who reads *Time* knows that they, having "successfully" unified weak, electromagnetic, and strong interactions within the framework of quantum field theory, are presently subduing the last obstacle, quantizing gravity by unifying all interactions into a quantum field theory of strings. And they are doing so in spite of the fact that the existence of classical gravitational radiation, let alone that of the quantized version (gravitons), has not been established experimentally.

An even older controversy, which in the opinion of some physicists has long ceased to be an interesting problem, concerns the ergodic hypothesis, the subject of this discussion. I will try to elaborate on this topic as fully as my knowledge will allow, but, by way of introduction, let me just say that the ergodic hypothesis is an attempt to provide a dynamical basis for statistical mechanics. It states that the time-average value of an observable—which of course is determined by the dynamics—is equivalent to an ensemble average, that is, an average at one time over a large number of systems all of which have identical thermodynamic properties but are not identical on the molecular level. This hypothesis was advanced over one hundred years ago by Boltzmann and Maxwell while they laid the foundations of statistical mechanics. [3–8] The general consensus that the hypothesis, still mathematically unproven, is probably true yet irrelevant for physics. The purpose of this article is to review briefly the status of the ergodic hypothesis from mathematical and physical points of view and to argue that the hypothesis is of interest not only for statistical mechanics but for physics as a whole. Indeed the mystery of quantum mechanics itself may possibly be unraveled by a deeper understanding of the ergodic hypothesis. This last remark should come as no surprise. After all, the birth of quantum

mechanics was brought about by the well-known difficulties of classical statistical mechanics in explaining the specific heats of diatomic gases and the blackbody radiation law.

The FPU Problem
Excerpts from "Studies of Nonlinear Problems" by Fermi, Pasta, and Ulam

Fermi
Pasta
Tsingou
Ulam

This article is reprinted with permission from *Los Alamos Science*, *Special Issue: Stanislaw Ulam 1909–1984* (Los Alamos National Laboratory, 1987), pp. 272–273.

This report is intended to be the first one of a series dealing with the behavior of certain nonlinear physical systems where the nonlinearity is introduced as a perturbation to a primarily linear problem. The behavior of the systems is to be studied for times which are long compared to the characteristic periods of the corresponding linear problems.

The problems in question do not seem to admit of analytic solutions in closed form, and heuristic work was performed numerically on a fast electronic computing machine (MANIAC I at Los Alamos).* The ergodic behavior of such systems was studied with the primary aim of establishing, experimentally, the rate of approach to the equipartition of energy among the various degrees of freedom of the system. Several problems will be considered in order of increasing complexity. This paper is devoted to the first one only.

We imagine a one-dimensional continuum with the ends kept fixed and with forces acting on the elements of this string. In addition to the usual linear term expressing the dependence of the force on the displacement of the element, this force contains higher order terms. For the purposes of numerical work this continuum is replaced by a finite number of points (at most 64 in our actual computation) so that the partial differential equation defining the motion of this string is replaced by a finite number of total differential equations. ...

The solution to the corresponding linear problem is a periodic vibration of the string. If the initial position of the string is, say, a single sine wave, the string will oscillate in this mode indefinitely. Starting with the string in a simple configuration, for example in the first mode (or in other problems, starting with a combination of a few low modes), the purpose of our computations was to see how, due to nonlinear forces perturbing the periodic linear solution, the string would assume more and more complicated shapes, and, for t tending to infinity, would get into states where all the Fourier modes acquire increasing importance.

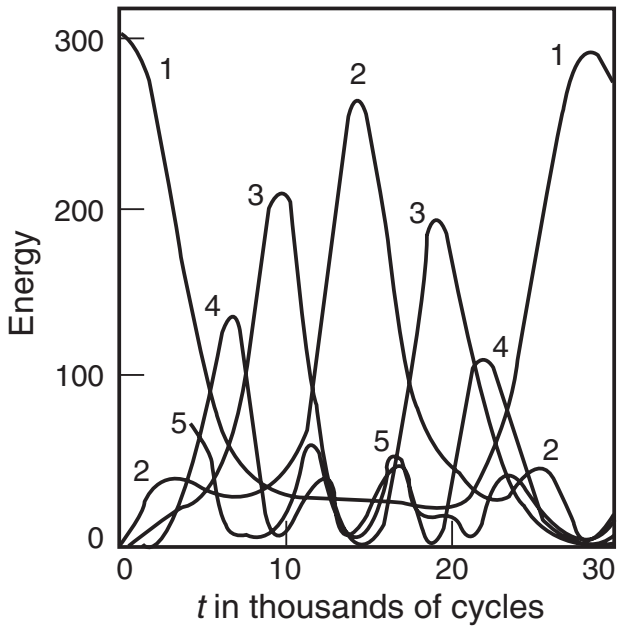
*We thank Miss Mary Tsingou for efficient coding of the problems and for running the computations on the Los Aamos MANIAC machine.

In order to see this, the shape of the string, that is to say ... [its displacement,] and the kinetic energy ... were analyzed periodically in Fourier series...

Let us say here that the results of our computations show features which were, from the beginning, surprising to us. Instead of a gradual, continuous flow of energy from the first mode to the higher modes, all of the problems show an entirely different behavior. Starting in one problem with a quadratic force and a pure sine wave as the initial position of the string, we indeed observe initially [see figures] a gradual increase of energy in the higher modes as predicted (e.g., by Rayleigh in an infinitesimal analysis). Mode 2 starts increasing first, followed by mode 3, and so on. Later on, however, this gradual sharing of energy among successive modes ceases. Instead, it is one or the other mode that predominates. For example, mode 2 decides, as it were, to increase rather rapidly at the cost of all other modes and becomes predominant. At one time, it has more energy than all the others put together! Then mode 3 undertakes this role. It is only the first few modes which exchange energy among themselves and they do this in a rather regular fashion. Finally, at a later time mode 1 comes back to within one per cent of its initial value so that the system seems to be almost periodic. All our problems have at least this one feature in common. Instead of gradual increase of all the higher modes, the energy is exchanged, essentially, among only among a certain few. It is, therefore, very hard to observe the rate of "thermalization" or mixing in our problem, and this was the initial purpose of the calculation.

FIGURES FROM THE FERMI-PASTA-ULAM PAPER

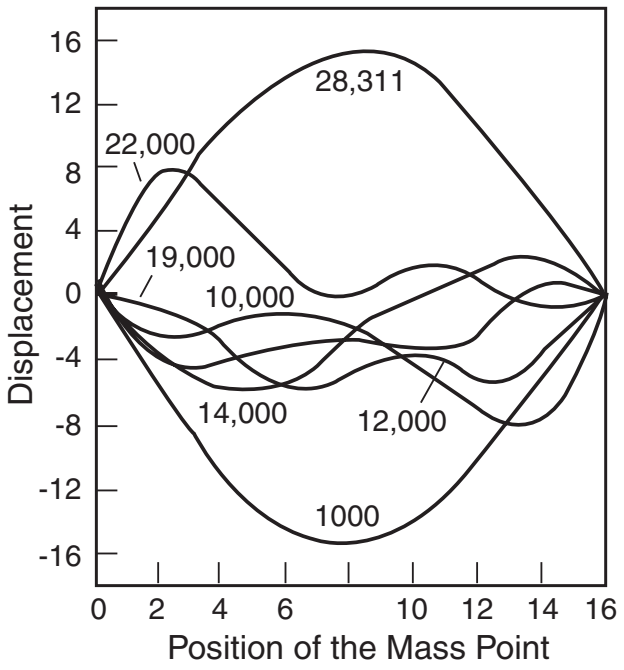
Fig. 1. The quantity plotted is the energy (kinetic and potential in each of the first five modes). The units for energy are arbitrary. *N* (the number of parts into which the string is divided) = 32; α (the coefficient of the quadratic term in the force equation) =1/4; $\delta t^2 = 1/8$ (δt is the length of the computational cycle). The initial form of the string was a single sine wave. The higher modes never exceeded in energy 20 of our units. About 30,000 computation cycles were calculated.



If one should look at the problem from the point of statistical mechanics, the situation could be described as follows: the phase space of a point representing our entire system has a great number of dimensions. Only a very small part of its volume is represented by the regions where only one or a few out of all possible Fourier modes have divided among themselves almost all the available energy. If our system with nonlinear forces acting between the neighboring points should serve as a good example of a transformation of the phase space which is ergodic or metrically transitive, then the trajectory of almost every point should be everywhere dense in the whole phase space. With overwhelming probability this should also be true of the point which at time $t = 0$ represents our initial configuration, and this point should spend most of its time in regions corresponding to the equipartition of every among various degrees of freedom. As will be seen from the results, this seems hardly the case. ...

In a linear problem the tendency of the system to approach a fixed "state" amounts, mathematically, to convergence of iterates of a transformation in accordance with an algebraic theorem due to Frobenius and Perron. ... Such behavior is in a sense diametrically opposite to an ergodic motion and is due to a very special character, linearity of the transformations of the phase space. The results of our calculation on the nonlinear vibrating string suggest that in the case of transformations which are approximately linear, differing from linear ones by terms which are very simple in the algebraic sense (quadratic or cubic in our case), something analogous to the convergence to eigenstates may obtain statistical relevance of time averaging for macroscopic observables.

Editor's note: The interpretation of the unexpected recurrences is now different. See David Campbell's discussion in this "Mathematics Science"chapter.



FIGURES FROM THE FERMI-PASTA-ULAM PAPER
Fig. 8. This drawing shows not the energy but the actual *shapes*, i.e., the displacement of the string at various times (in cycles) indicated on each curve. The problem is that of Fig.1.

Neural Networks

Lapedes

by Alan S. Lapedes

Artificial neural networks are abstract computer models designed to capture certain aspects of real biological brains, and they are used to implement novel ways of computing. We have investigated the mathematical structure of artificial neural networks, in order to understand them as a pure mathematical system without fuzzy anthropomorphic appeals to "thinking." We have shown that artificial neural networks are not limited to fuzzy or qualitative pattern-recognition tasks. Also, we have demonstrated in head-to-head competition with the most accurate available mathematical techniques that neural networks are remarkably accurate predictors of complicated chaotic time series. Artificial neural networks stand on their own as flexible and accurate mathematical approximation techniques, the understanding of which has given rise to additional new techniques (e.g., radial basis functions) that have additional advantages over the neural network formalisms of the 1980s.

Research in artificial neural networks underwent a resurgence of interest in the 1980s that continues to this day. Artificial neural networks are computer models (also sometimes built in hardware, but usually simulated in software) that are very loosely based on how real biological neurons, in e.g., the human brain, work. They are not necessarily intended to be realistic models of biological neurons but are designed to capture certain aspects of real brains and to use these aspects to create novel ways of computing. Recognition of patterns, something that real brains do well and traditionally artificial "brains" (computers) do poorly, is a problem that artificial neural networks have made great advances in addressing.

Perhaps because artificial neural networks do capture some aspects of real brains (although neuro-biologists might possibly prefer to replace the word "some" with "a very very few"), there has been a mystique about them and the principles underlying how they work. "Learning from examples" is a phrase that has often been associated with artificial neural networks: give them some examples relating an "input" to an "output" and they "learn" from these examples and "generalize" to new situations i.e., give a network a new input after it has been "trained" on a set of examples and it will give a reasonable and appropriate output. There has also been a prejudice that the use of artificial neural networks is limited to attempting to get computers to act like "real" brains, i.e., that artificial neural networks are best used for "fuzzy" or qualitative tasks such as pattern recognition and not for quantitative computation where obtaining high numerical accuracy is important.

However, we were intrigued by the simple mathematical structure of artificial neural networks, and felt that they could be understood in detail as a pure mathematical system, without fuzzy anthropomorphic appeals to "thinking." In fact, we also found that artificial neural networks can be remarkably precise, and can be used for tasks such as predicting the future of a time series (a series of numbers, such as the daily returns of the stock market) with remarkable accuracy. Now the stock market of course is notoriously hard to predict, and many people think it has a huge totally random component. However, there are other systems in nature that exhibit very complex, essentially random-looking behavior, but which have simple underlying rules. These systems, called chaotic attractors, are often described by nonlinear differential equations. It is such chaotic systems (if the stock market could be predicted with neural networks, we wouldn't be explaining how to do it here) that we found could be described remarkably well using neural networks. We also found that the methods neural networks use to accomplish such tasks are eminently understandable in terms of simple notions of fitting curves or surfaces to run through a given set of points (although in high dimensions, so visualization of this principle can be tricky for those people not used to thinking about dimensions other than three). Phrases such as "learning from example" only tend to mystify and obscure what is in essence a simple, but new and very useful, way to accomplish quantitative tasks.

Feedforward Design

One of the most popular neural network architectures is the so-called feedforward design, in which an initial layer of "input" neurons feeds into (possibly multiple) layers of "hidden" or processing neurons, and these hidden neurons in turn feed into an "output" layer. When the neurons on the input layer take on certain values, these values are fed through the hidden or processing layer(s), and ultimately a new value (the "answer," or output) appears as the values of neurons on the output layer.

By changing the way neurons are connected to each other between these layers, i.e., by changing what are called "synaptic weights," one can change the value of the output for any given input. So, if one has a "training set" of input/output examples, then it is possible to "train" the network (adjust the synaptic weights, using an algorithm called "back propagation") so that if an example "input" from the training set is presented on the input layer then the associated "output" is correctly reported on the output layer. After training one can insert new inputs on the input layer and see if the output appearing on the output layer is correct, i.e., if the network "learned from example" and "generalized" to new situations or inputs. But this is a needlessly anthropomorphic way to think about a system that merely transforms "input" to "output."

A simple function such as a straight line, y = a + b*x, where a and b are constants and x is a variable, also transforms an input, "x", to an output "y", depending on "synaptic weights" a and b. By adjusting a and b in ways quite familiar to mathematicians, we can make this function, y = f(x), map a set of x values to an associated set of y values with the least error possible. Although neural networks are nonlinear (the above example is linear), and also operate with high dimensional inputs (not just one "x"), the principle is the same. Therefore we decided to try neural networks on a very hard input/output mapping problem: attempt to accurately predict the future of a nonlinear, chaotic time series by "training" on a set of past values, to be used as input, of the time series. The value of the time series in the future is the output. Typically, if one looks at such chaotic time series, then there are features of the time series that seem to be changing in an almost random and unpredictable way (to the human eye, at least). We found that neural networks could predict such chaotic time series remarkably well. The reason why they can is not that they "think" about the time series better than human brains do, but that they are seen to be, when analyzed properly, very flexible and accurate input/output estimators. In fact they are "universal" function approximators: given enough hidden or processing neurons they can approximate almost any function arbitrarily well. [1,2]

In summary, artificial neural networks are not limited to fuzzy or qualitative pattern recognition tasks, and in head to head competitions with the most accurate available mathematical techniques have proven to be remarkably accurate predictors of complicated chaotic time series. Although they were originally motivated by real biological brains, we found that they can be understood in terms of quite familiar mathematical principles. Understanding in detail how they work, however, does not detract from an appreciation of their power. They stand on their own as a remarkably flexible and accurate mathematical technique, the understanding of which has given rise to additional new mathematical techniques (e.g., radial basis functions) that have additional advantages over the neural network formalisms of the 1980s.

Mathematics at Los Alamos

by J. Mac Hyman, William A. Beyer, and George A. Baker, Jr.

To show how both pure and applied mathematics emerged from Los Alamos, we have organized it by years in this article. A selection is made of a few papers, out of many, to exhibit some of the mathematical work of Theoretical Division.

Introduction

The Manhattan project began with the best scientists available. So too, mathematics at Los Alamos began with two world-class mathematicians: John von Neumann and Stanislaw Ulam. Both were products of Central Europe before World War I, the birthplace of many great scientists in the first half of the twentieth century. Von Neumann was born in 1903 in Budapest. Ulam was born in 1909 in Lwów, Poland. At that time, both cities were part of the Austro-Hungarian Empire. Both men had a classical education. For example, Ulam had eight years of Latin and four years of Greek. Ulam and von Neumann in later years sometimes corresponded in Latin.

Both men made important contributions to the foundations of mathematics. Von Neumann developed new axioms for mathematics. Ulam searched the world of infinite numbers and found what he called inaccessible cardinal numbers, numbers so large that they could not be reached by ordinary mathematics.

Von Neumann was in the U.S. at the beginning of World War II. Ulam barely made it out of Europe and was on the high seas when the war started.

With war starting in 1939, it became clear that the atomic bomb was a possibility. The best brains were assembled. Einstein was left out because it was thought he could not keep a secret. Von Neumann was not recruited at first because it was thought that bomb work was more physics than mathematics. Little did the organizers foresee the contributions von Neumann would make to both the atomic bomb and the hydrogen bomb.

Early in the war Ulam was teaching mathematics at the University of Wisconsin. He wanted a war job and thought von Neumann could help. Von Neumann was already working at the secret city of Los Alamos. He asked Ulam to meet him at the Union Station in Chicago to discuss job possibilities. Von Neumann told Ulam there was a chance for Ulam to do some very secret work. And so Ulam ended up at Los Alamos during the war. It is clear that these two great mathematicians started Los Alamos on its way to doing important mathematics.

In a list of members of T-Division in May 1945 are names of men who later became outstanding mathematicians. Among them are Peter Lax, Paul Stein, Milton Wing, John Kemeny, and Richard Bellman.

Much mathematics, both pure and applied, emerged from Los Alamos. It is difficult to organize it in a rational way. For the lack of a better way, we organize it by years. A selection is made of a few papers, out of many, to exhibit some of the mathematical work of T-Division.

Papers by Ulam can be found in the book, *Stanislaw Ulam: Sets, Numbers, and Universes; Selected Works* (Massachusetts Institute of Technology Press, Cambridge, Massachusetts, 1974).

1944

S. M. Ulam and his collaborators, D. Hawkins and C. J. Everett, in the years 1944 to 1948, published five papers on what they called "multiplicative processes." Such processes are now famous and are called "branching processes." This subject began in 1873 with work by the Rev. Henry Watson on the problem of survival of family names as a branching process. The final solution to Watson's problem was obtained in 1948 by Everett and Ulam. It is probable that Everett, Hawkins, and Ulam did not know of previous work by Watson and others. The work of Ulam et al. was motivated by a system of neutrons or other particles undergoing birth, transmutation to other particles, and death. Such systems occur in nuclear reactors or explosions. These papers by Ulam et al. are seminal and important contributions to the theory of branching processes.

1945

Work on ordered groups produced a paper published by Everett and Ulam in 1945. An ordered group is a lattice on which there is a group operation defined and there is a certain connection between the lattice and the group. Everett and Ulam obtain results that were previously known only for the commutative case. They prove a conjecture of Garrett Birkhoff's that, if the commutator group of an ordered group is in its center, then the integral closure of the group implies commutativity.

1946

A paper published by Everett and Ulam in 1946 is important on two counts. First, it was a way-station in the development of algebraic versions of logic. Second, the theorem on completion was later extended to other algebraic forms of logic. The representation theorem of this paper is a model for representation theorems for other two-dimensional (2-D) algebraic logics.

1949

The modern Monte Carlo school seems to have originated at Los Alamos after the Second World War in work on nuclear reactions and equations of state. It seems clear that Ulam was the inventor of Monte Carlo, and he gave it its name. Other early contributors were Everett, Fermi, Metropolis, Richtmyer, Teller, and von Neumann. It was the later appearance of the stored-program electronic computer that allowed the accuracy associated with large-scale sampling experiments.

1950

Richtmyer and von Neumann developed a method for one-dimensional (1-D) inviscid compressible fluid dynamics. Two key ideas were introduced. The first was the use of a staggered Lagrangian coordinate system (grid points move with the fluid). This enabled the accurate tracking of abrupt changes in materials (discontinuities) (like sonic booms or tidal waves). The second was artificial viscosity. This enabled the accurate modeling of shock waves by spreading the discontinuity over a small distance while still retaining the basic properties of the solution. This method is the foundation for all Lagrangian codes. The explicit artificial viscosity term allowed the method to calculate shock waves in gas dynamics. It was not second-order accurate and was not conservative. Someone observed that time-centering the velocity fixes conservation. In real fluids, the discontinuities are not infinitely sharp but extend over a thin layer due to viscosity and diffusive properties. In order to reduce nonphysical oscillations at the mesh points of a numerical solution near the shock, Richtmyer and von Neumann added small amounts of artificial numerical viscosity. The goal was to keep the representation of the shock sharp, while minimizing oscillations.

1952

S. Ulam and Hyers, "Approximately Convex Functions," *Proc. Amer. Math.* Soc. **3.**

1953

N. Metropolis, A. Rosenbluth, N. Rosenbluth, A. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines," *J. Chem. Phys.* **21**, 1087.

This paper must have a record for citations, about 6,200 of them. Even the paper of Crick and Watson published the same year, and maybe the most famous scientific paper ever, has only 1,200 citations. This paper of Metropolis et al. is famous because it deals with the use of Monte Carlo in solving a continuum problem in chemistry by means of the new electronic computers, something very new at the time.

1954

S. Ulam and Hyers, "Stability of Differential Expressions," *Math. Mag*. **28.**

1955

E. Fermi, J. Pasta, and S. Ulam, "Studies of Nonlinear Problems," Los Alamos Scientific Laboratory report LA-1940.

This important Los Alamos report, LA-1940, marks the beginning of understanding of solitary waves, now called solitons. The report first appeared with the words "Work done by Mary Tsingou," now Menzel, who still lives in Los Alamos. Fermi, Pasta, and Ulam made a nonlinear perturbation of a linear problem. They expected to see the energy become asymptotically divided among the various degrees of freedom of the system. Instead, the solution kept returning to near its original state. An analysis of this made in 1965 by Zabusky and Kruskal led to the modern understanding of solitons.

A discussion of this paper appears in an editorial by Cornell University mathematician, Steven Strogatz, in the *New York Times* for March 4, 2003. Strogatz dates the work to 1953. The report is dated 1955. However, Ulam in *Stanislaw Ulam: Sets, Numbers, and Universes; Selected Works,* page 490, states that, "We planned the work in the summer of 1952 and performed the calculations the following summer."

As Strogatz says,

> Fermi and his colleagues made a computer model to better understand entropy, the tendency of all systems to decay to states of ever greater disorder. What the computer revealed was astonishing. When they investigated plucking a computer model of a slightly non-ideal string, starting from a single harmonic, the computer added a series of higher harmonics, replacing one with another, gradually changing the sound. Then it suddenly reversed direction, deleting harmonics in the opposite order, and then returning to almost the original harmonic.

Jerry Erpenbeck points out that the numerical results of this paper are explained by the Kolmogorov, Arnold, and Moser (KAM) theorem, with the small amplitude of the nonlinear forces relative to the linear leading to the observer recurrence. For larger amplitudes outside the KAM regime, the behavior of the solution is not particularly regular.

Thomas Beyer states that some of the graphs in this report remind him of an ocean phenomenon called "rogue waves." These are waves of extremely large heights. A meeting on this subject was held at Brest, France, in November 2000.

1956

Gardiner, Lazarus, N. Metropolis, and S. Ulam, "Integer Sequences Defined by Sieves," *Math. Mag.* **29.**

The sieve of Eratosthenes that produces the prime numbers is the earliest sieve on the integers. This paper constitutes the first detailed study of other sieve processes. The work aroused considerable important theoretical and computational work on the asymptotic density of sequences produced by sieves. Paul Stein wrote a commentary that appears in the book called *Stanislaw Ulam: Sets, Numbers, and Universes; Selected Works* by S. Ulam.

1957

J. Holladay, "Cubic Splines Interpolants on Nonuniform Grids," *Math. Tables and Aids to Comp*. **11.**

This is an early paper that defines cubic spline interpolants on irregular grids by means of their variational property of minimizing the mean-square (linearized) curvature among all curves that pass through a given set of data points.

G. M. Wing, Klaba, and Bellman, "Invariant Embedding and Neutron Multiplication," *Proc. Nat. Acad. Sci.* **43.**

1958

K. A. Brueckner and J. L. Gammel, "Properties of Nuclear Matter" and "Properties of Liquid Helium-3 at Low Temperatures," *Phys. Rev.* **109,** 1023 and 1040 (1958).

The authors combined the phenomenological nuclear potentials of Gammel, Christian, and Thaler, which were developed at Los Alamos, and a method of computation of the ground-state energy, which was a generalization and application of the multiscattering theory of K. Watson. Using the then most powerful computers available (which were naturally at Los Alamos) the authors solved the complex set of nonlinear equations to obtain the ground-state properties of infinite nuclear matter. The same methods were applied in the second paper to the many-Fermion system, liquid helium-3, with semiquantitative agreement with experiments.

1959

T. Jordan and B. Swartz, "Calculating Best (minimax) Approximations to Elementary Functions by Polynomials and by Ratios of Polynomials," presented to the local Association for Computing Machinery chapter.

For a given maximum error requirement (or, even, percent error requirement) compute the polynomial (or rational function) of minimal degree(s) that best approximates a given elementary function (such as the exponential, the logarithm, or trigonometric functions) within that error level over a given interval for the function's argument. The advent of a more precise computer (STRETCH, 14-digits) prompted the work. Swartz's algorithm was Newton-like. Jordan's, similar to an algorithm subsequently published by others, involved a sequence of weighted linear least-squares approximations whose weights changed so that convergence to the desired minimax approximation occurred.

E. D. Cashwell and C. J. Everett, "A Practical Manual on the Monte Carlo Method for Random Walk Problems," (Pergamon Press, New York, 1959).

J. Pasta and S. Ulam, "Heuristic Numerical Work in Some Problems of Hydrodynamics," *Math. Tables and other Aids to Computations* **13.**

G. Pimbley, "An Initial Value Problem for the Multivelocity Neutron Transport Equation and Slab Geometry," Los Alamos Scientific Laboratory report LAMS-2116.

This work is representative of work done at Los Alamos in the 1950s by Lehner, Pimbley, and Wing on the study of neutron behavior in reactor physics. The problem is governed by a system of integro-differential equations with time and space as independent variables. It is shown among other things that a half plane forms the continuous spectrum, and the other half contains only the point spectrum.

1960

P. Lax and B. Wendroff, "Systems of Conservation Laws," Commun. Pure Appl. Math ***13,*** 217 (1960).

Lax and Wendroff introduced the conservation form of a finite difference scheme for 1-D Eulerian fluid dynamics. The second-order method preserved the conservation laws and automatically created an artificial viscosity term that allowed the method to calculate shock waves. This form is more closely tied to the weak formulation of the partial differential equation (PDE) than the PDE itself. Due to the special form, the scheme expresses numerical conservation, which implies that discontinuities propagate correctly, at least in the sense of weak solutions. The Lax-Wendroff theorem proved in this paper states that if a consistent and conservative scheme converges boundedly, it does so to a weak solution of a conservation law. Since weak solutions are not necessarily unique, entropy conditions should be imposed. With entropy conditions imposed, the Lax-Wendroff theorem can be reformulated such that bounded convergence for a conservative and consistent scheme satisfying an entropy condition implies convergence to the entropy solution.

The Lax-Wendroff method applies to inviscid compressible fluid dynamics. A stable second-order accurate finite difference scheme in Eulerian coordinates was introduced. The major points were the notion of difference schemes in conservative form and the use of the differential equations to convert high-order time derivatives to spatial derivatives, thereby producing a more accurate explicit scheme. It was also proved that if a consistent and conservative scheme converges boundedly, it does so to a weak solution of the conservation law. This theorem was later extended by Lax to show that a numerical entropy condition would imply the limit was an entropy solution.

1961

G. A. Baker, Jr., "Application of the Padé Approximant Method to the Investigation of Some Magnetic Properties of the Ising Model," *Phys. Rev.* **124**, 768 (1961).

This paper is the first to apply the Padé approximant method to the high-temperature series expansions that had been derived for various magnetic properties, e.g., the magnetic susceptibility, in an effort to determine their behavior at the critical point. The dramatic success of this method led to a whole industry among theoretical physicists that continues to this day. The results of these endeavors produced the broad outline of critical behavior. These results formed the basis for subsequent synthesis. This paper was reprinted in 1973 as part of the "Series of Selected Papers in Physics, Critical Phenomena" by the Physical Society of Japan. The editors were R. Abe and M. Suzuki.

G. A. Baker, Jr., J. L. Gammel, and J. G. Wills, "Investigation of the Applicability of the Padé Approximant Method," *J. Math. Anal. Appl.* **2**, 405 (1961).

This paper puts forward the famous Baker-Gammel-Wills conjecture concerning the pointwise convergence of the most powerful form of Padé approximants. It has taken 40 years of effort of many excellent mathematicians to resolve this conjecture. The fruit of these endeavors has been a very significant increase in the understanding of the convergence properties of what has now become a standard tool to obtain quantitative results. In the 40^{th} year, a counter-example was finally obtained. However, this counter-example suggests many new paths of fruitful inquiry.

1962

W. A. Beyer, "Hausdorff Dimension of Level Sets of Rademacher Series," *Pac. J. Math.* **12.**

At the time this paper appeared, there were only a handful of papers on the subject of sets of fractional dimension, now called fractals. Mandelbrot discovered this paper and wrote a summary of it for his

book "The Fractal Geometry of Nature." This is the book that brought fractals to the attention of the world. Ulam said at the time that Rademacher series reminded him of the Ising model in statistical mechanics. In 2001 a visitor to CNLS, Peter Peiffer, discovered that indeed there is an exact relation between Rademacher series and the Ising model. This work is ongoing.

G. A. Baker, Jr., "Certain General, Order-Disorder Models in the Limit of Long-Range Interactions," *Phys. Rev.* **126,** 2071 (1962).

In this paper, the solution of a long-range-force Ising model originated by Mark Kac is reduced to the solution of a certain functional equation. The exact numerical solution to this equation is given. This paper was reprinted in *Mathematical Physics in One Dimension, Exactly Soluble Models of Interacting Particles,* E. H. Lieb and D. C. Mattis, Eds. (Academic Press, San Diego, California, 1966). The context of the mysterious remark of Ulam, quoted in the previous entry, that the Rademacher series reminded him of the Ising model has lately become clear. It is that he had learned from Mark Kac of this version of the Ising model, which is indeed the very one that is related to Rademacher functions.

1965

B. Wendroff, "Eigenvalue Problems for Ordinary Differential Equations," *Math. Comp.* **19**, 218 (1965); Birkhoff, deBoor, Swartz, and Wendroff, *SIAM J. Num. Anal.*, **3**, 188 (1966); deBoor and Swartz, *Math. Comp*. **35**, 679 (1980); **36**, 1, 21 (1981).

The first proofs and analytic explanations of unexpectedly accurate approximate eigenvalues using piecewise polynomial methods for such problems—initially for Rayleigh-Ritz (Galerkin in this context) and for self-adjoint problems. The first two papers also develop strict numerical upper and lower bounds on eigenvalues by combining the Rayleigh-Ritz property with estimates of appropriate errors. The final three papers treat the general case of nonself-adjoint problems. They also describe the large variety of ways available to determine a piecewise polynomial approximate solution that is particularly accurate at its "breakpoints"—the ways include mixtures of Galerkin's method, least squares, and collocation at special points in the fashion of deBoor and Swartz that appeared in 1973—both for ordinary boundary value problems and for eigenvalue problems.

1967

B. Swartz and B. Wendroff, "Finite Elements for Well-Posed Evolutionary Problems," *Math. Comp.* **23**, 37 (1969).

The approximate solution of hyperbolic, parabolic, and similar evolutionary problems (those that involve semi-bounded spatial operators), using piecewise polynomials and Galerkin's method in space with

trapezoidal differencing in time, is proposed and analyzed. Stability and convergence were proved and illustrated. The use of piecewise polynomials for the spatial approximations yields finite difference-like algorithms. But the proofs, using analogs of energy methods in finite differences, are much simpler and—rediscovered by many—now form the standard introduction to finite elements for evolutionary problems.

1968

J. Hayes, R. Kellner, and D. Kahaner, "Approximate Boundary-Integral Methods for the Solution of Laplace's Equation and for Conformal Maps," Refs: Los Alamos Scientific Laboratory report LA-4004 (September 1968); Los Alamos Scientific Laboratory report LA-4423 (April 1970); *SIAM J. Appl. Math.* **22**, 503 (1972); Conformal maps: *Math. Comp.* **26**, 327 (1972); **29**, 512 (1975).

Hayes developed this initially for the approximate solution of the Rayleigh-Taylor problem. This led to the two reports describing his computer code for the various boundary value problems for Laplace's equation. Hayes and Kellner later proved ill-conditioning was not a problem even when the integral equations associated with the given boundary conditions were necessarily of the first kind. For conformal maps one can set up equations of either the first or second kind. The latter had been used since the 1950s because of its presumed relative insensitivity to numerical error, but the former turns out to be demonstrably more efficient in extreme geometric circumstances and/or for coarse boundary discretizations (Hayes, Kahaner, and Kellner).

1971

W. A. Beyer, M. L. Stein, and S. M. Ulam, "The Notion of Complexity," Los Alamos Scientific Laboratory report LA-4822 (December 1971).

The notion of arithmetic complexity of an integer n is defined as the minimum number additions, multiplications, and exponentiations required to combine 1s to form n. The value of this complexity is calculated for n less than 1,024. It is proposed that complexity and entropy are the same. A computation is presented to support this conjecture. This conjecture was proved in 2000 by S. Galatolo.

1972

W. A. Beyer and Wells, "Lower Bound for the Connective Constant of a Self-Avoiding Walk on a Square Lattice."

1973

N. Metropolis, M. Stein, and P. Stein, "On Finite Limit Sets for Transformations on the Unit Interval."

This paper is perhaps the earliest work dealing with a broad class of iterated nonlinear, 2-to-1, one-parameter transformations of the unit interval into itself. New universal properties of these iterates are found, including harmonics and anti-harmonics. Metric properties of these patterns were later generalized in the work of Louck, Feigenbaum, Li Wang, and Beyer. No theorems are proved in this paper. Only examples are provided. Many theorems suggested by the examples were later proved by Paul Stein, Beyer, Li Wang, Louck, and Myron Stein.

1974

W. Beyer, T. Smith, M. Stein, and S. Ulam, "A Molecular Sequence Metric and Evolutionary Trees."

This paper is representative of a number of papers written by the authors in the 1970s. Later Waterman joined the Los Alamos biological sequence club. The story of how this work evolved into the national human DNA database is told in the book, *The Gene Wars,* by Robert Cook-Deegan. Much of modern biology is based on sequences. As a mathematician, Ulam had a lifelong interest in metric spaces. He proposed we should investigate the concept of distance between points consisting of finite sequences over finite alphabets. In this paper, several metrics are used, but generally the triangle inequality could not be proved. Numerical evidence indicated the inequality did hold. The protein cytochrome c is found in almost all life forms. This paper uses cytochrome c for 33 amino acid sequences ranging from wheat to man. A tree is constructed with these sequences as nodes. The sequence pairs are provided with a metric. Linear programming is used to select the tree that minimizes the sum of the distances in the tree. This tree turned out to be similar to what a biologist would select using conventional methods.

C. Hirt, A. Amsden, and J. Cook, "An Arbitrary Lagrangian-Eulerian Computing Method for All Flow Speeds," *J. Comp. Phys.* **14**, 227 (1974).

W. Beyer and Waterman, "Error Analysis in Computation of Euler's Constant."

1975

G. Bell, "Mathematical Models for the Immune System."

1976

W. Beyer, Smith, and Waterman, "Biological Sequence Metrics."

Faltin, N. Metropolis, Ross, and G.-C. Rota, "The Real Numbers as a Wreath Product."

1977

J. Mac Hyman, "Automated PDE Software Design."

1978

M. Feigenbaum, "Quantitative Universality for a Class of Nonlinear Transformations," *J. Stat. Phys.* **19,** 25 (1978).

This is a famous paper that presents a property of a one-dimensional, one-parameter set of quadratic mappings of the unit interval into itself. As the parameter $\lambda(n)$, $n = 1,2,3,...$, increases, the mappings go through a sequence of period doublings. The ratio $(\lambda(n)–\lambda(n-1)/(\lambda(n+1)–\lambda(n))$ seems to approach as n increases a limit 4.6692..., now called the Feigenbaum ratio. This behavior was called "universal" because it seems to be a property of a large class of similar mappings. Feigenbaum's observation was later proved by O. Lanford in 1982. His proof was assisted by a computer. There have been further developments of Feigenbaum's theory.

Feigenbaum Universality

A mapping f of the interval [–1,1] is called unimodal if (1) f is continuous, (2) $f(0) = 1$, and (3) f is strictly increasing on [–1,0] and strictly decreasing on [0,1]. We say f is C^1– unimodal if in addition f is once continuously differentiable and $f'(x) \neq 0$ if $x \neq 0$. We consider a map f_μ from μ to the space of C^1 unimodal maps, continuous in the C^1 topology. More precisely, this means that

$$\sup_{x\in[-1,1]} \left|f_\mu(x) - f_{\mu_0}(x)\right| + \left|f'_\mu(x) - f'_{\mu_0}(x)\right| \to 0 \quad ,$$

when $\mu \to \mu_0$.

By investigating numerically a number of one-parameter families, such as $\mu \to 1 - \mu x^2$, Feigenbaum discovered a universality property: for large j, $\mu_\infty - \mu_j$ is asymptotic to

$$const \times \delta^{-j} \quad ,$$

where $\delta = 4.66920\ldots$ is apparently the same whatever one-parameter family is considered. So this is Feigenbaum's discovery of the universality of the number δ.

Feigenbaum went on to propose an explanation of this from the renormalization group approach to critical phenomena in statistical mechanics.

Most of this material is taken from a book by Collet and Eckmann, *Iterated Maps of the Interval as Dynamical Systems*.

N. Metropolis and G.-C. Rota, "Combinatorial Structure of the Faces of the n-Cube," *SIAM J. Math. Anal.* **35** (4), 689 (1978).

This paper contains two results: (1) An algebraic characterization of the lattice of faces of the n-cube; and (2) Explicit construction of the partition of the lattice of faces into the minimal number of chains.

G. A. Baker, Jr., B. G. Nickel, and D. I. Meiron, "Critical Indices for Perturbation Analysis of the Callan-Symanzik Equation," *Phys. Rev. B* **17**, 1365 (1978).

This paper gives the full report of an earlier *Physical Review Letter*, which was the first paper to make quantitative the results of the renormalization theory for which Ken Wilson was later to win the Nobel prize in physics. In this paper, the series in the renormalized coupling constant at fixed dimension is analyzed to give accurate estimates of the various critical properties. It is the renormalization group theory that provided a synthesis of the previous decade of theoretical work.

1980

D. D. Holm, "Lie-Poisson Brackets."

1981

J. Mac Hyman, "Soliton in Polyacetylene Adaptive Mesh Methods."

Scott and J. M. Hyman, "Quasiperiodicity Renormalization Group Davidov Solitons."

L. Biedenharn and J. Louck, *Angular Momentum in Quantum Physics, Theory and Application*, Vol. 8 (Addison-Wesley, Boston).

L. Biedenharn and J. Louck, *The Racah-Wigner Algebra in Quantum Theory*, Vol. 9 (Addison-Wesley Publishing Company).

1982

D. D. Holm and B. A. Kupershmidt, "Lie-Poisson Brackets."

In *Poisson Structures for Superfluids* by D. D. Holm and B. A. Kupershmidt, *Phys. Lett. A* **91**, 425 (1982), the authors reported that they discovered the semidirect-product nature of the Lie-Poisson brackets underlying the classical Hamiltonian formulation of superfluid dynamics. These noncanonical Poisson brackets turned out to be fundamental also for the Hamiltonian formulation of any *classical* ideal Eulerian fluid motion. Later they became a useful tool for formulating new models in fluid and plasma dynamics, as well, including multicomponent flows. This Lie-Poisson framework was particularly useful in obtaining nonlinear stability conditions for fluid and plasma equilibria. See "Nonlinear Stability of Fluid and Plasma Equilibria," D. D. Holm, J. E. Marsden, T. S. Ratiu, and A. Weinstein, *Phys. Rep.* **123**, 1–116 (1985), which currently has received more than 400 citations.

J. Brackbill and J. Saltzman, "Adaptive Zoning for Singular Problems in Two Dimensions," *J. Comp. Phys*. **46**, 342 (1982).

Farmer and Cruchfield, "Fractal Dimension Algorithms."

E. Jen, "Cellular Automata."

1983

W. A. Beyer and P. Stein, "Period Doubling for Trapezoid Function Iteration."

1986

G. Pimbley, "On Using Semigroup and the Trotter Product Formula to Solve Quasilinear Systems."

J. D. Louck, *Symbolic Dynamics of Trapezoidal Maps* (D. Reidel Publishing Company)

D. Farmer, "Predicting Chaotic Time Series."

A. Perelson, "Immunology and Adaptation."

J. Mac Hyman and Nichols, "KS Equation."

1988

G. A. Baker, Jr., "Pluralism in the Critical Phenomena of the One-Dimensional, Continuous-Spin Ising Model," *Phys. Rev. Lett.* **60**, 1844.

For some time previous to this publication, there had been a number of indications that the concept of universality was overly broad. It was felt that critical problems can be divided into classes differentiated only by the dimensionality of the system and the symmetry group of the order

parameter. This was a concept that permitted the computation by field-theoretic methods of the universal critical properties of a representative model of class. As expressed by M. Barma and M. E. Fisher, "It is commonly held that the $\lambda\Phi^4$ lattice model is always in the same universality class as the pure Ising model." In this short paper, it is shown that the one-dimensional Ising- and Gaussian-model universality classes do not exhaust the universality classes of the 1-D, continuous-spin Ising model, and hence the normal universality hypothesis fails in this simple, readily analyzable model. Many other examples have been found subsequently.

J. M. Hyman, Stanley, S. Colgate, and A. Perelson in "HIV Epidemiology," use stochastic scaler advection coupled-map lattices.

N. Metropolis and G.-C. Rota, "*Symmetric Functions: A Bijective Identity*." A bijective proof of a classical identity, called the cycltomic identity, is given.

1989

Scovel and D. D. Holm, "Symplectic Integration."

1990

G. D. Doolen and B. Hasslacher, in "Lattice Boltzmann" discuss optimal shadowing evolution on rugged landscapes.

1991

N. Metropolis, G.-C. Rota, and J. Stein, "Theory of Symmetry Classes."

This theory provides two characterizations showing that elements belonging to any symmetry class can be described in one of two ways: (1) as solutions of explicitly given sets of linear equations and (2) as linear combinations of "simple" elements of a given symmetry class.

1992

Chen, "Reaction Diffusion Patterns in 512^3 3-D Turbulence Simulation."

J. M. Hyman and Rosenau, "Compactons."

1993

Camassa and Holm, "Camassa-Holm Equation" *Phys. Rev. Lett.* **71**, 1661.

In *Phys. Rev. Lett.* **71**, 1661 (1993), Camassa and Holm (CH) derived their equation from shallow-water theory, found its Lie-Poisson Hamiltonian structure, interpreted its dynamics as being geodesic motion, discovered its peakon (weak) solutions and proved that the peakons are solitons by finding the Lax pair representation of the peakon isospectral problem. (This matrix Lax pair is a special case of the differential isospectral problem and compatible eigenfunction evolution equation for their completely integrable partial differentiation equation.) The CH equation is one order more accurate in asymptotic approximation for shallow-water wave dynamics beyond Korteweg-de Vries (KdV), yet it still preserves complete integrability and has soliton solutions that may be obtained via the inverse scattering transform method. Its soliton solutions contain both the KdV solitons and the CH peakons as limiting cases.

1994

I. Gabitov, D. D. Holm, B. Luce, and Camassa, "Nonlinear Optics."

The U.S. Patent Office granted U.S. PATENT #6,157,762 for the idea of using optical pulse reshaping devices called nonlinear amplifying loop mirrors (NALMs) in an iterated mapping to stabilize, shape and regenerate optical pulses in fibers at high bit rates for telecommunications. The idea treats the pulse propagation and reamplication process as an iterated mapping, in the Los Alamos T-Division tradition established earlier by Metropolis, Stein, and Stein, and by Feigenbaum. The theoretical design concepts were published in I. Gabitov, D. D. Holm, B. Luce, and A. Mattheus, *Optics Lett.* **20**, 2490–2492 (1995). These theoretical design concepts that led to this patent were successfully proof-tested by German Telekom using the commercial optical fiber network already installed in Europe to achieve signal propagation at 40 Gb/s over distances of thousands of kilometers.

G. A. Baker, Jr., "The Markov Property Method Applied to Ising Model Calculations," *J. Stat. Phys.* **77**, 955.

In this paper, a number of magnetic properties, not available in the Onsager, zero-magnetic-field solution of the 2-D Ising model are evaluated exactly on the square lattice for squares up to 10 by 10 with periodic boundary conditions. In the latter case there are 1.268×10^{30} separate terms to be evaluated, which is way beyond the capabilities of any projected computer. This triumph was accomplished by the elimination of redundant computations. It stood as a world record until 2001, which is a very long time, given the pace of computational development.

1995

S. A. Perelson, "Spiral Defect Chaos in HIV Dynamics."

N. Metropolis, G.-C. Rota, and J. Stein, "Symmetry Classes of Functions."

1996

G. A. Baker, Jr., and P. R. Graves-Morris, "Padé Approximants," *Encyclopedia of Mathematics and its Applications*, 2nd ed., G.-C. Rota, Ed. (Cambridge University Press, New York), Vol. 59, p. i and p. 1.

The field of Padé approximants since its modern rise in the early 1960s has grown tremendously. This book is encyclopedic in its scope and has been referred to by other workers in the field as the "red bible." When the modern period started, the information on Padé approximants was scattered throughout the literature, some of which was in rather obscure places. Thanks to the efforts of many scholars of the subject, a great deal of it has been brought to light and a great many researchers have added much that is new and valuable. The authors have combined the cream of the crop in this volume.

G. A. Baker, Jr., and N. Kawashima, "Renormalized Coupling Constant in the Ising Model," *J. Phys. A* **29**, 7183.

Over the years, there has been considerable interest in whether the so-called hyperscaling relations between the exponents of the limiting behavior of various quantities as the critical point was approached were correct for the famous 3-D Ising model. In the light of the success of the Nobel-prize-winning renormalization group theory, which in a least one major school of this theory predicted the validity of the hyperscaling relations here, more and more people began to believe them to be correct. By a hyperscaling relation is meant one that involves the spatial dimension. Yet there was no definitive computation that decided the issue. The authors have completed a large-scale, Monte Carlo calculation, which differed from others because they introduce a new estimator that was much better behaved. The results of their computation showed clearly that a quantity, the renormalized coupling constant, was different from zero, which is a sufficient condition for the hyperscaling relations to be valid.

1997

D. D. Holm and Shkoller, "Quantifying Uncertainty Euler-Poincaré Alpha-LANS Fluid Equations."

1998

D. D. Holm, J. E. Marsden, and T. S. Ratiu, "The Euler-Poincaré Equations and Semidirect Products with Applications to Continuum Theories," *Adv. in Math*. **137**, 1.

Euler-Poincaré equations are the Lagrangian analog of Lie-Poisson equations in the Hamiltonian formulation of fluid dynamics. This paper extends the idea of symplectic Hamiltonian reduction by Lie groups of symmetries in classical mechanics to reduction of *Hamilton's principle* by Lie symmetries. It also extends its applications to the case of continuum mechanics and reduction of Hamilton's principle by the infinite dimensional diffeomorphism group. The result is a powerful mathematical tool that facilitates the applications of asymptotics and averaging methods in formulating and analyzing new fluid dynamics models such as the Lagrangian-averaged Euler equations. The latter equations form the basis for deriving new Lagrangian-averaged turbulence models for a wide range of applications.

S. Chen, C. Foias, D. D. Holm, E. J. Olson, E. S. Titi, and S. Wynne, "The Camassa-Holm Equations as a Closure Model for Turbulent Channel and Pipe Flows," *Phys. Rev. Lett.* **81**, 5338.

The authors of this paper propose and provide the first experimental tests of the equations that later became known as the Lagrangian-averaged Navier-Stokes-alpha (LANS-alpha) equations for modeling and analyzing turbulence. In regularizing the Navier-Stokes equations, the LANS-alpha equations provide a turbulence model that is accurate at scales larger than the length-scale alpha and is rigorously computable because it has a finite-dimensional global attractor. The LANS-alpha equations are currently being implemented at various turbulence centers, including the Los Alamos Climate, Ocean, and Sea Ice Simulation project for high-resolution modeling of ocean circulation dynamics—an important component of Earth's climate.

1999

J. M. Hyman and Shashkov, Mimetic Discretizations for Maxwell's Equations," *J. Comp. Phys*. **151**.

The authors construct reliable fine difference methods for approximating the solution to Maxwell's equations using accurate analogs of differential operators that satisfy the identities and theorems of vector and tensor calculus in discrete form.

J. D. Louck and W. Y. C. Chen, "Enumeration of Cubic Graphs by Inclusion-Exclusion," *J. Comb. Theory A* **86**.

The authors use the principle of inclusion and exclusion to enumerate cubic graphs. The work was motivated by studies of generating functions of $3n - j$ coefficients in angular momentum theory.

2000

W. Y. C. Chen, J. D. Louck, and J. Wang, "Adjacency and Parity Relations of Words in Discrete Dynamical Systems," *J. Comb. Theory A* **91**.

Words on two letters label the branches of the inverse graph of the *n*th iterate of the parabolic map of the real line. The subset of lexical words on length $n - 1$ are key entities in this description, as are the divisor set of lexical words of degree D such that $1 + D$ divides n. It is proved that adjacent sequences in the ordered divisor set alternate in the parity of their lengths. This corresponds to an oscillatory motion of the central point back and forth throughout the central point.

2001

Compacton Solutions in a Class of Generalized Fifth-Order Korteweg-de Vries Equations

Solitons play a fundamental role in the evolution of general initial data for quasilinear dispersive partial differential equations, such as the Korteweg-de Vries, nonlinear Schrödinger, and the Kadomtsev-Petviashvili equations. We derive and investigate a new KdV-like Hamiltonian partial differential equation from a four-parameter Lagrangian where the nonlinear dispersion gives rise to solitons with compact support (compactons).

2002

G. A. Baker, Jr., "Effects on the Structure of the Universe of an Accelerating Expansion," *General Relativity and Gravitation* **34**, 767 (2002).

In this paper the author shows that if the universe is expanding, as is currently found experimentally, then there will be local clusters of matter, the size of which is limited by their total mass. If the cluster were larger, it would necessarily be torn apart by the expansion of the universe. This result is expressed by a domain-size mass relationship. In some separate results, the author points out that in the Einstein-Straus "Swiss-cheese model" there exist orbits that are discontinuous in their initial position and velocities. This model was previously known to be unstable under perturbations. The "Swiss-cheese model" was introduced in an effort to reconcile the apparently static space seen in the

solar system with the expanding space that seems to separate the galaxies. The author gives a new, homogeneous, mass-energy density solution to the Einstein field equations that matches the large-scale expansion and has only corrections of the order of the square of the Hubble constant to the behavior of the solar system. These corrections are many orders of magnitude smaller than can be currently observed. A static special case of this solution is the curved-space extension of the Schwarzschild-Lemaître solution.

T-Division's Emil Mottola has some more recent material on general relativity that has attracted considerable notice.

Fig. 1. The M31 galaxy (Andromeda galaxy).

Solitons, Peakons, and Nonlinear Integrability

by Darryl D. Holm

Modern scientists call a solitary wave that propagates without dispersing its energy over larger and larger regions of space "solitons" to emphasize their coherent, particle-like property. Solitons are used today as a paradigm for the coherent dynamical behavior of waves and pulses throughout science: from hydrodynamics to nonlinear optics, from plasmas to cell membranes, from tornados to the Great Red Spot of Jupiter.

The Fermi-Pasta-Ulam Problem

In a famous 1955 paper that became the starting point for many subsequent nonlinear studies, Fermi, Pasta, and Ulam (FPU) examined waves in the vertical motion of particles on a lattice that are attached horizontally by nonlinear springs. In his introduction [1] to the *Collected Papers of Enrico Fermi* written in 1965, Ulam remarks that

> Fermi expressed often a belief that future fundamental theories in physics may involve nonlinear operators and equations, and that it would be useful to attempt practice in the mathematics needed for the understanding of nonlinear systems. The plan was then to start with the possibly simplest such physical model and to study the results of the calculation of its long-time behavior.

FPU numerically simulated the resulting wave motion on their discrete lattice using numerical algorithms for which they point out, "The corresponding partial differential equation ... is the usual wave equation plus nonlinear terms of a complicated nature."

In his autobiography, Ulam gave a summary of the surprising "recurrences" that FPU found in some of their experiments,

> It was the consideration of an elastic string with two fixed ends, subject not only to the usual elastic force [of strain] proportional to strain, but having, in addition, a physically correct small nonlinear term. The question was to find out how this nonlinearity after very many periods of vibration would gradually alter the well-known periodic behavior ... and how, we thought, the entire motion would ultimately thermalize, imitating perhaps the behavior of fluids which are initially laminar and become more and more turbulent. ... The results were entirely different qualitatively from what even Fermi, with his great knowledge of wave motions, had expected. The original objective had been to see at what rate the energy of the string, initially put into a single sine wave ... would gradually develop higher tones with the harmonics, and how the shape would

> finally become 'a mess' ... To our surprise the string started playing a game of musical chairs, only between several low notes, and ... came back almost exactly to its original sinusoidal shape." [2], pp. 226–227.

These FPU "recurrences" were a mystery until the mid 1960s when scientists using computer simulations to study nonlinear wave propagation began finding hints for their explanation. In 1967 Kruskal and Zabusky showed that the continuum limit of the FPU equation is the Korteweg de Vries equation (KdV), which was known to describe nonlinear evolution of shallow water waves. By a glorious accident, it turned out the discrete lattice equation that FPU had chosen had so many conservation laws that it was nearly integrable, being a finite dimensional approximation of KdV, which is an integrable partial differential equation.* By the KAM theorem due to Kolmogorov, Arnold and Moser, perturbations of an integrable system will have many remaining periodic orbits and some traces of these orbits are what FPU discovered in their famous recurrences. A brilliant summary of how this circle of ideas stood in 1975 was presented by Peter Lax at a conference in Los Alamos in honor of Stan Ulam's sixty-fifth birthday. [4]

The Korteweg de Vries Equation and Solitons

The KdV equation for shallow water motion has a solitary traveling wave solution called a "soliton" that has a long history with its own mystique in science.

Over one hundred and fifty years ago, while conducting experiments to determine the most efficient design for canal boats, a young Scottish engineer named John Scott Russell (1808–1882) made a remarkable scientific discovery. As he described it in his "Report on Waves" [5]:

> I was observing the motion of a boat which was rapidly drawn along a narrow channel by a pair of horses, when the boat suddenly stopped—not so the mass of water in the channel which it had put in motion; it accumulated round the prow of the vessel in a state of violent agitation, then suddenly leaving it behind, rolled forward with great velocity, assuming the form of a large solitary elevation, a rounded, smooth and well-defined heap of water, which continued its course along the channel apparently without change of form or diminution of speed. I followed it on horseback, and overtook it still rolling on at a rate of some eight or nine miles an hour, preserving its original figure some thirty feet long and a foot to a foot and a half in height. Its height gradually diminished, and after a chase of one

Fermi felt very strongly that every physicist should be able to make back-of-the envelope calculations of anything and obtain an answer which is correct to within an order-of-magnitude. For example, at tea one afternoon, he asked his associates to estimate the number of railroad locomotives in the United States. A typical solution would involve estimating the number of miles, on the average, that you would have to drive before you crossed a railroad track; from this, you obtain the total number of miles of railroad in the United States; then you estimate the number of miles of track per locomotive. By stressing the orders-of-magnitude of things, Fermi wanted his boys to understand what is important and what is negligible. He felt that we learn all of the complicating factors in the problems that we study, but we seldom take the time to evaluate their relative significance.

—Joseph O. Hirschfelder [former group leader in T-Division], "The Scientific and Technological Miracle at Los Alamos," in Reminiscences of Los Alamos: 1943–1945, pp. 81–82.

*A classical Hamiltonian mechanical system with N degrees of freedom is said to be integrable, if it has N functionally independent constants of motion whose mutual Poisson brackets all vanish on the $2N$-dimensional phase space. In this case, the motion can be reduced to canonically conjugate action-angle variables with N constant actions and N periodic angles that change linearly with time. KdV is an example of an infinite dimensional integrable Hamiltonian system.

> or two miles I lost it in the windings of the channel. Such, in the month of August 1834, was my first chance interview with that singular and beautiful phenomenon which I have called the Wave of Translation.

This event took place on the Union Canal at Hermiston, very close to the present-day campus of Heriot-Watt University, Edinburgh. Following this discovery, John Scott Russell built a 30-foot-long wave tank in his back yard so he could continue making observations of the properties of the solitary wave in shallow water that he called the "wave of translation."

Modern scientists call such solitary waves "solitons" to emphasize their coherent, particle-like property. Solitons are used today as a paradigm for the coherent dynamical behavior of waves and pulses throughout science: from hydrodynamics to nonlinear optics, from plasmas to cell membranes, from tornados to the Great Red Spot of Jupiter. For a more detailed and technical account of the solitary wave, see [6].

The soliton first discovered in 1834 by John Scott Russell is governed by Euler's equations for an incompressible fluid with a free surface moving under gravity. At linear order in the two small parameters for shallow water asymptotics, Euler's equations may be approximated by the KdV. The KdV soliton arises as a traveling wave whose steady $\mathrm{sech}^2(x)$ profile in elevation results from a balance between nonlinear steepening and third order linear dispersion. At the next, quadratic, order in the shallow water asymptotic expansion beyond KdV, the soliton is governed by the Camassa-Holm equation (CH). Both KdV and CH are completely integrable by the inverse scattering transform (IST) method.

The IST method may be used to solve certain nonlinear integrable evolutionary partial differential equations by a series of linear operations. For these nonlinear equations, the IST method is in some ways an analog of the Fourier transform method for linear equations. That is, nonlinear integrable equations that admit the IST approach may be solved by a series of linear operations involving the evolution of the eigenfunctions ψ belonging to an isospectral eigenvalue problem $L\psi = = \lambda\psi$ for a linear operator L with eigenvalue spectrum λ. Typically, the discrete spectrum of L comprises the speeds of the solitons and its continuous spectrum comprises other, slower, radiation that is encoded in the initial condition for the nonlinear evolution equation and is left behind when the solitons emerge. To everyone's delight, the linear isospectral problem for KdV turned out to be the Schrödinger equation. The corresponding isospectral problem for CH is a linear combination of the Schrödinger equation and the equation for the vibrations of a nonuniform string. [7,8] For reviews of the IST approach for solving soliton equations, see [9] and [10].

Solitons, Peakons, and Future Research Directions

To be integrable, a Hamiltonian system typically must have a great deal of mathematical structure. The inverse scattering transform is just the beginning. The discovery of additional mathematical structures possessed by integrable equations such as KdV and CH for shallow water waves has provided new insights for the dynamical systems approach to partial differential equations. These mathematical structures include deep connections to two other branches of mathematics.

- Variational formulation: CH was generalized to higher spatial dimensions in the Euler-Poincaré framework for Hamilton's principle on the Lie algebra of vector fields. [11]
- Geometric aspects, CH was understood as geodesic flow on the infinite dimensional Lie group of smooth maps, with respect to the metric given by its kinetic energy. This understanding lies at the heart of fluid dynamics. [11]

In turn, discovery of these mathematical structures in the fundamental nonlinear nature of fluid dynamics has led to new technological bases for modeling turbulence in fluids. The prime example is the LANS-alpha turbulence closure model. The acronym LANS-alpha stands for Lagrangian-Averaged Navier-Stokes, and alpha is the correlation length scale for the turbulent Lagrangian fluid trajectories. The LANS-alpha turbulence model is discussed elsewhere in this volume and more recent results are reviewed in [12].

We shall conclude by illustrating the remarkable "peakon" solutions of the CH equation, discovered in [7]. The peakons are weak solutions for the zero-dispersion limit of CH and they are also genuine solitons. That is, they propagate as confined pulses that interact elastically and they correspond to spectral data of the CH isospectral eigenvalue problem. For example, the N–peakon solution provides the dominant emergent pattern in the solution of the initial value problem for the CH equation with spatially confined initial conditions. This is confirmed in Fig. 1. The coherent nonlinear interactions of the peakons for the CH equation seen in this figure raise vital questions about aspects of balance and evolution whose answers must be pursued before one can fully understand their three-dimensional behavior as a basis for modeling turbulence. Pursuing the answers to these questions will be the objective of future research.

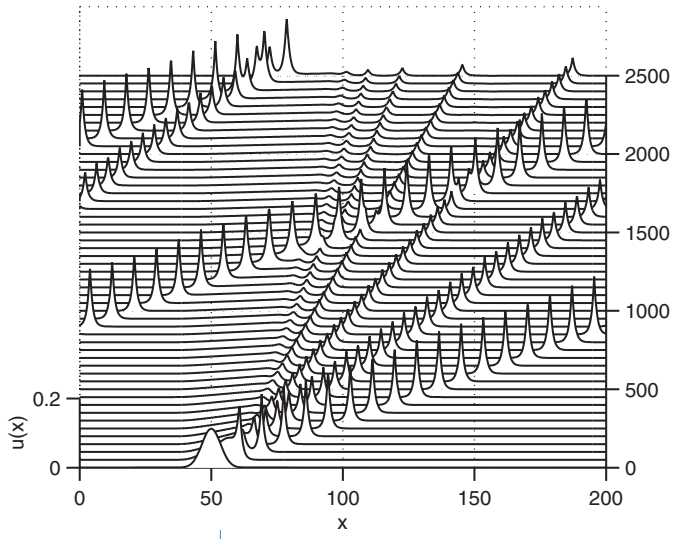


Fig. 1. CH evolution of peakons emerging from a Gaussian initial velocity distribution of unit area and width 5α with α = 1. The peakon shape is $g(\chi) = e^{-|\chi|/\alpha}$, which is the Green's function for the Helmholtz operator $Q_{op} = 1 - \alpha^2\partial^2$ in one spatial dimension. Several peakons emerge from the initial Gaussian. The speed of each peakon is equal to its height and the peakons interact eleastically as they cross and recross the periodic domain. (Fig. 1 was kindly produced by Martin Staley.)

Alcouffe
Dendy
Gauss-Seidel
Jacobi
Knoll
Mousseau
Oliphant

Solution of Linear and Nonlinear Equations

by Joel E. Dendy, Jr.

The modeling of physical processes requires the solution of systems of partial differential equations, which are usually too complicated to be solved analytically. The numerical solution of the process by which a continuum system is converted into a system of algebraic equations approximates the solution of the continuum system. In most situations, one is interested in modeling how a physical process evolves in time. In such cases the time period of interest is divided into many small time steps and the computational algorithm proceeds from one step to the next by performing a small number of arithmetic operations for each unknown.

In the early history of computing, one encountered the problem of stability. For certain computations to give meaningful answers, stability requires that the number of time steps be impossibly large. The alternative is to model some of the physics in "steady state" mode, giving rise to the need to solve implicit algebraic equations for every time step of the calculation. In the early days of computing, if these equations were nonlinear, it was absolutely essential to linearize them. Even so, the direct solution of the linear system of equations, usually called Gaussian elimination, was too time-consuming to be tractable, requiring too many numerical operations and too much storage per unknown.

An alternative to direct solution is iterative solution. One guesses an answer and then performs an "iteration" of some process to improve the solution. Early methods, for example, Gauss-Seidel and Jacobi, required a small number of arithmetic operations for each iteration but many iterations to achieve a good approximation to the answer. Even so, storage considerations made these methods attractive. The history of numerical analysis in this area is full of advances in special cases, all directed towards making the number of operations as small as possible. Several of these advances took place at Los Alamos, some in T-Division.

One of these advances was anticipated by T. A. Oliphant of T-Division. The paper [1] is cited as being one of the early papers on approximate or incomplete factorization methods. [2] However, it took until 1977 for these methods to become widely used. [3] They are still used extensively in such areas as petroleum reservoir engineering. This paper is also an example of the idea of preconditioning: to accelerate the solution of a given linear system, form a related linear system that is easier to solve.

One very important topic at the Laboratory is the numerical solution of the multigroup neutron transport equation, with dependent variables of space, angle, and energy. Discretization in energy leads to many energy groups for each of which discretization in space and angle occurs. The coupling of energy groups is via an integral with respect to space, angle, and energy. Typically, this integral is frozen for current values. Then a sweep through each energy group occurs, with each group requiring a sweep through the spatial mesh for each angle. Each of these sweeps is explicit, requiring a small number of arithmetic operations for each point in the spatial mesh. This very complicated iteration strategy converges rapidly except in the important case of optically thick regions with scattering ratio near unity. For such regions, however, the multigroup diffusion equation is a good approximation. An early idea was to solve this equation approximately and to use the approximate solution to improve convergence for the solution of the transport equation. Unfortunately, early attempts were unstable for exactly the case where they were needed most. Alcouffe's contribution was to show how to difference the multigroup diffusion equation consistent with the multigroup transport equation so as to arrive at a stable and very effective iterative method. [4]

The next issue, given this stability, is to solve the multigroup diffusion equation efficiently. Again, coupling between energy groups is via an integral, which is calculated approximately for current values. Then a diffusion equation for each energy group is solved approximately. Alcouffe first used successive over-relaxation (SOR) for this task. The paper showed how to use the multigrid method instead. [5]

Many iterative methods rely on local solution techniques. Gauss-Seidel, for example, sweeps through the mesh, changing the value of the approximate solution at each point so as to make the residual zero at that point, i.e., so as to satisfy the equation at that point exactly. In the process, of course, the residual at neighboring points becomes nonzero. Such iterative methods require very few numerical operations per grid point but many sweeps through the mesh are required. Technically, high frequency errors are annihilated quickly, but smooth errors are annihilated very slowly. The idea behind the multigrid method is to use coarser grids to annihilate these smooth errors on coarser grids, where this process is much cheaper. In fact, the total work required for multigrid to solve the linear system to a given tolerance is proportional to the number of unknowns; one can do no better, as just writing down the answer takes that much work. At the time [5], the multigrid method had been successfully applied to many problems with smooth coefficients. The major advance [5] was to show how to apply it to problems with discontinuous coefficients, these arising because of interfaces between different materials. One of the methods in Alcouffe et al. [5] The method was extended [6] in what some consider to be the first of the Algebraic Multigrid Methods. These methods attempt to use just coefficient

information to generate the auxiliary coarse grids needed for a multigrid solution. This holy grail is still an area of active research. In terms of applications at Los Alamos, the method is still being used as a preconditioner for the solution of the Morel diffusion scheme. [7]

One of the recent exciting developments in which T-Division has participated is a new technique for the solution of nonlinear equations. Simulations at Los Alamos usually involve many kinds of physics and thus lead to systems with several types of coupled unknowns, e.g., energy, temperature, velocity, etc. To solve these systems in the past, one resorted to "split physics," linearization, and successive substitution. It was deemed too expensive to solve the whole system "honestly" by forming the Jacobian, for solution by Newton's method. And it is still too expensive today, to be honest. The exciting alternative is called Jacobian Free Newton-Krylov. [8] The idea is to approximate the application of the Jacobian to the current iterate to generate the next iterate and to employ Krylov subspace techniques to accelerate convergence. Further acceleration is achieved through preconditioning. But the key idea, in which T-Division has played a leading role, is to use knowledge of the physics to form the preconditioner. Examples include Mousseau et al [8] and Knoll et al, [9], where the linearized, split form of the system is the "physics-based" preconditioner. This preconditioner is solved approximately with the aid of a simple form of multigrid. Further development of these ideas is an active area of research.

Nonlinear Science from Paradigms to Practicalities

by David K. Campbell

This article is excerpted with permission from *Los Alamos Science*, *Special Issue: Stanislaw Ulam 1909–1984* (Los Alamos National Laboratory, 1987) pp. 218–262.

No tribute to the legacy of Stan Ulam would be complete without a discussion of "nonlinear science," a growing collection of interdisciplinary studies that in the past two decades has excited and challenged researchers from nearly every discipline of the natural sciences, engineering, and mathematics. Through his own research Stan played a major role in founding what we now call nonlinear science, and through his encouragement of the work of others, he guided its development. In this survey article I will try to weave the thread of Stan's contributions into the pattern of recent successes and current challenges of nonlinear science. At the same time I hope to capture some of the excitement of research in this area.

Introduction

Let me start from a very simple, albeit circular, definition: nonlinear science is the study of those mathematical systems and natural phenomena that are *not* linear. Ever attuned to the possibility of *bons mots*, Stan once remarked that this was "like defining the bulk of zoology by calling it the study of 'non-elephant animals'." His point, clearly, was that the vast majority of mathematical equations and natural phenomena are nonlinear, with linearity being the exceptional, but important, case.

Linear versus Nonlinear

Mathematically, the essential difference between linear and nonlinear equations is clear. Any two solutions of a linear equation can be added together to form a new solution; this is the *superposition principle*. In fact, a moment of serious thought allows one to recognize that superposition is responsible for the systematic methods used to solve, independent of other complexities, essentially *any* linear problem. Fourier and Laplace transform methods, for example, depend on being able to superpose solutions. Putting it naively, one breaks the problem into many small pieces, then adds the separate solutions to get the solution to the whole problem.

In contrast, two solutions of a nonlinear equation *cannot* be added together to form another solution. Superposition fails. Thus, one must consider a nonlinear problem *in toto*: one cannot—at least not obviously—break the problem into small subproblems and add their solutions. It is therefore perhaps not surprising that no general analytic

approach exists for solving typical nonlinear equations. In fact, as we shall discuss, certain nonlinear equations describing chaotic physical motions have *no* useful analytic solutions.

Physically, the distinction between linear and nonlinear behavior is best abstracted from examples. For instance, when water flows through a pipe at low velocity, its motion is *laminar* and is characteristic of linear behavior: regular, predictable, and describable in simple analytic mathematical terms. However, when the velocity exceeds a critical value, the motion becomes *turbulent*, with localized eddies moving in a complicated, irregular, and erratic way that typifies nonlinear behavior. By reflecting on this and other examples, we can isolate at least three characteristics that distinguish linear and nonlinear physical phenomena.

First, the motion itself is qualitatively different. Linear systems typically show smooth, regular motion in space and time that can be described in terms of well-behaved functions. Nonlinear systems, however, often show transitions from smooth motion to chaotic, erratic, or, as we will see later, even apparently random behavior. The quantitative description of chaos is one of the triumphs of nonlinear science.

Second, the response of a linear system to small changes in its parameters or to external stimulation is usually smooth and in direct proportion to the stimulation. But for nonlinear systems, a small change in the parameters can produce an enormous qualitative difference in the motion. Further, the response to an external stimulation can be different from the stimulation itself: for example, a periodically driven nonlinear system may exhibit oscillations at, say, one-half, one-quarter, or twice the period of the stimulation.

Third, a localized "lump," or pulse, in a linear system will normally decay by spreading out as time progresses. This phenomenon, known as dispersion, causes waves in linear systems to lose their identity and die out, such as when low-amplitude water waves disappear as they move away from the original disturbance. In contrast, nonlinear systems can have highly coherent, stable localized structures—such as the eddies in turbulent flow—that persist either for long times or, in some idealized mathematical models, for all time. The remarkable order reflected by these persistent coherent structures stands in sharp contrast to the irregular, erratic motion that they themselves can undergo.

Continuing from p. 223:

Paradigms of Nonlinearity

Before examining in some detail how this challenge is being confronted, we need to respond to some obvious but important questions. First, why study nonlinear *science*, rather than nonlinear *chemistry*, or nonlinear *physics*, or nonlinear *biology*? Nonlinear science sounds impossibly broad, too interdisciplinary or "the study of everything." However, the absence of a systematic mathematical framework and the complexity of

natural nonlinear phenomena suggest that nonlinear behavior is best comprehended by classifying its various manifestations in many different systems and by identifying and studying their common features. Indeed, both the interest and the power of nonlinear science arise precisely because common concepts are being discovered about systems in very different areas of mathematics and natural sciences. These common concepts, or *paradigms,* give insight into nonlinear problems in a large number of disciplines at once. By understanding these paradigms, one can hope to understand the essence of nonlinearity as well as its consequences in many fields.

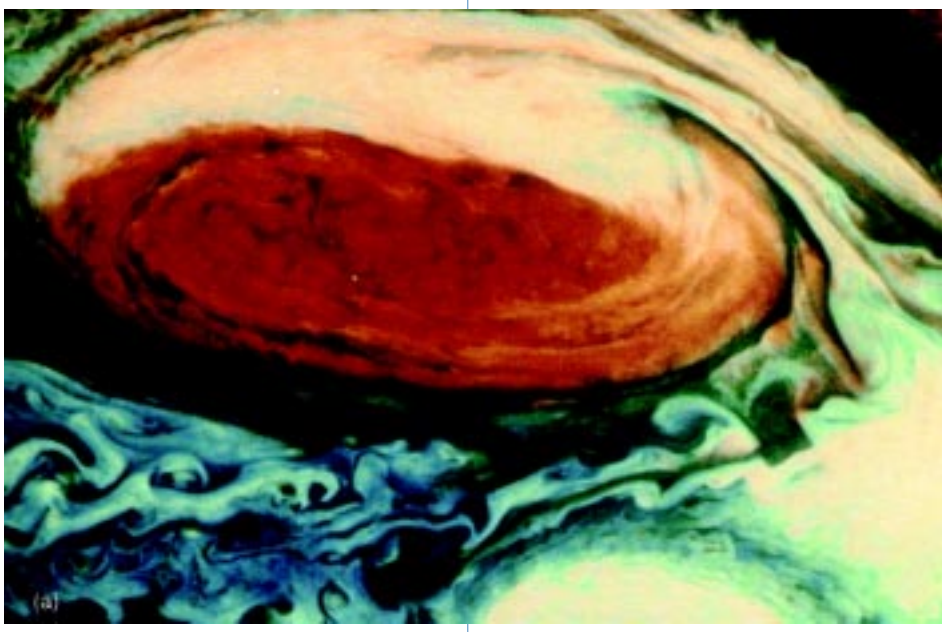
Second, since it has long been known that most systems are inherently nonlinear, why has there been a sudden blossoming of interest in this field in the past twenty years or so? Why weren't many of these fundamental problems solved a century ago? On reflection, one can identify three recent developments whose synergistic blending has made possible revolutionary progress.

The first, and perhaps most crucial, development has been that of high-speed electronic computers, which permit quantitative numerical simulations of nonlinear systems. Indeed, the term *experimental mathematics* has been coined to describe computer-based investigations into problems inaccessible to analytic methods. Rather than simply confirming quantitatively results already anticipated by qualitative analysis, experimental mathematics uses the computer to generate qualitative insight where none has existed before. As the visionary of this development, John von Neumann, wrote (in a 1946 article called "On the principles of large scale computing machines"):

"Our present analytic methods seem unsuitable for the solution of the important problems rising in connection with nonlinear partial differential equations and, in fact, with virtually all types of problems in pure mathematics. ... really efficient high-speed computing devices may, in the field of nonlinear partial differential equations as well as in many other fields which are now difficult or entirely denied of access, provide us with those heuristic hints which are needed in all parts of mathematics for genuine progress."

Stan Ulam, together with many of his Los Alamos colleagues, was one of the very first to make this vision a reality. Among Stan's pioneering experimental mathematical investigations was the seminal study of the [Fermi, Pasta, Ulam] FPU problem [in deleted material]. Another example was his early numerical work on nonlinear mappings, carried out in collaboration with Paul Stein... Both of these studies will figure in our later discussion.

The second crucial development has been the experimental observation of "universal" nonlinear characteristics in natural systems that range from chicken hearts and chemical reactors to fluids and plasmas. In the past decade these experiments have reached previously inaccessible levels of precision, so that one can measure *quantitative* similarities in,



COHERENT STRUCTURES IN NATURE
Fig. 1(a). A closeup of the Red Spot of jupiter, taken from the Voyager spacecraft. False color is used to enhance features of the image. In addition to the celebrated red Spot, there are many other "coherent structures" on smaller scales on Jupiter. (Photo courtesy of NASA.)

for example, the route to chaotic behavior among an enormous variety of nonlinear systems.

The third and final development has been in the area of novel analytical mathematical methods. For instance, the invention of the *inverse spectral transform* has led to a systematic method for the explicit solution of a large number of nonlinear partial differential equations. Similarly, new methods based on the theory of Hamiltonian systems allow the analysis of nonlinear stability of a wide range of physically relevant mathematical models.

Continuing from p. 224:

Coherent Structures and Solitons

From the Red Spot of Jupiter through clumps of electromagnetic radiation in turbulent plasmas to microscopic charge-density waves on the atomic scale, spatially localized, long-lived, wave-like excitations abound in nonlinear systems. These nonlinear waves and structures reflect a surprising orderliness in the midst of complex behavior. Their ubiquitous role in both natural nonlinear phenomena and the corresponding mathematical models has caused coherent structures and solitons to emerge as one of the central paradigms of nonlinear science.

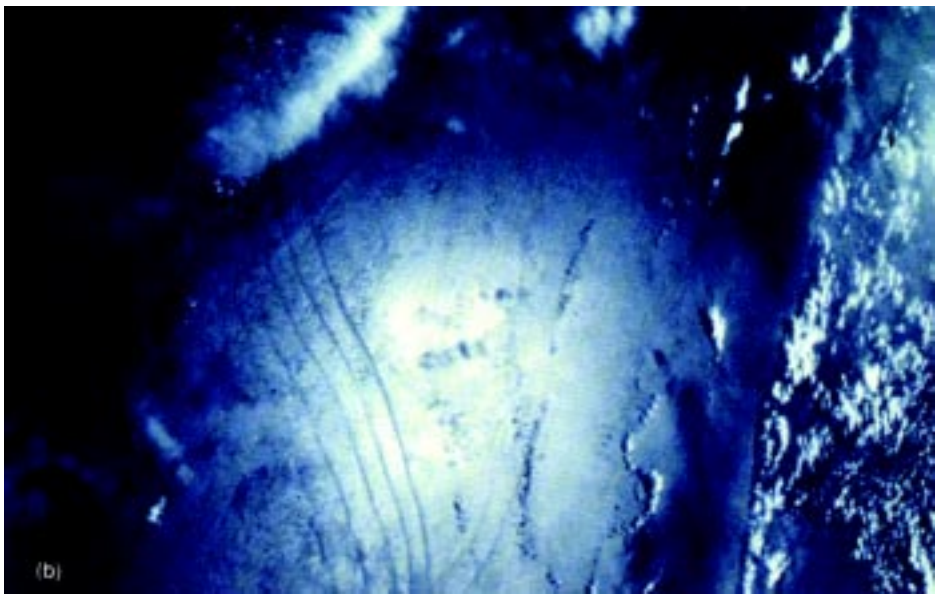


Fig. 1(b). Nonlinear surface waves in the Andaman Sea off the coast of Thailand as photographed from an Apollo-Soyuz spacecraft. (Photo courtesy of NASA.)

Coherent structures typically represent the natural "modes" for understanding the time-evolution of the nonlinear system and often dominate the long-time behavior of the motion.

To illustrate this, let me begin with one of the most familiar (and beautiful!) examples in nature. namely, the giant Red Spot (Fig. 1a). This feature, first observed from earth in the late seventeenth century, has remained remarkably stable in the turbulent cauldron of Jupiter's atmosphere. It represents a coherent structure on a scale of about 4×10^8 meters, or roughly the distance from the earth to the moon.

To give an example at the terrestrial level, certain classes of nonlinear ocean waves form coherent structures that propagate essentially unchanged for thousands of miles. Figure 1b is a photograph taken from an Apollo-Soyuz spacecraft of a region of open ocean in the Andaman Sea near northern Sumatra. One sees clearly a packet of five nearly straight surface waves; each is approximately 150 kilometers wide, so the scale of this phenomenon is roughly 10^5 meters. Individual waves within the packet are separated from each other by about 10 kilometers. The waves, which are generated by tidal forces, move in the direction perpendicular to their crests at a speed of about 2 meters per second. Although the surface deflection of these waves is small—about 1.8 meters—they can here be seen from orbit because the sun is directly behind the spacecraft, causing the specular reflection to be very sensitive to variations of the surface. These visible surface waves are actually a manifestation of much larger amplitude—perhaps ten times larger—internal waves. The internal waves exist because thermal or salinity gradients lead to a stratification of the subsurface into layers. *A priori* such large internal waves could pose a threat to submarines and to offshore structures. Indeed, the research on these waves was initiated by Exxon Corporation to assess the actual risks to the oil rigs they planned to construct in the area. Fortunately, in this context the phenomenon turned out to be more beautiful than threatening.

Continuing from p. 231:

Deterministic Chaos and Fractals

Deterministic chaos is the term applied to the aperiodic, irregular, unpredictable, random behavior that in the past two decades has been observed in an incredible variety of nonlinear systems, both mathematical and natural. Although the processes are strictly deterministic and all forces are known, the long-time behavior defies prediction and is as random as a coin toss.

That a system governed by deterministic laws can exhibit effectively random behavior runs directly counter to our normal intuition. Perhaps it is because intuition is inherently "linear;" indeed, deterministic chaos *cannot* occur in linear systems. More likely, it is because of our deeply ingrained view of a clockwork universe, a view that in the West was

forcefully stated by the great French mathematician and natural philosopher Laplace. If one could know the positions and velocities of all the particles in the universe and the nature of all the forces among them, then one could chart the course of the universe for all time. In short, from exact knowledge of the initial state (and the forces) comes an exact knowledge of the final state. In Newtonian mechanics this belief is true, and to avoid any possible confusion, I stress that we are considering only dynamical systems obeying classical, Newtonian mechanics. Subsequent remarks have nothing to do with "uncertainties" caused by quantum mechanics.

However, in the real world exact knowledge of the initial state is not achievable. No matter how accurately the velocity of a particular particle is measured, one can demand that it be measured more accurately. Although we may, in general, recognize our inability to have such exact knowledge, we typically assume that if the initial conditions of two separate experiments are *almost* the same, the final conditions will be the same. For most smoothly behaved, "normal" systems this assumption is correct. But for certain nonlinear systems it is false, and deterministic chaos is the result.

At the turn of this century, Henri Poincaré, another great French mathematician and natural philosopher, understood this possibility very precisely and wrote (as translated in *Science and Method*):

"A very small cause which escapes our notice determines a considerable effect that we cannot fail to see, and then we say that that effect is due to chance. If we knew exactly the laws of nature and the situation of the universe at the initial moment, we could predict exactly the situation of that same universe at a succeeding moment. But even if it were the case that the natural laws had no longer any secret for us, we could still only know the initial situation *approximately*. If that enabled us to predict the succeeding situation *with the same approximation*, that is all we require, and we should say that the phenomenon had been predicted, that it is governed by laws. But it is not always so; it may happen that small differences in the initial conditions produce very great ones in the final phenomena. A small error in the former will produce an enormous error in the later. Prediction becomes impossible, and we have the fortuitous phenomenon."

Despite Poincaré's remarkable insight, deterministic chaos remained virtually unexplored and unknown until the early 1960s. As the ensuing discussion will reveal, the reason for this long hiatus is that chaos defies direct analytic treatment. The seeds planted by Poincaré could only germinate when the advances in interactive computation made experimental mathematics a reality.

Continuing from p. 247:

Practicalities

The impacts of deterministic chaos and fractals are only now beginning to be felt throughout science. The concepts that even simple systems can exhibit incredibly complicated behavior, that simple rules can lead to enormously intricate geometric objects, and that this behavior and these objects can be quantified are now all widely appreciated and are being applied in many fields.

The fractal viscous-fingering phenomenon (Fig. 2) is of enormous technological interest, for it represents a major barrier to the development of efficient advanced oil-recovery techniques. Nearly half the oil deposited in limestone or other porous media is typically unrecovered because it remains stuck in the pores. To force out this oil, water is injected into a second well nearby. Viscous fingering limits the effectiveness of this technique, because when one of the thin fingers of water breaks through from the injector to the recovery well, only injected water rather than oil is thereafter recovered. Clearly a full understanding of this fractal phenomenon and ways to control it are of considerable economic importance.

Similarly, a direct application of fractals occurs in the design of the toughened ceramics used as engine parts. These special ceramics are designed to tolerate flaws, such as voids and cracks, without breaking into pieces. The flaws arise primarily from voids that develop during the sintering process and fractures that arise chiefly from the use of hard materials when machining the ceramics. By adding secondary constituents to the ceramics, propagating cracks can be forced to move through the ceramic along tortuous, convoluted routes, causing more energy to be expended than if the route were smooth and regular. Hence, for a given impulse, an irregular crack does not propagate as far through the ceramic and does less overall damage. Convoluted routes should lead to cracks in the form of complex fractal patterns. Indeed, microscopic studies of high performance ceramics have revealed such patterns and established that the higher the fractal dimension of the cracks, the tougher the ceramics.

The results of deterministic chaos are also being applied across a broad range of disciplines. Experimentally, high-precision measurements of chaotic dynamics in many types of fluid flows, current and voltage responses of semiconductors and other solid-state electronic devices, and cardiac arhythymias have established the importance of dissipative chaos in fluid dynamics, condensed-matter physics, and medicine. Indeed, recent medical experiments have suggested that many physiological parameters vary chaotically in the *healthy* individual and that greater regularity can indicate a pathological condition; for example,

VISCOUS FINGERING
Fig. 2. A fractual structure formed by injecting water under pressure into a high-viscosity fluid. The fractal dimension of this object has been calculated to be $d_f = 1{,}70 \pm 0.05$. (Figure courtesy of Gerard Daccord and Johann Nittmann, Etudes et Fabrication Dowell Schlumberger, France, and H. Eugene Stanley, Boston University.)

normally chaotic oscillations of the densities of red and white blood cells become periodic in some anemias and leukemias. Hamiltonian chaos finds a direct application in accelerator design, where the potential loss of an apparently stable beam due to subtle long-time phenomena such as "Arnold diffusion" is a vital issue of technology.

The central theoretical challenge in "applied chaos" is to develop deterministic chaotic models to explain these diverse phenomena. Rather than focusing on the details of specific applications, let me describe two broader problem areas of current research.

First, although we have stressed the randomness and unpredictability of the *long-time* behavior of chaotic systems, it nonetheless remains true that these systems are deterministic, following laws that involve *no* external randomness or uncertainty. Thus, it is possible to predict the behavior for *short* times, if the equations of motion are known. The analytic solution of the logistic map for $r = 4$ is a clear illustration, given two initial conditions known to, say, 10-bit accuracy, one can predict the relative positions—albeit with exponentially decreasing accuracy—for 10 iterations of the map. The subtler problem, currently under intense investigation, occurs when one observes that a system is deterministically chaotic but does not know the form of the underlying equations: can one nonetheless use the basic determinism to make *some* prediction? In view of the clear value of such predictive techniques—consider the stock market—substantial efforts are being focused on this question.

Second, and at a still broader level, are the related issues of universality and mode reduction in chaos, both of which we mentioned previously. Universality implies that for certain chaotic phenomena—such as the period-doubling cascade—the details of the system and the equations describing it are *irrelevant*: the observed complex behavior develops in a similar manner in every context, be it fluid dynamics, condensed-matter physics, or biology. Indeed, the term universality is borrowed from the statistical mechanics of phase transitions, where it has been shown that the details of the microscopic interactions are irrelevant for most of the important properties of the transitions. In the context of chaos, universality also lends tremendous power to analyses of certain phenomena; in essence, the simplest example—for instance, the logistic map for period doubling—contains the critical features of the entire effect.

The central idea of mode reduction can most easily be visualized in fluid flows. In any given fluid motion not all the (infinitely!) many possible modes are "active," so the *effective* phase-space dimension is much smaller than the full dimension of the equations. The case of laminar flow in which fluid moves *en bloc* is a trivial illustration. A more interesting and much less obvious example is observed in experiments on Couette-Taylor flows, in which fluid is contained between two concentric rotating cylinders. As the speed of relative rotation is increased, the flow forms bands of Taylor vortices. Further increases in the relative rotation cause the bands to develop "wobbling" instabilities and finally to be replaced by fully developed turbulence. In these experiments clever techniques (sometimes referred to as "geometry from a time series" and related to rigorous mathematical embedding theorems) have been used to extract phase-space information directly from a time series of measurements on a single dynamical variable. Such techniques have revealed strange attractors with effective phase-space dimensions on the order of five. In such experiments there are, in one sense, on the order of only five active modes. Mode reduction reduces the number of degrees of freedom being modeled to the minimum necessary to capture the essence of the motion.


Julius Ashkin
Wayne A. Bowers
Paul Olum
Robert Serber
William R. Sarita
Earl D. Pau

Developing Nuclear Weapons

by Francis H. Harlow

Harlow
Huneter
Mark
Robinson
Wechsler

The Los Alamos Scientific Laboratory was established in 1943, with the specific goal of developing nuclear weapons, if indeed such a development was possible. The staff chosen for this purpose came from many parts of the country, with some of the most significant contributions resulting from the involvement of scientists from other parts of the world.

The project required the close collaboration of three kinds of technological disciplines: theory, experiments, and system engineering, together with numerous supporting activities in the fields of security, construction, transportation of materials, and personnel matters. A detailed history of the project is given in [1] for the years 1943–1945.

During those hectic (perhaps even frantic) years, the collaborations across the Laboratory were so closely entangled that it is often difficult to separate the activities in the Theoretical Division from those of the rest of the Laboratory. Experiments involving high explosives, for example, were almost always associated with T-Division calculations and analysis, while theoretical challenges often led to both experimental and systems-design engineering.

The core staff of T-Division in the early years is listed in the earliest surviving roster (See Fig. 1.) for the early part of 1945, and badge photos of some of those people are shown in along the bottom of the introductory pages and Fig. 2.


Richard H. Stark


Philip M. Whitman


William G. Penney


William H. Page


Ralph E. Nelson


Harold Ninger

Fig. 1. Earliest surviving roster of T-Division's core staff.

THEORETICAL DIVISION
MAY 10, 1945

H.A. Bethe, Division Leader, E-208, Ext 71
V.F. Weisskopf, Deputy Division Leader, E-217, Ext 74
J. von Neumann, Consultant, E-205½, Ext 72 R3

GROUP T-1
R. Peierls, E-119, Ext 178
R.F. Christy, Section Leader E-120, Ext 178 R2
K. Fuchs, Section Leader, E-118, Ext 77
Baroody, E.M., E-121, Ext 178 R2
Calkin, J.W., E-117, Ext 77
Inglis, D.R., T-35, Ext 54
Keller, J., E-120, Ext 178 R2
Penny, W.G., E-101A, Ext 470
Podgor, T/5 S., E-116, Ext 469
Roberts, T/5 A.E., E-117, Ext 77
Skyrme, T.H.R., E-118, Ext 77
Stark, R.H., E-116, Ext 469
Stein, T/5 P.R., E-121, Ext 178

GROUP T-2
R. Serber, Leader, E-109, Ext 177
L.I. Schiff, Alt. Leader, E-108, Ext 76
Case, K.M., E-107, Ext 76
Glauber, R., E-107, Ext 76
Kurath, T/3 D., E-202, Ext 205
Rarita, W., E-111, Ext 177
Richman, C., E-110, Ext 177
Stehle, T/5 P., E-107, Ext 76

GROUP T-3
V. Weisskopf, Leader, E-217, Ext 74
R.E. Marshak, Alt. Leader, E-218, Ext 74 R2
Bellman, Pvt. R., E-222, Ext 468
Cohen, T/5 S., E-219, Ext 74 R2
Lennox, T/5 E., E-218, Ext 74 R2
Olum, Paul, E-216, Ext 74
Smith, J.H,, E-219, Ext 74
Wing, Milton, E-222, Ext 468
Bowers, W.A., E-216, Ext 74

GROUP T-4
R. Feynman, Leader, E-206, Ext 72
J. Ashkin, E-209, Ext 72 R2
Ehrlich, R., E-210, Ext. 72 R2
Peshkin, T/4 N., E-203, Ext. 79 R3
Reines, F., E-210, Ext 72 R2
Welton, T.A., E-209, Ext 72 R2

GROUP T-5
D. Flanders, Leader, E-205, Ext 69 R1
P. Whitman, Alt. Leader, E-204, Ext 69 R2
Atkins, A.L., E-211, Ext 73 R2
Davis, R.R., E-215, Ext 74
de le Vin, E., E-201, Ext 79
Elliott, J., E-213, Ext 73 R1
Hauser, T/5 F.H., E-214,
Huber, T/4 .C., E-120, Ext 70
Hudson, H., E-214, Ext 73 R1
Inglis, B., E-212, Ext 73 R2
Johnson, M., E-212, Ext 73 R2
Kellogg , T/5 H., E-203, Ext 69 R3
Langer, B., E-201, Ext 79
Page, T/3 W., E-214, Ext 73 R1
Rau, E., T/3, E-120, Ext 70
Staley, T/3 J., E-212, Ext 73 R2
Teller, M., E-214, Ext 73 R1
Vuletic, T/5 V., E-211. Ext 73 R2
Wilson, F., E-212, Ext 73 R2
Wright, T/5 E., E-213, Ext 73 R1
Young, T/Sgt., G., E-211, Ext 73 R2

GROUP T-6
E. Nelson, Leader, E-116
N. Metropolis, Alt. Leader, E-115, Ext 78
R. Feynman, Consultant, E-206, Ext 72
Ewing, F.E., E-112, Ext 75
Goldberg, T/5., E-108, Ext 76
Kemeny, Pvt. J., E-105, Ext 75
Hamming, R.W., E-114, Ext 78
Heermans, Corp. A., E-105, Ext 75
Heller, T/5 A., E-105, Ext 75
Hurwitz, T/5 D., E-105, Ext 75
Johnston, T/3 J., E-103, Ext 75
Kington, T/4 J., E-105, Ext 75
Livesay, N., E-112, Ext 75
Ninger, H., E-105, Ext 75
Noah, F.E., E-105, Ext 75
Vorwald, T/5 A., E-105, Ext 75
Zimmerman, T/3 W., E-105, Ext 75

GROUP T-7
J. Hirschfelder, Leader, T-30, Ext 206
J. Magee, Alt. Leader, T-30, Ext 206
Brummer, T/5 E., T-28, Ext 206
Feckete, T/4 P., T-26, Ext 206
Larson, T/4 L., T-28, Ext 206
Ostrow, E., T-30, Ext 206
Schwartz, T/4 P., T-28, Ext 206

GROUP T-8
G. Placzek, Leader, E-220, Ext 468
Mark, C., E-221, Ext 468 R2
Carlson, B., E-221, Ext 468 R2
Day


Richard P. Feynman


Donald A. Flanders


John R. Von Neumann


John S. Keller


Joseph Hirschfelder


Verna M. Zimmerman

The people associated with this program included many of the smartest scientists in the world. Some came to stay for the duration of the Second World War; others were consultants who came only occasionally. One, who was in the thick of the activity, furnished a major amount of information to Russian scientists about many of the most secret developments regarding the issues surrounding the technology of the atomic bomb.

Much of the technology is still classified as SECRET, and cannot be included in this book. There are, nevertheless, enough unclassified (open) items that a fairly complete discussion of the generic issues can be presented.

The Challenge

In order to produce massive amounts of explosive energy from uranium or plutonium, the metal must be compressed to high enough density, or assembled fast enough that neutrons are not able to escape more rapidly than they multiply from fission reactions. For example, the mass associated with the borderline between supercritical and subcritical for a bare sphere at normal density is 10.5 kg for α-phase plutonium and 16 kg for δ-phase plutonium. One of the difficulties associated with obtaining a suitable supercritical state lies in achieving fast enough assembly that the fission energy cannot commence blowing the metal apart before it has had time to reach a high degree of supercriticality.

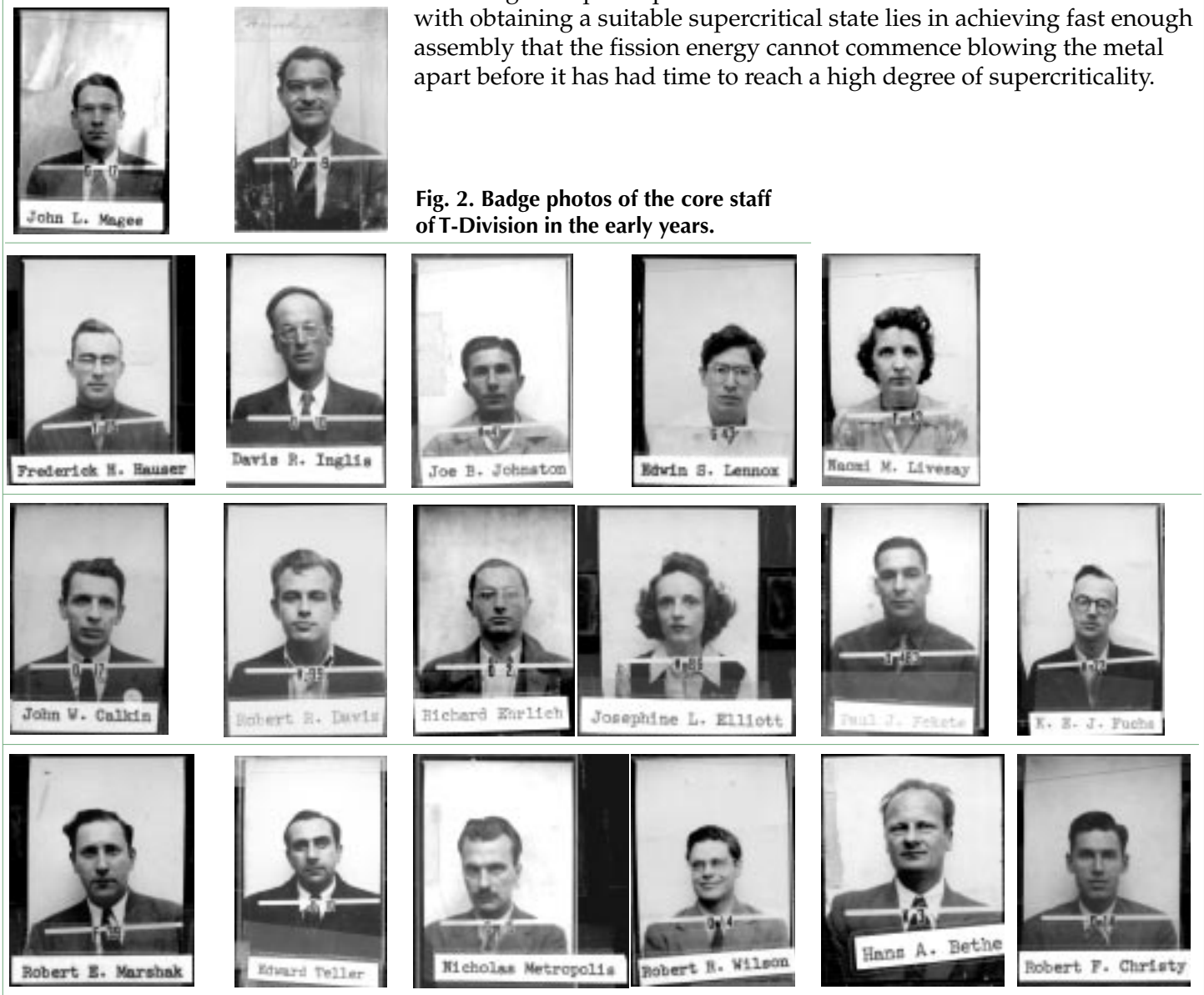


Fig. 2. Badge photos of the core staff of T-Division in the early years.

On the other hand, it is possible that simple hydrodynamic rebound could commence the disassembly before the fission reactions start, thereby also lowering the efficiency (defined as the ratio of metal consumed by fission reactions to the total amount of metal). Thus, the challenge was to develop a rapid assembly technique and to make sure that the reactions were initiated at the optimal stage of assembly. Three kinds of ideas were explored: gun assembly, high-explosive compression of a solid sphere of metal (initially subcritical), and high-explosive compression of a spherical shell.

The Gun Device

The basic principle is quite simple (See Fig. 3.) Two subcritical hemispheres are slammed together, driven by a charge of propellant behind one of them.

Among the issues addressed by T-Division were the behavior of the propellant and the neutronics associated with both predetonation (commencement of energy release before complete assembly) and the techniques required to give controlled initiation at the optimal time. Other issues were to determine the optimal height above the surface of the Earth for detonation and the nature of expected damage. These and many other issues were resolved by the Laboratory with sufficient confidence that this type of weapon was used at Hiroshima, where it produced the energy equivalent of fifteen thousand tons of high explosive.

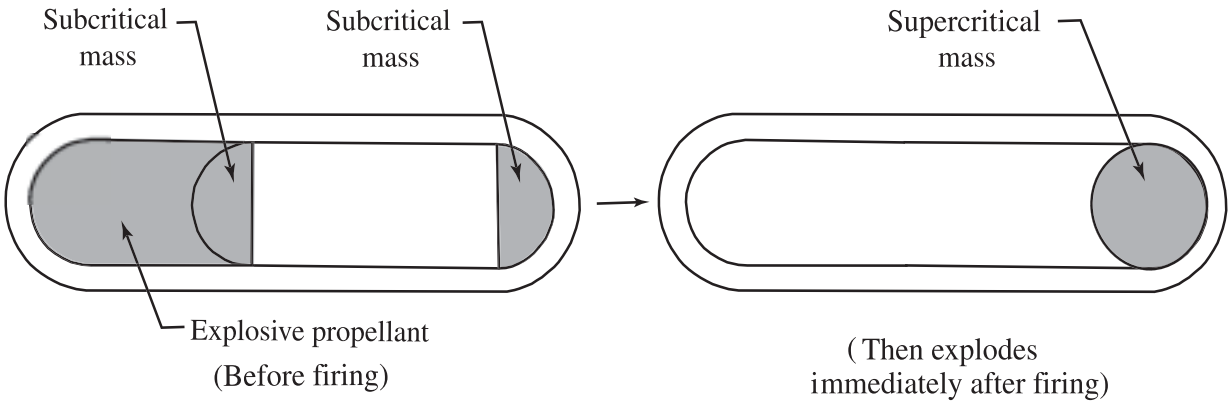


Fig. 3. Illustration of a gun-assembled weapon.

The Implosion Device

The basic principles of the implosion device are more complicated than those of the gun device (See Fig. 4.) A subcritical sphere of metal is surrounded by chunks of high explosive, which are simultaneously detonated in such a way as to maintain sphericity during compression to a supercritical configuration.

In this type of device, the issues of explosive behavior and shock transit through the metal, coupled to the simultaneous neutronics, required extensive tedious calculations to resolve. At this time, the calculations were all accomplished by hand. (See *Computers and Computational Science* in this book). Only later came the development of powerful computers, capable of giving answers to problems of this type in a few hours instead of many months.

In a variation of this implosion device, the high-explosive energy is transmitted to a hollow spherical annulus of the metal. It was believed that much greater levels of criticality would be attainable by this procedure. The overwhelming disadvantage of this approach is the tendency for the initially spherical shell to develop hydrodynamic instabilities that could preclude the spherical implosion deemed necessary for achieving the desired level of efficiency. T-Division tried hard to produce calculations that would calm this fear. Cylindrical implosions were examined theoretically and experimentally using a variety of metals, in order to determine the circumstances of disastrous instability in this simpler configuration to measure. Eventually the idea of imploding a hollow spherical annulus was considered too risky to pursue on the time scale that was required. Instead, virtually the entire Laboratory concentrated its energy on developing the solid-sphere implosion ideas, containing an initiator that commenced the fissioning process at the optimal time in the implosion.

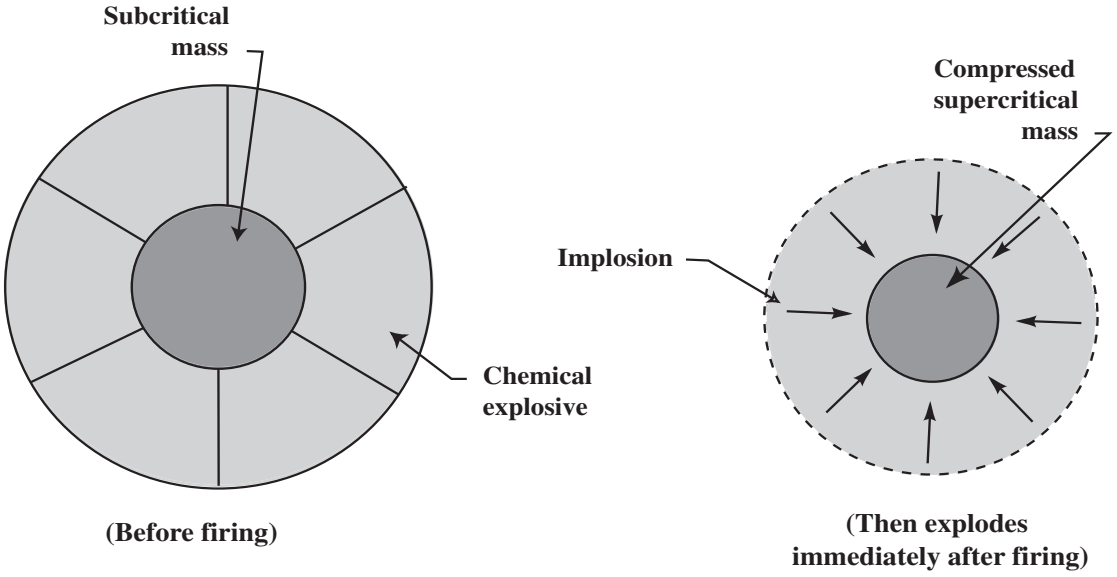


Fig. 4. Illustration of an implosion-assembled weapon.

By the summer of 1945, the confidence for success of solid-sphere implosion was great enough to indicate that a test would be appropriate. Named Trinity, the test was carried out in southern New Mexico. With an energy production equivalent to twenty-one thousand tons of high explosive, the test was considered successful. Two closely similar bombs were exploded over Hiroshima and, a few days later, over Nagasaki the following month, producing about the same amount of energy as the Trinity experiment.

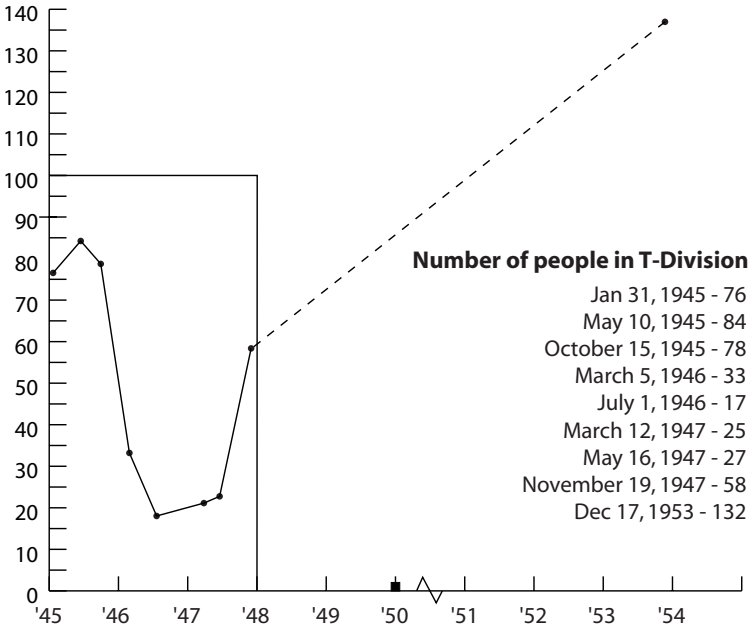


Fig. 5. Change in number of people at T-Division in early years.

Boosting and the Super

To many of the Laboratory personnel, the end of the Second World War in August 1945 signified the end of their assignment to Los Alamos. Over the next year, the staff of T-Division dropped from an average of about 80 people to an average of about 20. (See Fig. 5.)

But there was a large amount of work still to be done, which would result in major contributions to the U.S.'s national defense for the next 50 years (the collapse of the Soviet Union) and beyond.

The central theme of this additional work was associated with a completely different approach to the production of nuclear energy. The focus during the war years was primarily on the extraction of *fission* energy, in which the nucleus of a very heavy plutonium or uranium atom was fragmented by neutron bombardment, with the release of more neutrons (to give a runaway chain reaction) and large amounts of energy. The alternative process, called thermonuclear *fusion*, occurs when the nuclei of hydrogen atoms (especially the isotopes called deuterium and tritium) combine together to form the nucleus of an atom of helium, together with enormous amounts of energy and a very energetic neutron. This is the dominant reaction occurring in the sun and requires that the materials be compressed and heated to densities and temperatures never before attained on Earth.

The key to developing thermonuclear fusion lay in the capability of fission energy to provide the mechanism for compressing and heating the hydrogen isotopes. As an extension to the already developed fission bombs of the war years, the first exploration of this potential capability occurred in the return to high-explosive implosion of a hollow metal shell of plutonium or uranium, filled with a deuterium-tritium mixture of gases. Compressed to a supercritical state, the metal produced enough energy to initiate the thermonuclear-fusion reaction, which flooded the metal with neutrons and thereby enhanced the fission

process with a high degree of efficiency. This process is called "boosting." Paul Robinson (*LA Science*, Winter/Spring 1983) writes:

> Since this process has no analogy within other realms of science, essentially all of the physics—including new mathematical methods—had to be created to provide theoretical understanding and design techniques. The science of nuclear weapon phenomena has always been a driving factor for larger and faster computers, and boost physics will continue to require major advances in computing hardware, as well as in physical theory, before a complete description from first principles will be possible. This statement is still accurate in spite of the fact that today Los Alamos has what is believed to be the largest scientific computing center in the world.

The second exploration of thermonuclear fusion actually started as early as 1942 but was not a major part of the Laboratory effort until the late 1940s. At that time the hypothetical device was called the Super. T-Division was much involved with the development of the concept until 1944, when all of the division's activities were focused exclusively on the problems associated with the development of a pure fission bomb. After the successful completion of that work, T-Division again turned to the thermonuclear-fusion concept for a large focus of its activities. In 1951, a new insight was developed regarding the technique that should be employed to make the Super actually work. (See Fig. 6.)

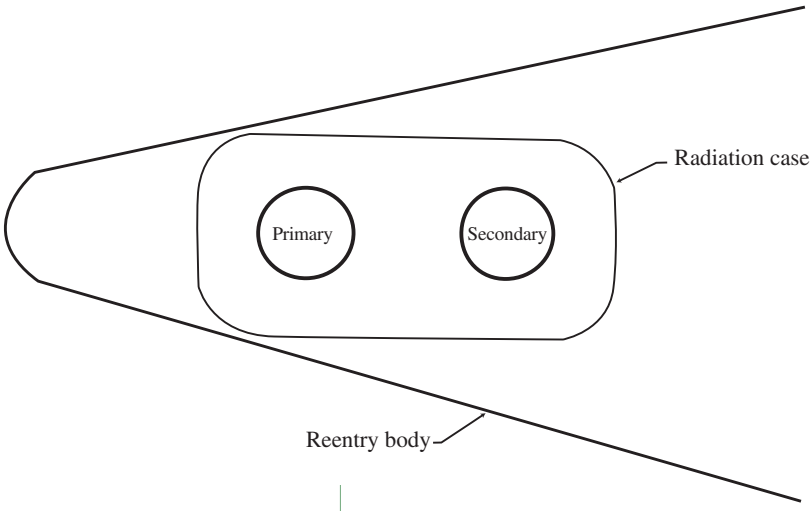


Fig. 6. The two-stage device.

Instead of achieving fusion within the fission bomb itself, the idea was to place the thermonuclear fuel in an adjacent secondary, there to be compressed to the required fusion conditions by means of the contained radiation from the fission bomb, called the primary.

In T-Division and indeed through many parts of the Laboratory, the two-stage device became a major challenge for development. In addition to the material-dynamics and neutronic issues, the Laboratory faced another major theoretical challenge. Paul Robinson's assessment of twenty years ago (*LA Science*, Winter/Spring 1983) is as relevant today as it was at that time:

> Radiation fields of such magnitudes or characteristics had not been encountered in any other field of science, so the physical theories and methods had to be created. Similarly, full mathematical description is limited by the computing power of today's best scientific computers.

Nevertheless, the Laboratory soon was able to overcome the problem to a sufficient degree to warrant a full-scale program for development of the two-stage device. As described in the paper by Mark, Hunter, and Wechsler (*LA Science*, Winter/Spring 1983):

> In an impressively short time, considering the amount and novelty of the design work and engineering required, the Mike shot, with a yield of about 10 megatons, was conducted in the Pacific on November 1, 1952.
>
> As tested, Mike was not a usable weapon: it was quite large and heavy, and its thermonuclear fuel, liquid deuterium, required a refrigeration plant of great bulk and complexity. Nevertheless, its performance simply confirmed the validity of the new approach. In the spring of 1954, a number of devices using the new pattern were tested, including the largest nuclear explosion (about 15 megatons) ever conducted by the U.S. Some of these devices were readily adaptable (and adopted) for use in the stockpile.
>
> Since 1954 a large number of thermonuclear tests have been carried out combining and improving the features first demonstrated in the Mike shot. The continuing objective has been weapons of smaller size and weight, of improved efficiency, more convenient and safe in handling and delivery, and more specifically adapted to the needs of new missiles and carriers.

T-Division continued to play a major role in nuclear weapons design until 1971, when the activities had enlarged to the extent that a new Laboratory division (Nuclear Weapons Analysis [TD] Division) was formed to carry on this work. Much supporting research, however, has remained in T-Division, including fundamental studies of material properties, neutronics, nuclear physics, atomic physics, material dynamics (especially instabilities, turbulence, and the mixing of materials), and numerical solution techniques. The Laboratory's computers were also a T-Division responsibility until 1968, when the task became so overwhelming that a new Computing Division (C-Division) was formed.

Laboratory Fusion

A challenge that is closely related to the achievement of thermonuclear fusion in bombs is the attainment of fusion in a laboratory setting. As described by Paul Robinson, the Inertial Confinement Fusion (ICF) program

> seeks to utilize laser radiation or ion beams to compress and heat fusion fuels to achieve thermonuclear reactions in the laboratory. Although directed toward different objectives, inertial fusion experiments embody significant aspects of

> radiation-transport physics. However, since the scale is significantly reduced, the physical characteristics of the processes and the resulting "micro-explosions" do not precisely replicate the phenomena exhibited in nuclear explosions. Nevertheless, the similarities have brought new approaches and techniques, and indeed new scientists, to explore these complex phenomena.

This program resulted in continuing and active collaboration with laboratories in many parts of the world. The problems have turned out to be excruciatingly difficult. At the present time, the new National Ignition Facility at Lawrence Livermore National Laboratory may be powerful enough to attain this elusive goal.

Conventional Munitions

As described by Paul Robinson in *LA Science*, 40th Anniversary issue, Winter/Spring 1983:

> Although the Laboratory has always devoted some effort to the development of nonnuclear weapons, we have expanded these activities in response to ever-widening threats. The subjects of improved lightweight armors and armor penetrators have substantially benefited from our research and innovation.

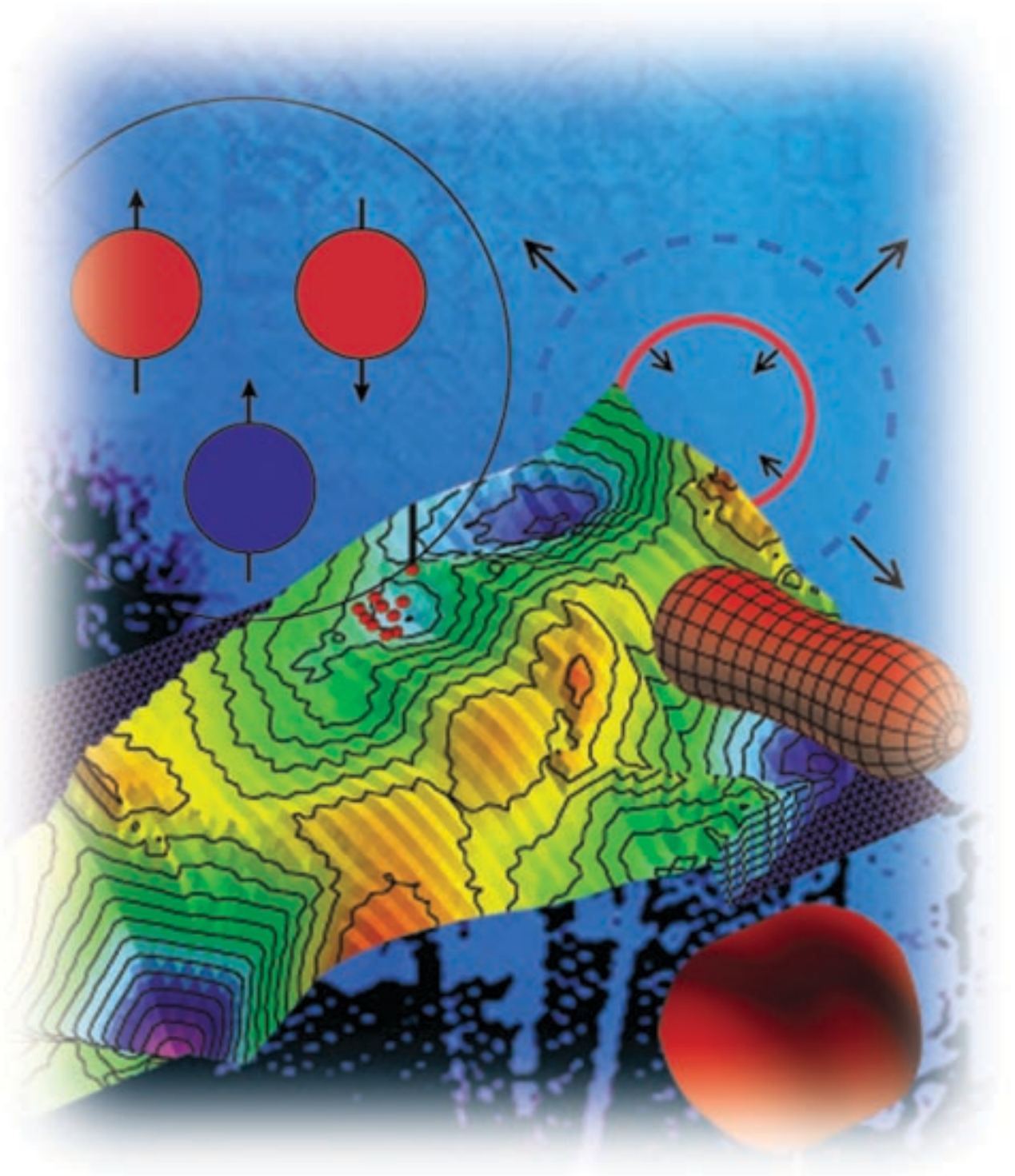
A major T-Division role has been to investigate the response of materials to strong impacts in a collaboration between the Department of Energy and the Department of Defense.

T-Division Involvement

The Laboratory, with T-Division playing a major role, was created with the exclusive goal of helping the United States win World War II. With that accomplished, the Division and the Laboratory could have closed. Instead, as shown by the sampling of subsequent activities in this book, the tremendous discoveries of new technology have been continuously expanded in numerous ways for solving major problems of national concern.

I naturally dealt a lot with the British group—Chadwick, Peierls, Tuck, Penny. The man who eventually turned out to be a spy, Fuchs, managed to put himself into a very important position. He was a member of the Theoretical Physics Division, like Professor Hirschfelder, but he also arranged to be appointed liaison between Theoretical Physics and X Division, so that he sat in on all our discussions and planning meetings.

—George B. Kistiakowsky [chief or the Explosives Division in World War II Los Alamos], "Reminiscences of Wartime Los Alamos," in Reminiscences of Los Alamos: 1943–1945, p. 60.



Introduction to Nuclear Physics Research in Theoretical Division

by Mark B. Chadwick

Since the Manhattan project, Los Alamos has been known throughout the scientific community as a leader in nuclear physics research. This research has spanned both applied and basic science. The applied work has supported a range of technologies, including nuclear weapon design, weapons diagnostics, reactors, accelerator-driven systems, medical physics, astrophysics, and nuclear nonproliferation.

To support these programmatic needs, nuclear reaction and structure theories have been developed, together with the codes that implement these theories. Particular scientific strengths that have developed are nuclear properties and reaction scattering for actinides (fissile and fissionable materials), light nuclei (thermonuclear reactions), and medium-mass nuclei (important in radiochemical weapons diagnostics). These tools were used to develop nuclear cross section and decay evaluated nuclear data files (ENDFs)—that integrate experimental data and sophisticated theory predictions—and underpin the Laboratory's simulation codes. The ENDF databases undergo continual improvement to respond to emerging needs in nuclear science. Their development has taken place in concert with the Laboratory's simulation codes that model nuclear particle transport.

The Laboratory has supported a very strong research group in basic nuclear theory. As noted elsewhere in this book in the article "Catching the Free Neutrino" in the "Particle Physics" chapter, the neutrino was first discovered in an experiment conducted by Reines, a member of T-Division. Since that time the Laboratory has played a leadership role in neutrino and weak interaction physics, both experimentally and theoretically, culminating in recent years in the liquid (large) scintillator neutron/neutrino detector measurement suggesting neutrino mass, and the landmark Sudbury Neutrino Observatory experiment that observed neutrino oscillation (and hence neutrino mass).

Following the construction of the Los Alamos Meson Physics Facility in the 1970s there has been a focus on medium energy nuclear physics. The theory group is known for its work in light-nucleus nuclear structure where methods have been developed to predict nuclear properties based on a realistic nucleon-nucleon force. The theory group also studies physics beyond the Standard Model, the role of symmetries in understanding hadron structure, heavy-ion reactions, effective field theory, and quantum chromodynamics.

Preequilibrium Nuclear Reactions

Arthur
Barschall
Chadwick
Feshbach
Griffin
Kerman
Koonin
Young

by Mark B. Chadwick

Preequilibrium nuclear reactions become important when the incident particle energy exceeds about 10 MeV, and therefore are important in reactions involving fusion 14 MeV neutrons. They play a crucial role in Laboratory nuclear technologies such as weapons, and accelerator-driven systems developed for nuclear waste transmutation and for neutron and proton cancer therapy.

When modeling nuclear reactions important in weapons, including reactions on actinide fuels and on radiochemical detectors, preequilibrium processes have been important for accurately predicting the magnitude of (n,2n) cross sections. Because (n,2n) reactions on radiochemical detectors are a diagnostic for 14 MeV fusion neutrons, the GNASH nuclear modeling code has been instrumental in underpinning nuclear cross-section evaluations supplied to Chemistry Division and to Applied Physics Division.

Preequilibrium nuclear reaction theory describes processes in which a projectile particle incident on a target nucleus initiates a multistep scattering, where its energy is shared among nucleons in the target via nucleon-nucleon scattering, and particle emission may occur prior to equilibration. These preequilibrium particles are typically of higher energy than those emitted from an equilibrated compound nucleus and are emitted in the forward direction.

Preequilibrium nuclear reactions become important when the incident particle energy exceeds about 10 MeV, and therefore are important in reactions involving fusion 14 MeV neutrons. They play a crucial role in Laboratory nuclear technologies such as weapons, and accelerator-driven systems developed for nuclear waste transmutation (see the article "The Accelerator Transmutation of Waste Effort" in the "Environmental Studies" chapter in this book) and for neutron and proton cancer therapy.

Interestingly, the existence of preequilibrium reactions was first discovered in 1966 by a Los Alamos theorist, Jim Griffin, who proposed a semiclassical model to understand them. The first experimental signature of preequilibrium scattering, analyzed by Griffin, was made by Heinz Barschall—another Los Alamos scientist who was a young researcher during the Manhattan project, after which he moved to Wisconsin. In 1980, preequilibrium physics was advanced when Arthur Kerman—a long-time visitor and consultant to Los Alamos' T-Division—presented a quantum mechanical formalism for representing these processes with his coauthors from MIT, Feshbach and Koonin.

Both semiclassical and quantum preequilibrium formalisms have been implemented in Los Alamos nuclear reaction modeling codes, for predicting ejectile emission spectra in nuclear reactions. The GNASH nuclear modeling code utilizes these theories, together with extensions to the original quantum formalism to expand its applicability. It also implements other relevant processes, such as Hauser-Feshbach compound nucleus decay and fission, allowing the code to be used in a wide range of applications that require accurate predictions of nuclear cross sections. Preequilibrium models also contributed to the high predictive capability attributed to Los Alamos intranuclear cascade codes LAHET and CEM, which are used at higher energies.

When modeling nuclear reactions important in weapons, including reactions on actinide fuels and on radiochemical detectors, preequilibrium processes have been important for accurately predicting the magnitude of (n,2n) cross sections. Since (n,2n) reactions on radiochemical detectors are a diagnostic for 14 MeV fusion neutrons, the GNASH code has been instrumental in underpinning nuclear cross-section evaluations supplied to Chemistry Division and to Applied Physics Division. Work in T-Division in these areas by Phil Young and by Ed Arthur (now in the Laboratory's Associate Director for Strategic Research Directorate) is particularly noteworthy.

A Unified Theory of Nuclear Masses, Fission Barriers, and Superheavy Nuclei

Bethe
Bolsterli
Chadwick
Fiset
Frankel
Kodama
Krappe
Kratz
Metropolis
Möller
Myers
Nazareth
Nilsson
Nix
Norton
Sierk
Strutinsky
Swiatecki
Treiner

by Peter Möller and Arnold J. Sierk

This more-than-30-year effort has had a major impact on the world of nuclear physics. Los Alamos research has provided some of the earliest and most complete estimates of lifetimes and decay modes of superheavy nuclei, an extremely successful generalization of the liquid-drop model, and a nuclear mass model with the best extrapolation properties of any such model.

In the 10 years since the last version was completed, it has not been equaled for predictability of nuclear masses and deformations of nuclei measured subsequent to its development. Recent fission calculations that computationally were one million times larger than 34 years ago described in a single consistent calculation features such as barrier heights throughout the periodic system, the several modes of fission for nuclei near ^{228}Ra and ^{258}Fm, and the mean masses of the heavy and light fragments in asymmetric fission.

Some Early Discoveries in Nuclear Physics

Nuclear radioactivity was discovered in 1896 when Henri Becquerel noticed that a photographic plate that had been left in a drawer together with a uranium compound had been clouded, although it had been completely in the dark. He soon established that the uranium compound emitted unknown rays, given the name α rays, which were able to penetrate the cover of the photographic plate. The α rays were soon demonstrated to have a short range, that is they could be stopped with a relatively small amount of matter. Two other types of radioactivity, β rays having an intermediate range, and γ rays with quite a long range were discovered soon after.

It was later established that these α rays were identical to the nucleus of a helium atom, β rays were electrons or positrons, and γ rays were electromagnetic quanta. At that time, the structure of the helium nucleus was not understood. This was clarified after James Chadwick's discovery of the neutron in 1933; this discovery was delayed by decades because the uncharged neutron has very little electromagnetic interaction with matter, and was much harder to detect with the techniques available at the time. It was only after the discovery of the neutron that the picture of the atom was completed: an atom consists of a nucleus of protons and neutrons with electrons moving around this nucleus at very large distances compared to the size of the nucleus. Each element in the periodic table has a unique number of protons. Several *isotopes* of each element exist, each isotope with a different neutron number. Because of the nature of the nuclear force only *a few* stable isotopes exist of each element. Above the element bismuth with 83 protons,

no completely stable isotopes exist for any element. And there are two elements below Bi for which no stable isotopes exist either.

Before the discovery of the neutron, artificial transmutations of one element into another had been studied by bombarding nuclei with α rays. Immediately after the discovery of the neutron, a new type of transmutation in which heavy-element targets were bombarded with neutrons became possible and many such measurements were actively pursued. In particular, uranium (U), which with 92 protons is the heaviest element in the periodic system that occurs naturally in quantity, was the target of choice. The idea was to form a new more neutron-rich isotope of uranium, namely uranium-239, which would be unstable with respect to beta decay and decay to an element with 93 protons and 146 neutrons. Unexpectedly, the products of such reactions were found to yield a large number of different substances with different half-lives. Elaborate schemes were proposed to explain the observations, none very satisfactory. It took about 6 years before the correct interpretation was deduced following an experiment in Berlin by Hahn and Strassmann that shook the world. They established in December 1938 after careful chemical analyses that isotopes of the element barium with 56 protons in its nucleus were present in the reaction products. They had no explanation for this observation, but a theoretical interpretation followed almost immediately. Before discussing this, we will give a little background.

Early Nuclear-Structure Theory

An important tool for understanding many nuclear properties, in particular α– and β- decay properties, is the "liquid-drop" model of nuclear masses. In its earliest incarnation, which appeared in the mid 1930s, it went by the name of the Bethe-Weizäcker formula. Because this model was very important, is relatively simple, and because later developments in T-Division built on this model, we discuss it briefly here. The nuclear mass is approximately given by:

$$M(Z,N) = M_{\mathrm{H}}Z + M_{\mathrm{n}}N + a_{\mathrm{C}}Z^2/A^{1/3} - a_{\mathrm{v}}A + a_{\mathrm{s}}A^{2/3} + a_I(N-Z)^2/A \quad ,$$

which is valid for spherical nuclei. As is customary in nuclear physics, since Einstein allows us to convert between energy and mass, we will use these terms interchangeably. Normally the nuclear mass is given as an energy. The four unknown model constants a_{C}, a_{v}, a_{s}, and a_{I} in the above equation are determined by using the above expression to calculate the masses of the more than 1,000 known nuclei for different values of these constants and determining the set of values that yield optimum agreement between the model results and the measured masses of nuclei. There exist mathematical approaches that ensure that one finds the very best set of constant values so that there is no possibility that subsequently somebody by trial and error can find a set of constant values that will give better agreement with data.

We now interpret this simple formula. The mass of the neutron is M_{n}, the mass of the hydrogen atom (a proton plus an electron) is M_{H}, the number of protons in the nucleus for which we calculate the mass is Z, and the number of neutrons is N. Thus, a first try to get the mass of an atom would be to sum up the masses of the N neutrons and the Z protons and electrons, which would be given by the first two terms in the above expression.

These two terms actually give the major component of the nuclear mass and would be the mass of the constituent particles of the atom if it were broken up into N neutrons and Z hydrogen atoms. However, because like charges repel each other, some work or energy is required to push the Z protons together into the small confines of an atomic nucleus. This energy is equivalent to mass and is given by the third term in the mass expression. The constant a_{C} can be calculated if we determine the average distance between the protons in a nucleus, but in practice this constant and the three other *model constants* are determined by adjustments to data as discussed above.

Apart from the electrostatic Coulomb force, there also acts an attractive *nuclear force* among the neutrons and protons in the nucleus. To leading order, it does not differentiate between protons and neutrons. In contrast to the electromagnetic force in which the force between two charges Q_1 and Q_2 separated by a distance r is proportional to Q_1Q_2/r^2 and repulsive, the nuclear force is attractive and very strong at small distances, but for distances slightly larger than the "size" of a proton or neutron, it goes very rapidly to zero. One can show that because of the presence of this force, there is a contribution to the mass given by the fourth term above, proportional to the total number of nucleons A and with a minus sign, because the force is attractive.

Because the nucleons (the name given to either a proton or a neutron) are held together by the nuclear force, energy, or equivalently mass, has to be added to the nucleus to break it into its constituent protons and neutrons. Very simply, a nucleon attracts just the nucleons that immediately surround it. Each such cluster gives the same (negative) contribution to the mass, and there are A such clusters. However, nucleons at the surface of the nucleus find no nucleons outside the surface to attract, so the contribution of the nucleons at the surface does not give the full (negative) cluster contribution to the nuclear mass. Since all nuclei have approximately the same density, the volume is proportional to A and the surface area is proportional to $A^{2/3}$, which leads to the fifth term, the surface-energy term.

Finally, because of special laws governing the microscopic world, one can show that the energetically most favorable number of protons and neutrons in a nucleus, in the absence of an electrostatic Coulomb force, is an equal number of protons and neutrons, which gives rise to the last term above known as the symmetry-energy term. In summary, for nuclei that exist on Earth, the balance between the terms discussed

above is such that energy (mass) has to be added to break the nucleus into its constituents. Expressed in another way: The mass of a nucleus is smaller than the sum of the masses of its constituents.

In the 1930s, this very simple model was used to understand many nuclear properties, for example radioactive decay, which tends to transmute systems of high energy to systems of lower energy. Sometimes an energy barrier is present in the initial phases of such a transmutation and prevents it, although the end configuration is of lower energy. However, in the microscopic world, the strange but well understood laws of quantum mechanics permit such barriers to be *penetrated* with a certain probability that depends on, for example, the height of the barrier and the mass of the penetrating particle. This process gives rise to radioactive decay.

The *half-life* of the decay, which is the time it takes half of the original nuclei to decay, can be calculated once important issues governing the decay have been sufficiently understood. Even before the discovery of fission, scientists realized from simple arguments based on Eq. (1) that enormous amounts of energy were locked up in heavy nuclei. With the liquid-drop formula one can, for example, calculate that if a uranium atom is split into two equal-size fragments then the sum of the fragment masses is smaller than that of the original uranium atom due to the third term in Eq. (1). The difference is released as energy. And the energy released in this division of the nucleus is enormously much larger than when, say, one coal atom burns in a chemical reaction, about 100,000,000 times larger, in fact. But scientists originally thought there was no way to release this huge amount of trapped energy.

Explanation and Consequences of Fission

When the results of the Berlin experiment were communicated to Meitner and Frisch, colleagues of Hahn and Strassmann who were by then in exile in Sweden, Meitner and Frisch immediately realized that the atomic nucleus had been split. In a famous paper published in *Nature* on February 11, 1939, they speak of the "surface tension of a charged liquid drop" and further state: "It seems therefore possible that the uranium nucleus has only a small stability of form, and may, after neutron capture divide itself into two nuclei … These two nuclei will repel each other and should gain a total kinetic energy of 200 MeV." Then in another paper Frisch described a direct measurement of the kinetic energies of the fission fragments, thus providing independent supporting evidence of the fission process.

A few months later Bohr and Wheeler in their seminal paper, "The Mechanism of Nuclear Fission," put the description of fission on a firm theoretical basis. The basis of their discussion is Eq. (1). However, they modified it to apply to deformed shapes. They realized that the energy required to pack charged particles into a deformed shape is different from the energy required to pack them into a spherical shape, which means that term 3 in Eq. (1) has to be modified. Because a deformed

nucleus has more nucleons in the surface region than a spherical nucleus, term 5, the surface-energy correction to the volume energy also has to be modified for deformed nuclei. Thus Bohr and Wheeler wrote

$$M(Z,N) = M_H Z + M_n N + a_C Z^2 / A^{1/3} B_C(\text{shape}) - a_v A + a_s A^{2/3} B_S(\text{shape}) + a_I (N-Z)^2 / A \quad . \qquad \text{(Eq. 1)}$$

B_C and B_S are dimensionless functions both having the value 1 for a sphere.

This analysis assumes that the nuclear density remains constant during deformation, just as is the case for a drop of water. By use of this liquid-drop model of deformed nuclei, one can calculate the energy of a nucleus as it deforms from a spherical shape through intermediate shapes to separated fragments. In 1939, before the age of the computer, it was only possible to evaluate this expression for small deviations from a spherical shape.

But it could be shown that as a nucleus deforms from a spherical shape the energy initially rises then decreases as the deformation increases. Thus, the nuclear energy curve versus deformation, the fission barrier, has a minimum for a spherical shape, then a maximum followed by a steep decline in energy as the system evolves towards separated fragments. It could further be shown that the fission barrier height would gradually decrease as the charge of the system increases, and that the barrier would disappear altogether at around proton number $Z = 100$. For the isotope U-235, the height of the barrier is just low enough that the energy imparted to the system by capturing a neutron will raise the energy of the nucleus just sufficiently for the nucleus to fission immediately.

It was also discovered that two or three neutrons are emitted in connection with the fission of one uranium nucleus. It was immediately realized that these neutrons could initiate new fission reactions and that therefore in uranium a self-sustaining chain reaction might be possible, provided that the amount of uranium was large enough so that on the average more than one of the emitted neutrons were captured by nuclei and initiated fission reactions instead of escaping through the surface of the volume of uranium. Such a chain reaction would make it possible to release the vast amounts of electrostatic energy locked up in uranium nuclei.

Founding of Los Alamos Scientific Laboratory

It was realized already in 1939 that self-sustaining chain reactions in uranium might make possible a new and extremely powerful explosive device, and many scientists were increasingly concerned about this prospect. On August 2, 1939, Albert Einstein signed a letter to President Roosevelt warning of this possibility (see the “Introduction” chapter in this book). The letter specifically mentioned that such an explosive

device on a boat might blow up an entire port. It also advised the president to fund further research into uranium and that he make arrangements so that the government could closely follow developments. After further research had made it increasingly plausible that an explosive device sufficiently small to be used as an instrument of war could be constructed by the first half of 1945, President Roosevelt on December 28, 1942, gave the go-ahead for such a project. Los Alamos Scientific Laboratory, as it was then called, was established immediately afterwards, in early 1943. It is of interest to note that Hans Bethe, whose name is associated with Eq. (1) above, became the first leader of the Theoretical Division. Two and a half years after the foundation of the Laboratory, the first atomic bombs had been manufactured, tested, and used to expedite the end of the war.

Nuclear Science Developments, 1945–1970

In one of the first applications of the new electronic computer ENIAC, former Los Alamos scientists Frankel and Metropolis in 1947 explored the liquid-drop model of fissioning nuclei in a way unimaginable to Bohr and Wheeler only eight years earlier. In particular, the computer made it possible to numerically determine very accurately fission-barrier heights and the corresponding shapes of nuclei as given by this model. Metropolis would soon afterwards return and participate in the construction of the first electronic computer at Los Alamos, the MANIAC.

In atoms, properties of the electronic orbitals around the central nucleus could already at this time be calculated fairly accurately by use of the laws of quantum mechanics that govern the behavior of electrons, with the force exerted on the electrons by the central nucleus given by the purely electric attraction between opposite charges. The solution of these equations shows that certain chemical elements, the noble gases, are unusually stable and nonreactive, due to particularly stable electron configurations, which occur for 2, 10, 18, 36, 54, and 86 electrons, corresponding to helium, neon, argon, krypton, xenon, and radon.

At these particularly stable numbers, which are called magic numbers, large gaps occur in the energy-level spectrum corresponding to the individual electron orbits. It was long thought that in nuclei a similar picture could not apply because all nucleons move close together, more like in a liquid, and the short-range, very complex nuclear force is very different from the long-range electromagnetic interaction. In addition, it is impossible to solve the microscopic quantum-mechanical equations that result from an implementation of the full-fledged nucleon-nucleon force. However, experimental evidence gave increasing evidence that for nuclei, particularly stable particle configurations also exist, in this case, particularly stable proton and/or neutron configurations. For example, for such neutron and proton numbers, the nucleus has a mass that is up to 10 MeV lower (indicating a more strongly bound system) than given by the liquid-drop mass expression above.

Unlike Oppenheimer or Bohr, Bethe was a jolly fellow. My office was down the hall, and I could always tell when Bethe and Feynman were working together because there would be frequent bursts of guffaws and belly laughs. When there was a comic moving picture in the base theater, you could always tell by the laughter whether Bethe was in the audience.

—Joseph O. Hirschfelder [former group leader in T-Division], "The Scientific and Technological Miracle at Los Alamos," in Reminiscences of Los Alamos: 1943–1945, p. °81.

It turned out that many such effects "beyond the liquid-drop model" were described to a high accuracy in a model in which the protons and neutrons are considered to move independently of each other in a potential originating from the center of the nucleus. This is very surprising, because the nucleus has no obvious fixed center that could give rise to such a force, in contrast to the situation with the electron orbitals in the atom, in which case the nucleus is the center from which the force on the electrons originates. That a similar model works for nuclei was unexpected, and this *single-particle model* of nuclei emerged only in the 1950s.

The simple force that is used in the model that describes the motion of nucleons is called an *effective force*. When a physicist tries to solve a complex problem a common approach is to replace real forces, for which the equations would be unsolvable, with effective forces for which the equations can actually be solved. Ideally the effective forces should be such that the solutions that are obtained retain all essential features of the solutions corresponding to the real forces, if those solutions could be obtained. Unfortunately there is no well-defined rule of how to find effective forces that are useful in practice; that is, forces that well describe the system under study. They are found by insight, intuition, physics skills, and trial and error. The single-particle model of nuclei is an implementation of an enormously successful effective-force approximation to the true nuclear forces. Because of differences between the purely electromagnetic forces determining the electron levels and the nuclear forces, the first few magic numbers in spherical nuclei are 2, 8, 20, 28, 50, and 82 for both protons and neutrons.

In the mid-1950s there were two important developments. Sven-Gösta Nilsson extended the single-particle model to deformed nuclei and also solved the resulting equations on a computer [1]. At this time it had become clear from measurements of the charge field around nuclei that most nuclei are deformed, like an American football. Nilsson's deformed model correctly described the energy levels observed in such nuclei. At the same time Wladek Swiatecki used fission barriers obtained in the liquid-drop model to calculate fission half-lives for isotopes of plutonium, curium, californium, and fermium with proton numbers 94, 96, 98, and 100, respectively [2]. Spontaneous-fission half-life calculations based on these barriers did not agree with the observed half-lives. But Swiatecki noted that the ground-state masses of these isotopes were also not perfectly given by the liquid-drop model. For the barrier minimum, or ground-state energy, he therefore used the experimental mass/energy value, which sometimes changed the ground-state mass by several MeV. Half-life calculations based on these barriers agreed quite well with observed half-lives.

In the mid 1960s, a development of enormous importance occurred. Vilen Strutinsky observed, as others had done, that large gaps in the quantized energy levels calculated for spherical nuclei were strongly correlated with extra binding of nuclear ground-state masses, over and above what was given in the liquid-drop model. But he also observed

that similar gaps often occur for deformed shapes for specific nucleon numbers, and he postulated that extra binding was also associated with these gaps. And he provided a method, the "Strutinsky method" to calculate the effect of the microscopic structure, seen in calculated quantum levels, on the nuclear mass or energy for any deformation and any particle number.

Strutinsky's method gave an actual number, a "shell-correction" energy to the liquid-drop model for the nuclear energy. Thus, the total energy of a nucleus of a particular shape and particular proton and neutron numbers was obtained by (1) specifying the shape, (2) calculating the macroscopic liquid-drop-model energy for this shape, (3) calculating the quantum-mechanical energy levels for this shape, and (4) using these levels to calculate a "shell correction" by use of Strutinsky's method. The total nuclear energy for this particular shape is then obtained as a sum of the shell-correction and liquid-drop-model energies. This combination of a macroscopic model with shell corrections obtained from nucleon energy levels is called the macroscopic-microscopic method. Strutinsky used his model to explain fission-isomeric states that had recently been observed in some very heavy elements. These were states that fissioned, but with half-lives in the millisecond range, compared to billions of years for the normal spontaneous fission half-life of the same nucleus. In the liquid-drop model, the ground state is spherical, and the energy varies smoothly as the nucleus passes over the fission barrier. In contrast to this situation, Strutinsky found that the shape-dependent shell-correction energy, which fluctuates significantly over much smaller deformation ranges than does the liquid-drop energy, leads in the case of some heavy nuclei to a deformed ground state and a splitting of the macroscopic fission barrier into two peaks separated by a second local minimum. This became generally known as the "double-humped" fission barrier. The fission-isomeric states were interpreted to be nuclear states confined to the secondary minimum between the two barrier peaks. This interpretation was subsequently confirmed by detailed experiments and calculations.

At this time, the assumption that elements beyond uranium rapidly became more unstable to fission, and that the number of elements would end just above proton number $Z = 100$ was challenged by Myers and Swiatecki, who observed that the extra binding energy that would be associated with the next magic proton number beyond $Z = 82$, would be large enough so that nuclei would again be fairly long-lived in the vicinity of this magic number. Quantum-mechanical models suggested that the next magic proton number would be 114 and the next magic neutron number beyond 126 would be 184. A large number of groups all over the world then started to calculate the stability of these "superheavy" elements as they were called. In particular, during 1967–1968 at Lawrence Berkeley Laboratory, Wladek Swiatecki, Bill Myers, Ray Nix, and Sven-Gösta Nilsson of the Lund Institute of Technology in Sweden performed such a calculation, based on the liquid-drop model of Myers and Swiatecki and the Nilsson modified oscillator model [3,4].

The stage was now set for what would over the next 30 years become a sustained, systematic, and very successful effort in this field at Los Alamos.

The Los Alamos Macroscopic-Microscopic Model, 1970–2000

In early 1968 Ray Nix accepted a position with the Theoretical Division at Los Alamos Scientific Laboratory. He and his collaborators immediately started a program of systematic nuclear-structure investigations based on the macroscopic-microscopic model and of systematic improvements of the model. In the macroscopic-microscopic approach, as outlined in the previous section, it is first necessary to specify the shape of the nucleus. Nix introduced the three-quadratic-surface (3QS) shape parameterization, which was superior to any previous shape description for studies of highly deformed fissioning nuclei. With the 3QS, five independent degrees of freedom can be specified: (1) overall elongation of the fissioning system, (2) neck radius, (3) left-right mass asymmetry, (4) left nascent fragment spheroidal deformation, and (5) right nascent fragment spheroidal deformation. With the shape defined, it is then necessary to calculate the single-particle effective potential acting on the nucleons. Nix developed the folded-Yukawa potential which successfully generalized potentials with realistic surface properties from spheres to nuclei of arbitrary shape. With the shape and potential defined, the availability of efficient techniques for solving the Schroedinger equation made implementation of Strutinsky's method feasible for large-scale investigations of nuclear properties of elements throughout the periodic table.

The model is completely described in a 1972 paper [5]. This model was then immediately used to calculate the stability of a large number of nuclei in the region of the next magic numbers, which in this model as in the earlier one, were predicted to be $Z = 114$ and $N = 184$. In contrast to other more limited studies of this region, the Los Alamos calculation presented a table of all three important decay modes, namely fission, α decay, and β decay for all nuclei of interest in the superheavy region [6]. The same data were also provided for the decay chains of these superheavy elements. The people involved in this work during 1970–1972 were M. Bolsterli, E. O Fiset, J. R. Nix, and J. L. Norton, all of T-Division.

A new phase in the work on the macroscopic-microscopic model began in 1973. Hans-Jürgen Krappe from the Hahn-Meitner Institute in Berlin came to visit Lawrence Berkeley Laboratory and collaborated there with Wladek Swiatecki and with Ray Nix in Los Alamos on refinements of the surface-energy term of the liquid-drop model. The normal surface-energy model accounts for a reduction in the volume energy because there are no nucleons "outside" of the surface to be attracted. However, in some situations there are nucleons outside, sufficiently near to feel the nuclear force.

One such situation is when two colliding nuclei have approached each other so that they are almost touching. Then there is an attraction between the nuclei partially counteracting the repulsive force between the protons of the two nuclei. This effect makes it easier for the nuclei to merge in a collision. Another situation is when a fissioning nucleus has almost divided. Then the almost-formed fragments are connected by a thin region of protons and neutrons. Nucleons in one fragment near this region can interact across the surface with nucleons in the other region. These effects were now described in the *finite-range* surface-energy model [7]. The model was significantly improved in 1979 [8] by which time Arnold Sierk had joined the effort as a staff member of T-Division and H.-J. Krappe was visiting T-Division on a sabbatical.

Also in 1973 important work was carried out on the microscopic part of the model, when Sven-Gösta Nilsson visited Los Alamos on a sabbatical. He was accompanied by one of his students, Peter Möller, one of the authors of this paper. Together with Ray Nix, they made detailed studies of energy levels of deformed nuclei with the folded-Yukawa single-particle potential and compared model results with experimental data. This work led to a set of improved spin-orbit and diffuseness constants for the model [9]. This parameter set is still used today, 30 years later. The original set of constants had been determined from studies of levels in spherical nuclei only. The Nilsson perturbed-spheroid parameterization (ε-parameterization) was also incorporated into the folded-Yukawa macroscopic-microscopic model as an alternative to the 3QS parameterization, after Möller showed that for some nuclei the ground-state mass was more than 2 MeV lower in this parameterization. Möller also incorporated the new finite-range surface energy into the model. The three developments—(1) ε-parameterization, (2) the improved choice of spin-orbit and potential diffuseness constants, and (3) the finite-range model for the surface energy—all turned out to be very important for later work on nuclear mass models.

Möller and Nix used the model with its new optimized constants to calculate fission barriers of nuclei with proton number in the range $88 \leq Z \leq 108$. In such calculations, energies are calculated for a set of shapes that are thought to be relevant for the fission process. For example, energies are calculated for various shapes that lead from the ground-state shape to separated fragments. Of course, a nucleus can in principle divide in many different ways, for example one small fragment and one large fragment or into two fragments of equal size. To determine what is the energetically most favorable way to divide into two fragments, these calculations are carried out in a systematic manner. In this particular calculation, seven different elongations were investigated. For each elongation, five different neck radii were investigated. Five different left-right mass asymmetries were also investigated. This leads to 175 possible combinations or, equivalently 175 different shapes. In this calculation, the energy was consequently calculated for each of these 175 possible shapes for each nucleus investigated. In 1973, this was a substantial computation. The calculation showed the by-now well-known effect that the energetically most favorable path to separated

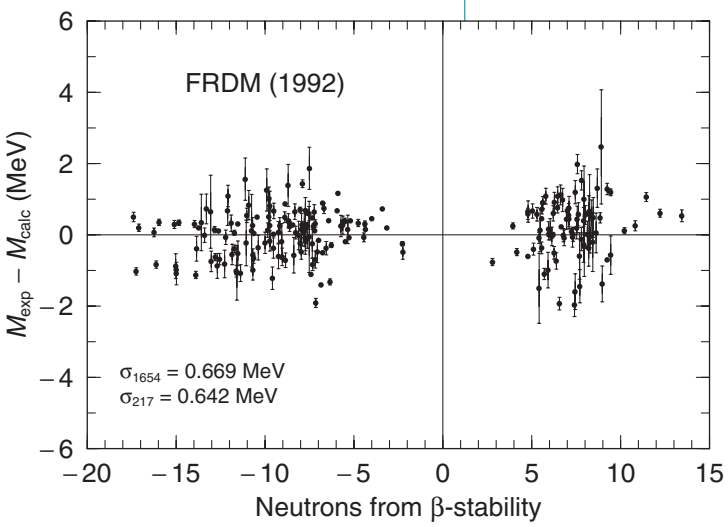


Fig. 1. Difference between measured and calculated masses for nuclei whose masses were not known when the model constants were determined. Because model errors often increase in regions far from where the model constants were determined, we plot the differences versus neutrons from stability. We see that there is no such trend of increasing errors in this mass model.

fragments involves nuclear shapes with left-right asymmetry (mass asymmetry).

In 1975–1977, Ray Nix, Möller (now as a postdoctoral visitor), and Sierk studied various topics of interest in heavy-element formation. For example, in one set of calculations the nature of the macroscopic potential-energy surface in fission and heavy-ion reactions was systematically studied for compound systems from beryllium-8 to $^{476}184$. In one case, for $^{300}120$ the calculation included microscopic corrections and showed that heavy-ion reactions with lead-208 give rise to a shell-stabilized cold-fusion valley [10,11]. Another interesting study was an exact calculation of the penetrability for a simple two-dimensional heavy-ion fusion barrier, in collaboration with T. Kodama and R. A. M. S. Nazareth of Centro Brasileiro de Pesquisas Físicas in Rio de Janeiro [12].

The most well-known work by Ray Nix and Peter Möller is their development of nuclear mass models based on the microscopic folded-Yukawa single-particle model and the macroscopic finite-range liquid-drop model. The first calculation in this series appeared in 1981 and covered 4,023 nuclei from oxygen-16 to $^{279}112$ [13,14]. The success of the work was in no small part due to the systematic and sustained development effort on the model during the previous 10 years. Some additional development work on the models resulted in further refinements. An interim calculation appeared in 1988 [15,16], and the current version in 1995 [17]. In the 1995 calculation, two versions of the mass model appear, one based on the previous finite-range liquid-drop model and one based on the droplet model of Myers and Swiatecki, now modified to account for the effect of the finite range of the nuclear force on the surface-energy term. These modifications were carried out in collaboration with W. D. Myers and W. J. Swiatecki of Lawrence Berkeley Laboratory, and with J. Treiner of Orsay. Some of the most noteworthy features of the 1995 mass model are that first, it covers all nuclei between the proton and neutron drip lines from oxygen-16 to $^{339}136$; second, for nuclei that were not known when the model parameters were determined, the average error remains about the same as for previously known nuclei. (See Fig. 1.)

Therefore, the model has been extremely useful in applications where nuclei whose properties have not been measured occur, for example, in astrophysical studies. A third feature of the calculation is that not only nuclear masses, but a substantial number of other nuclear-structure quantities, are obtained in the model—for example, nuclear ground-state shapes and quantum-mechanical energy levels. Properties of particular importance for astrophysical applications are given in a separate paper [18]. The calculation of these properties and the related astrophysical studies were undertaken in collaboration with K.-L. Kratz

of Mainz University. Several other highly cited studies related to the mass work have been also been carried out and presented in Refs. [19–22].

Most recently we have in the nuclear theory group revisited our old interest in nuclear fission. Over the years it had become increasingly clear that in fission studies of the shape evolution from the ground state to separated fragments, it is necessary to study many more possible shapes than the several hundred to about 1,000 that had been studied previously. Because in fission a nucleus may divide into fragments of different mass and different shapes, it is necessary to do calculations in which all five shape degrees of freedom that were enumerated above are varied independently.

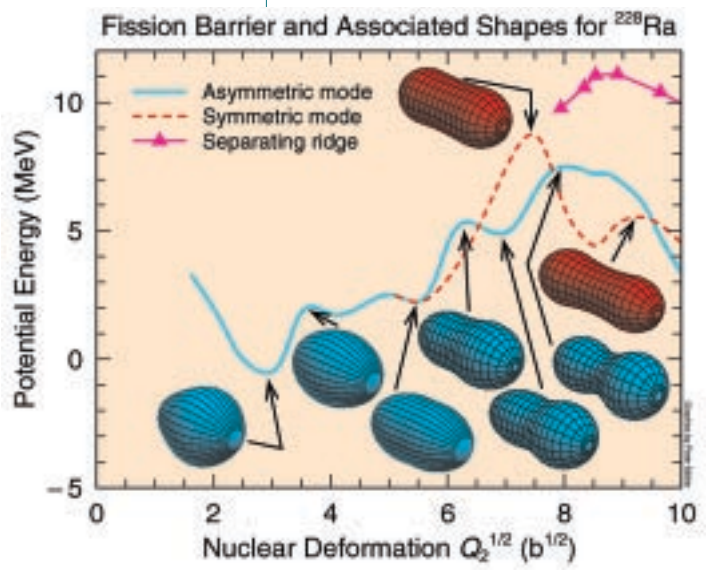


Fig. 2. When nuclei fission, they gradually change their shape from a ground-state shape to elongated shapes with an intended middle section and finally separate into two distinct fragments. In this process the energy initially increases up to a maximum *barrier height* and then starts to decrease. For Ra-228, two about equally probable types of division occur. One is into two fragments of different size, asymmetric fission. The shapes and energies associated with the initial part of this division are given by six blue nuclear shapes and an associated blue fission barrier. The possibility of division into fragments of equal size also exists. At a relatively early stage in the division process, the fission path divides and a symmetric path branches off from the asymmetric path. The shapes and energies associated with this type of fission are shown in red. The two paths are separated by a ridge shown in magenta. The calculated features shown in this figure agree very well with experimental observations.

To investigate all relevant combinations of these five degrees of freedom with a sufficiently close spacing of the shapes leads to about 3,000,000 shapes for which the nuclear energy must be calculated. This was previously impossible to do in a reasonable amount of computer time, since it would require more than a year of dedicated computer time on one of the early CRAY computers. However, recently two developments have made such a calculation feasible, namely (1) the steady increase in individual computer speed and (2) the availability of clusters of computers with a hundred or more processors. One can then distribute the shape configurations for which the energies need to be calculated to the different computers in the cluster and obtain the results in days instead of years.

In 1999 we performed such a calculation on the Los Alamos Avalon cluster of 140 separate CPUs. Once the 5-D energy landscape has been calculated, a nontrivial problem is how to identify the significant saddle points and other structures of importance in this energy landscape. Here we borrow a technique used in topographical studies in geography, namely imaginary water-immersion techniques. Our study solves several outstanding problems in fission and was published as an article in *Nature* [23]. We show some results of the investigation in Figs. 2 and 3. Figure 2 shows fission barriers associated with two observed modes, symmetric and asymmetric, observed in the fission of nuclei near radium-228.

Our calculated barrier heights and the nuclear shapes associated with the two fission modes agree with conclusions from experiments. The experiments also indicate that the two modes are "well separated" between saddle and scission. This is in agreement with the high ridge obtained in our calculation. By following the asymmetric mode until close to the point where the single nucleus divides into two fragments, we can determine the masses of these fragments, which are normally of

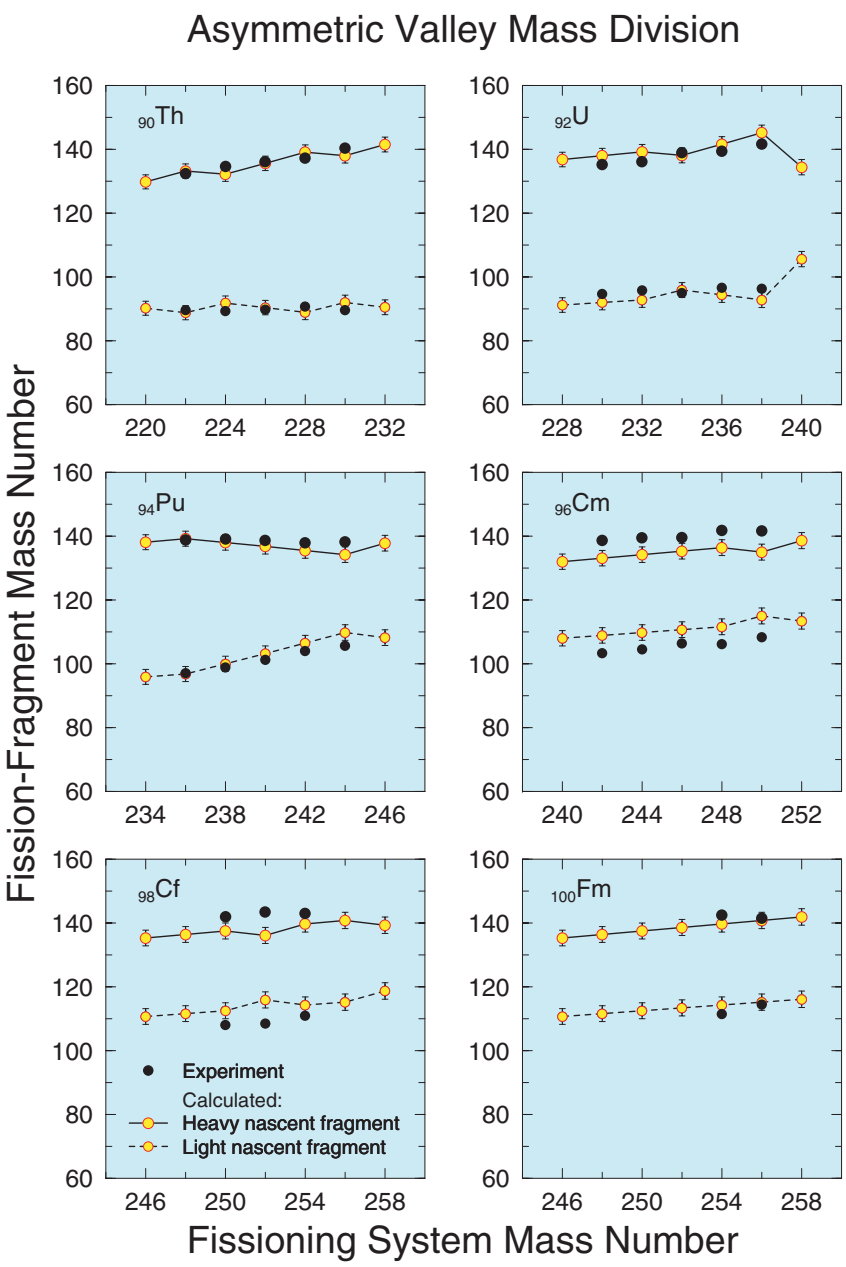


Fig. 3. Calculated heavy and light fission mean fragment masses compared to experimental data for even isotopes of six elements in the actinide region. The calculated mass divisions agree with experiment to within three mass units, about 1.3% of the total mass, which is an impressive agreement for this type of calculation.

unequal mass in heavy-element fission. Our results are compared to experiment in Fig. 3.

We remarked earlier that one of the earliest calculations with the Los Alamos macroscopic-microscopic model looked at the stability of very heavy elements, about 20 protons heavier than uranium. This was in 1973. By the time the the first mass calculation appeared in 1981 [13,14], numerous essential improvement had been incorporated into the model. This time also marked the beginning of a long series of new heavy-element discoveries at the Gesellschaft für Schwerionenforschung in Darmstadt, Germany, where elements with proton numbers 107–112 have now been discovered. The heaviest element, with proton number 112 was discovered in 1996. The nuclei of these elements are all *deformed* in shape, not spherical, which was the predicted shape of the superheavy elements near the next predicted magic proton and neutron numbers (Z = 114 and N = 184) beyond the last known elements.

Because these new elements lie between the last known doubly magic nucleus lead-208 and the predicted, next doubly magic nucleus ^{298}114, they were at first not anticipated to be sufficiently stable to be observed and therefore not studied theoretically for some time. However, the 1981 mass calculation showed very large gaps in the calculated deformed single-particle-level spectra for these nuclei and calculated shell corrections that increased the ground-state binding energy by well over 6 MeV, large enough to make these nuclei sufficiently long-lived to be experimentally observable. We show in Fig. 4 the calculated microscopic corrections in the heavy-element region and recently discovered heavy nuclei as red dots.

The small lake where the red dots are concentrated shows the area of particularly stable heavy deformed nuclei. The large lake beyond the flag shows the predicted region of spherical superheavy elements. Finally in Fig. 5 we compare calculated and experimentally observed energies of the α particles that are successively emitted when one of the heavy new elements, namely $^{272}111$, decays, and note the excellent agreement.

The calculated values are obtained directly from the calculated masses of the mothers and daughters of the decays. The calculated values were submitted for publication in 1993, well before these experimental results were available.

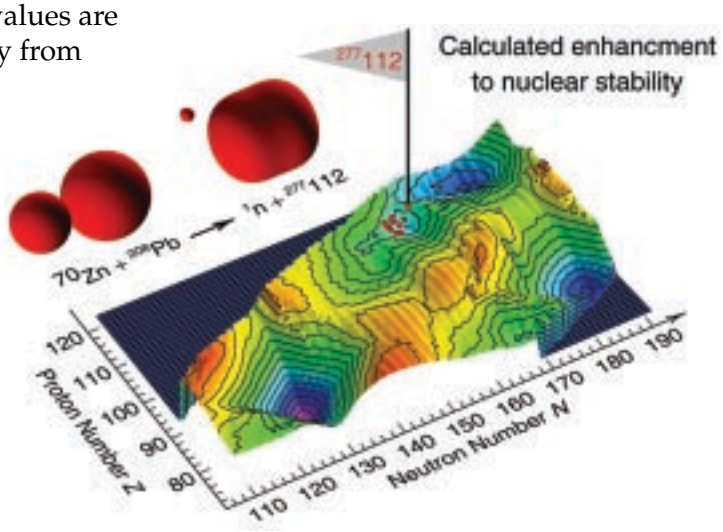


Fig. 4. Quantum-mechanical "shell-correction" effects on nuclear energy-level structure give rise to unusually strong binding for particular proton and neutron numbers. This contribution to the nuclear mass is calculated for several hundred heavy nuclei and shown, as is customary, vs the proton and neutron number of these nuclei. Local minima, indicated in various shades of blue correspond to increased contributions to nuclear stability from the shell-correction function. The last stable nuclei of the periodic system occur near lead (Pb) in the vicinity of proton number Z = 82 and neutron number N = 126, a so-called "doubly magic" nucleus. The red dots in another blue region show recently discovered very heavy nuclei. The decay half-lives of these nuclei range from a fraction of a millisecond to several tens of seconds. The deeper dark-blue lake beyond the flag shows a region of spherical superheavy nuclei that are predicted to have decay half-lives of years to thousands of years. Nuclei in this region have yet to be discovered.

This more-than-30-year effort has had a major impact on the world of nuclear physics. Los Alamos research has provided some of the earliest and most complete estimates of lifetimes and decay modes of superheavy nuclei, an extremely successful generalization of the liquid-drop model, and a nuclear mass model with the best extrapolation properties of any such model. In the 10 years since the last version was completed, it has not been equaled for predictability of nuclear masses and deformations of nuclei measured subsequent to its development. Recent fission calculations that computationally were one million times larger than 34 years ago [3,4] described in a single consistent calculation features such as barrier heights throughout the periodic system, the several modes of fission for nuclei near radium-228 and fermium-258, and the mean masses of the heavy and light fragments in asymmetric fission.

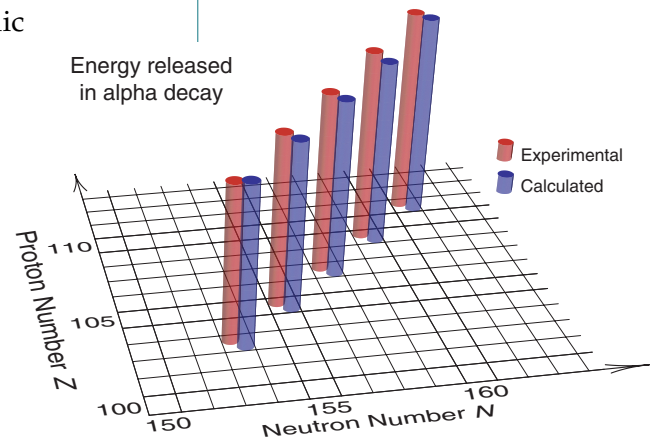


Fig. 5. Energy released in the α decay of the recently discovered heavy-element isotope $^{272}111$ and its decay products. Each alpha decay transforms the original nucleus into a heavy daughter with two neutrons and two protons less than the parent nucleus and a helium-4 atom. The decay energies are plotted versus the proton and neutron numbers of the original parent nucleus and the succession of four daughter decay products. The agreement between calculations (which were published before the discovery of this heavy element) and experiments is excellent.

From Nuclear Forces to Nuclei

J. Carlson

by Joseph A. Carlson

Although for many years it has been possible to determine the most basic properties of atoms and molecules from the underlying electromagnetic forces, it is very difficult to do the same thing for the nucleus. Solving for the ground state of a nucleus beyond the deuteron (one neutron and one proton) required the advent of modern computers. Theoretical Division invented the Monte Carlo algorithms used to understand the physics of light nuclei. We are developing new methods to study the reactions that produce solar neutrinos, weak parity-violating interactions in few-nucleon systems, and to analyze tests of physics beyond the Standard Model in light nuclei. Furthermore, new methods are being developed to study larger nuclei with realistic interactions and also the neutron-rich matter comprising neutron stars. Beyond understanding nuclear properties and reactions, our goals include using nuclei reliably as a probe of hadronic physics and physics beyond the Standard Model and to be able to calculate processes not accessible in any terrestrial laboratory.

For many years it has been possible to determine the most basic properties of atoms and molecules from the Coulomb interaction between electrons and with the atomic nucleus. It has proven very difficult to do the same thing for the nucleus itself, however. The average or mean-field potential of a nucleon is not easily obtained from the underlying nucleon-nucleon interaction, primarily because the interaction between nucleons depends very strongly upon the orientations of the nucleon's spin as well as their spatial positions.

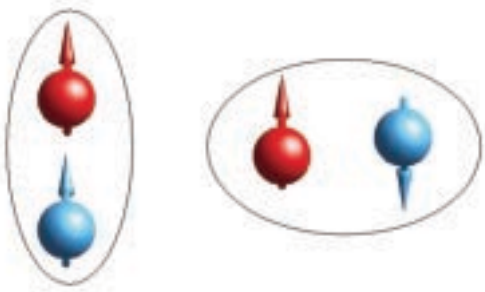


Fig. 1. Attractive spin and space orientation between a neutron and a proton. Flipping either the proton or neutron's spin would result in a repulsive interaction.

It is not yet possible to calculate the nucleon-nucleon interaction from quantum chromodynamics, the underlying theory of strong interactions. However, it has been extremely well studied experimentally, including important experiments conducted at Los Alamos and elsewhere. The chief features of the interaction are: (1) a strong repulsion at short distances, and (2) a spin- and isospin-dependent interaction at larger distance described by the exchange of pions. The spin and charge states of the pion imply that the force is similar to that between magnetic dipoles, as illustrated in Fig.1.

Solving for quantum states of many strongly correlated particles is often very difficult. A quantum solution for an N-particle state without spin-dependent interactions requires the solution of a 3N dimensional differential equation. For nucleons, which have two spin states and two isospin (neutron or proton) states, there are 4^N coupled differential

equations in 3N dimensions. Charge conservation and isospin conservation can be used to reduce the number of equations, nevertheless solving for beryllium-8 requires solving approximately 200,000 coupled differential equations to determine the ground-state of the system.

Solving for the ground state of a nucleus beyond the deuteron (one neutron and one proton) required the advent of modern computers. The study of the three-nucleon system was pioneered in the Theoretical Division (See the article "The Three Nucleon Systems" in this chapter.) in the mid-1980s. [1] This brought valuable insight into the structure and properties of the tri-nucleons: tritium and helium-3. However, solving for systems of more than three nucleons was impossible using traditional differential equation techniques because of the large number of coupled equations and their important spatial structure. Even today, it is just becoming possible to solve for four nucleons with similar methods.

Los Alamos and the T-Division had pioneered the use of Monte Carlo methods in physics, and quantum physics is a particularly important application. In the late 1980s Quantum Monte Carlo methods were developed at Los Alamos to handle systems with strong spin- (and isospin-) dependent interactions. The quantum problem is converted to a problem in statistical mechanics. Monte Carlo techniques are then used to integrate over the spatial coordinates of the nucleons, making it possible to study significantly larger nuclei. These methods were originally applied to determine the binding energy and structure of the alpha particle (helium-4).

These calculations required the addition of a small three-nucleon interaction to reproduce the binding of the three- and four-nucleon systems. The internal structure of the nucleon leads one to expect a three-nucleon interaction at the required level, about 10% of the nucleon-nucleon interaction. These three-nucleon interactions are being studied in neutron-deuteron and proton-deuteron scattering experiments.

Later these methods were refined and used to study larger nuclei. Along with colleagues at Argonne National Laboratory, the University of Illinois, and Jefferson Laboratory, calculations have been performed to study the spectra and structure of light nuclei as well as many of their important reactions. [2,3] With the advent of massively parallel computers, calculations were initially performed for six nucleons and later extended up to eight nucleons. Along the way it was realized that the interactions used to study the lightest nuclei were inadequate to describe heavier systems. In particular those with a large fraction of neutrons or protons were problematic, for example helium-8 has six neutrons and only two protons. Additional information is required, however, to determine the properties of neutron-rich systems. Theoretically plausible models of the 3-nucleon interaction have been constructed

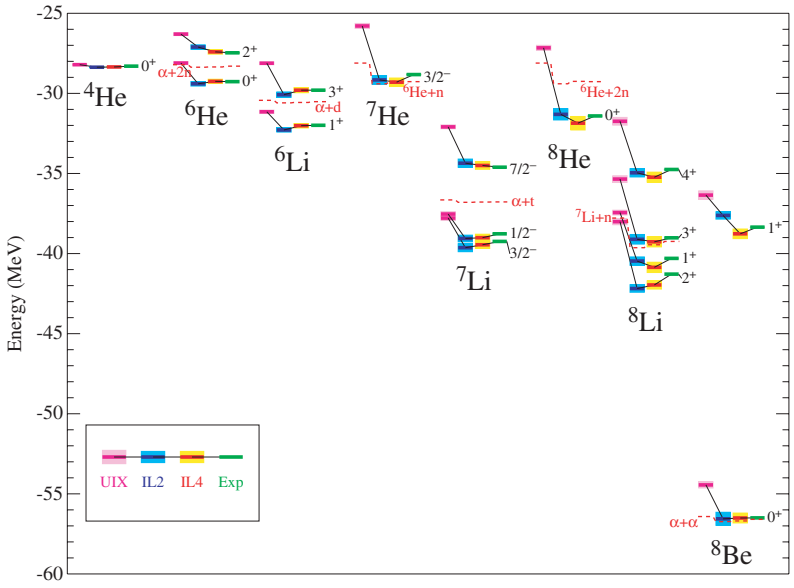


Fig. 2. The spectra of light nuclei (up to A = 8), theory versus calculations.

which can reproduce the spectra of light nuclei remarkably well (Fig. 2).

The structure of light nuclei are quite diverse. The nuclei from A = 5 to 8 are only weakly bound or not bound at all. The fact that there are no bound states of five or eight nucleons is important in describing nucleosynthesis. These nuclei manifest a significant shell structure arising from the nuclear mean field, but at the same time contain strong coupling into clusters of nucleons. Many of the basic properties beryllium-8 is an extreme case, consisting largely of two well-defined alpha particle clusters. This is shown in Fig. 3, where the ground-state density is plotted in the laboratory and intrinsic frames. In the laboratory, beryllium-8 is a spin-zero nucleus and has a spherical density distribution. However, in an intrinsic (internal) frame, the nucleus shows strong clustering into two alpha particles separated by approximately 4 fm. This structure is apparent in the excited states of beryllium-8, and is reproduced in these calculations even when starting from a mean-field picture.

In order to explain nuclear reactions or electron scattering experiments, it is crucial to use realistic models of the electroweak currents along with the realistic strong interactions. This combination allows one to successfully describe many properties of nuclei up to fairly high momentum transfers, including elastic electron scattering, sensitive to the charge and current distributions in the nucleus, and inclusive electron and neutrino scattering which can be used to obtain information about nuclear response and correlations.

A host of nuclear cross sections have also been calculated with realistic interaction models. Some of these reactions are difficult or impossible to measure in the laboratory because of their very small cross sections. One important example is proton-proton fusion, which provides the dominant source of energy in the sun. The reaction probability is very small at the solar energies of interest because of the Coulomb repulsion between protons. Calculations correlating this cross section with the rate of tritium beta-decay yield a result to an accuracy of << 1%. Calculations

have also been performed for a variety of electroweak capture reactions in the light nuclear regime.

These same methods are now being used to study the reactions that produce solar neutrinos, weak parity-violating interactions in few-nucleon systems, and to analyze tests of physics beyond the standard model in light nuclei. Furthermore, new methods are being developed to study larger nuclei with realistic interactions, and also the neutron-rich matter comprising neutron stars. Beyond understanding nuclear properties and reactions, goals include using nuclei reliably as a probe of hadronic physics and physics beyond the Standard Model, and to be able to calculate processes not accessible in any terrestrial laboratory.

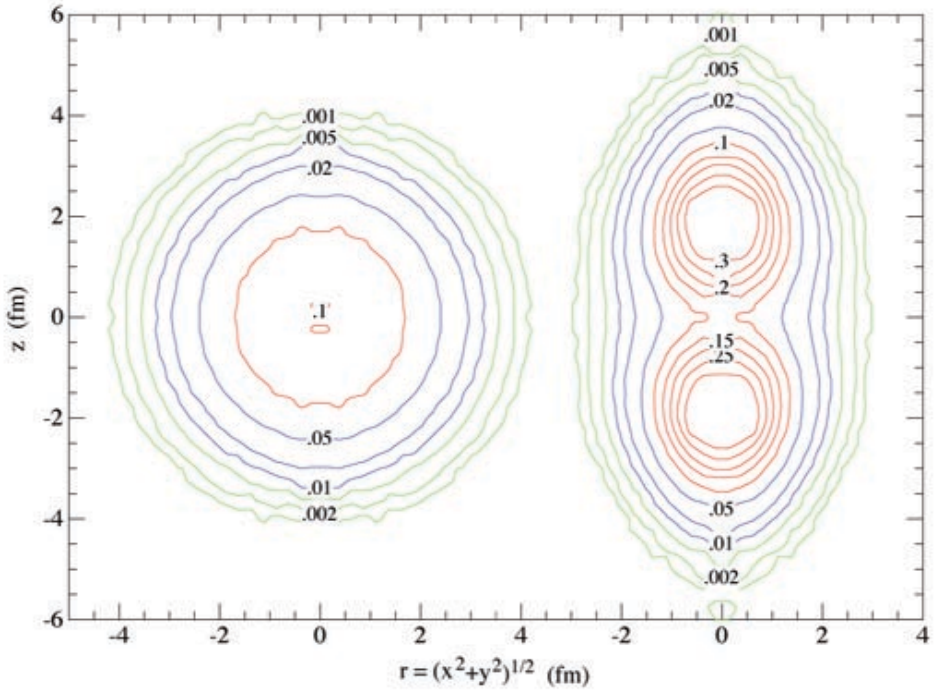


Fig. 3. Beryllium-8 nucleon densities in the laboratory (left) and intrinsic (right) frame.

The Three-Nucleon Systems

J. Carlson
Friar

by James L. Friar

Los Alamos theorists and their collaborators at the University of Iowa were the first to perform accurate calculations of the properties of helium-3 and hydrogen-3, nuclei that are composed of three nucleons. This work used the fastest supercomputers available at the time and introduced novel techniques for solving the equations (governing systems of three nucleons) that follow from the laws of quantum mechanics. Prior to this work only systems of two nucleons could be accurately calculated starting from the nuclear forces, and it took more than 30 years to proceed from two to three nucleons.

The strength of the binding of the nucleons in helium-3 (or in hydrogen-3) tells us about the forces between two nucleons or groups of nucleons. When only pairwise forces were used in the calculations (i.e., forces between individual pairs of nucleons), the computed binding was slightly weaker than the measured binding. Because these forces described the properties of a single pair of nucleons fairly well, our calculations provided a strong argument that three-nucleon forces (forces involving three nucleons simultaneously) were also necessary for a good description of nuclei.

Nuclear physics is the study of nuclei, which form the cores of atoms. Nuclei are about 100,000 times smaller than typical atoms. Just as atoms are composite systems comprised of electrons revolving around their nuclei, the nuclei themselves are composites whose primary constituents are called nucleons. Nucleons come in two different but closely related species: protons (with positive electric charge) and neutrons (which are electrically neutral). Together these two species are generically called nucleons, and differentiating between the two species of nucleon is called ''isospin.'' It is the positive charge of the protons attracting the negative charge of atomic electrons that permits atoms to form as bound systems.

The electrical interaction between charged particles (seen in experiments with pith balls charged by static electricity) is quite a weak force compared to the force that binds nucleons together to form nuclei. The electrical force also has a very long range, and can easily affect objects separated by distances of centimeters or even meters. The nuclear force between nucleons has a very short range and is tiny when the nucleons are more than 10^{-12} cm apart. The binding of nuclei by strong, short-range forces is in stark contrast with the binding of atoms by weak, long-range forces.

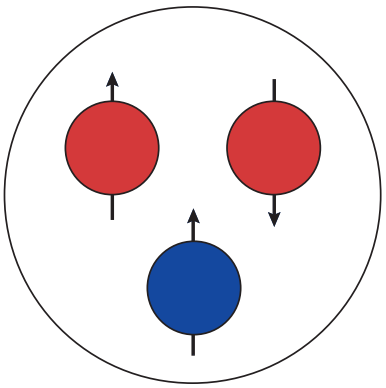
In addition to having two species of nucleons (i.e., isospin), nucleons have ''spin.'' They can be visualized as spinning balls of matter. The laws of quantum mechanics dictate that particles like the nucleons can have a fixed amount of spin directed either parallel or antiparallel to a particular direction. Thus there are four "intrinsic'' degrees of freedom in a nucleon: two for isospin (either a proton or a neutron) and two for spin (pointed either up or down).

A typical light nucleus (an isotope of helium, helium-3) is shown schematically in Fig. 1. We represent the isospin degree of freedom by color, with protons as red and neutrons as blue. These colors are only for purpose of illustration; any colors could have been used. The two (red) protons have their spins antialigned, one with spin pointed up and the other pointed down. The neutron (in blue) has its spin pointed up. All helium nuclei have two protons, while those with different numbers of neutrons are called helium isotopes (there was one neutron in our example, making a total of three nucleons, hence helium-3). Because the protons have cancelling spins, the resulting spin of the helium-3 nucleus is the same as the spin of the neutron.

The complexity of the nuclear force results largely from the many possible combinations of spin (either up or down) and isospin (either proton or neutron) that are possible. The force between the two protons in the example would be very different if their spins were aligned. In addition the force between a proton and a neutron is different from the force between two protons (or between two neutrons), and also depends strongly on the alignment of spins.

T-Division theorists and their collaborators at the University of Iowa were the first to perform accurate calculations of the properties of helium-3 and hydrogen-3 (the heaviest form of hydrogen, with one proton and two neutrons; simply interchange the red and blue colors in Fig. 1). This work used the fastest supercomputers then available, and introduced novel techniques for solving the equations (governing systems of three nucleons) that follow from the laws of quantum mechanics. Prior to this work only systems of two nucleons could be accurately calculated starting from the nuclear forces, and it took more than 30 years to proceed from two to three nucleons. Calculations of bound four-nucleon systems by Joe Carlson in T-Division were performed a few years later using a very different (but very powerful) numerical technique.

Fig. 1. The helium-3 nucleus, with two protons (red) and one neutron (blue), showing spin alignment.



What do these calculations tell us and why did we do them? The strength of the binding of the nucleons in helium-3 (or in hydrogen-3) tells us about the forces between two nucleons or groups of nucleons. When only pairwise forces were used in the calculations (i.e., forces between individual pairs of nucleons), the computed binding was slightly weaker than the measured binding. Because these forces described the properties of a single pair of nucleons fairly well, our calculations provided a strong argument that three-nucleon forces (forces involving three nucleons simultaneously) were also necessary for a good description of nuclei.

One can also perform three-nucleon scattering calculations and compare them to the corresponding experiments. They involve hitting a target containing a heavy isotope of hydrogen-2 (hydrogen, containing one proton and one neutron) with a beam of protons or neutrons. Such collisions also provide important information about nuclear forces. Some of the earliest calculations of this type were performed by the Los Alamos–Iowa collaboration. These calculations and the corresponding experiments provide additional evidence for the existence (and strength) of three-nucleon forces.

The combination of strong, short-range pairwise forces between nucleons supplemented by relatively weak (and also short-range) three-nucleon forces is now firmly entrenched as the standard set of ingredients necessary for understanding the properties of nuclei.

R-Matrix Theory of Nuclear Reactions in T-Division

by Gerald M. Hale and Donald C. Dodder

Collins
Dodder
Eisenbud
Hale
Lane
Schneider
Thomas
Wigner
Witte

When the R-matrix theory of nuclear forces was proposed in 1947, little was known about the nuclear forces governing them. In the 1960s, T-Division initiated a numerical project to use R-matrix theory to analyze nuclear reactions that has continued to this day. The theoretical promise of R-matrix theory proposed as a general framework for describing nuclear reactions was realized in numerical work begin in T-Division more than 30 years ago.

In the 1970s, comprehensive R-matrix analyses of data in the light nuclear systems were used to provide neutron cross sections for fission reactor applications and charged-particle cross sections for fusion energy applications. More recently, new systems have been analyzed to provide cross sections and rates of interest to astrophysics and the nuclear weapons program, including the helium-burning reactions in stars that made the carbon and oxygen on which terrestrial life is based and the reactions that created the lightest elements during the big bang. Information developed recently from these analyses forms the most complete database for the thermonuclear reactions ever available in the weapons program.

The exact nature of nuclear forces has eluded scientists for decades. When Nobel laureate Eugene Wigner first proposed R-matrix theory [1] with Leonard Eisenbud in 1947, little more was known about them other than that they were very strong, mostly attractive, short-ranged, and could depend on the spins of the interacting particles. Wigner capitalized on the short-ranged nature of these forces to introduce what is still a very modern idea: to build into the description what is known about the long-ranged behavior of the system and parameterize what is largely unknown about the short-ranged behavior. This idea forms the basis of present-day effective field theories of the strong interactions, and despite the fact that much more is now known about nuclear forces, it remains a viable approach to describing nuclear reactions phenomenologically.

The part of the theory that is known is the long-ranged (asymptotic) behavior of the wave function describing the relative motion of two interacting nuclei. Beyond the range of the forces causing the scattering, this wave function can be expressed as a linear combination of incoming and outgoing spherical waves. Coefficients giving the amplitudes of the outgoing waves relative to the incoming ones, called elements of the scattering (or *S*) matrix, contain all the information necessary to calculate the result of any possible scattering measurement on the system. Wigner's idea was that, because nuclear forces are so short-ranged, the

asymptotic form for the wave function holds down to very small separations (the order of a few femtometers, or 10^{-15} meters) between the interacting nuclei, known as "channel radii." These channel radii define a "nuclear surface," inside of which the wave function becomes much more complicated due to the many-body (usually taken to be protons and neutrons, or more generally, "nucleons") nature of the interacting system when the nuclei are close together. A quantity Wigner called the *R* matrix contains the information about this complicated region inside the nuclear surface, expressed through unknown parameters in a relatively simple mathematical form. Matching the parameterized solution to the known asymptotic form of the wave function at the nuclear surface relates the matrices *R* and *S*, allowing the experimental data from nuclear scattering experiments to be used to determine the parameters of the *R* matrix, using a least-squares fitting procedure.

It took some years for the full implications of this new theory to be realized. T-Division staff member Robert Thomas and British physicist Anthony Lane studied, sometimes in collaboration with Wigner, aspects of the "compound-nucleus" theory of nuclear reactions that had grown out of Niels Bohr's picture of the atomic nucleus. Lane and Thomas saw in this picture fundamental connections with Wigner's theory, and they labored for 6 years to write the "bible" of R-matrix theory [2] for *Reviews of Modern Physics*. Puzzling experimental anomalies, such as Wigner cusps (sharp spikes near reaction thresholds) and inverted ordering of nuclear energy levels, could be explained by the known long-ranged behavior built into R-matrix theory without any detailed knowledge of nuclear forces. Sadly, Thomas had died by the time the article appeared in 1958, and the subject was not pursued further at Los Alamos until several years later.

During the late 60s, Donald Dodder initiated in the division an ambitious numerical project to use R-matrix theory to analyze nuclear reactions that has continued to this day. Dodder's vision was to develop the world's most sophisticated and general computer program to analyze experimental data from nuclear reactions. The program would treat the reactions of any type of particles—neutral or charged, massive or massless—using a general quantum-mechanical formalism that allows the computation of any measured scattering "observable." This vision was implemented by Kathleen Witte, an expert computer programmer who had been trained as a physicist, and was working in the Computing (C) Division. Their code, named "EDA" (for Energy Dependent Analysis), has run on many different types of computers over the years, and remains, to our knowledge, the most general computer program of its type in existence.

Gerald Hale joined this effort as a post-doctoral fellow in 1970. He and Dodder first applied the code to the interpretation of the wealth of polarization data for light nuclei that was coming from experiments performed at Los Alamos and other laboratories around the world. These studies demonstrated for the first time that R-matrix theory could be used to describe fully with a single set of parameters all possible measurements for all nuclear reactions that occur in a given "compound

system" containing a fixed number of nucleons. (Wigner once said, "a dream of mine came true," upon hearing this work described at a conference.) During this same period, incidentally, R-matrix methods for doing calculations of electron scattering from atoms and molecules were pioneered in T-Division by Barry Schneider and Lee Collins.

Later, the motivation and support for the work became increasingly more applied, especially under Hale's direction after Dodder retired. Comprehensive R-matrix analyses of data in the light nuclear systems were used to provide neutron cross sections for fission reactor applications, and charged-particle cross sections for fusion energy applications. [3] Because so much experimental information was included, these analyses greatly increased the precision of the data needed for applications, while at the same time revealing the nuclear energy-level structure of these compound systems in more detail than ever before. During the late 80s and early 90s, Hale collaborated with other theorists in T-Division and elsewhere, using R-matrix methods to investigate the nuclear effects in muon-catalyzed (room-temperature) fusion of the light elements (see Jim Cohen's article "Muon-Catalzyed Fusion" in the "Atomic Physics" chapter in this book).

More recently, the R-matrix work in the division has diversified in new directions. The realization that the theory has a well-behaved numerical extension into the complex energy plane [4] led to a program of determining the energy levels in light nuclei in terms of the complex poles of the S matrix. The S-matrix elements at real energies, parameterized in terms of R-matrix elements by analyzing experimental data for light systems, have provided valuable tests of microscopic calculations in the few-body systems. These comparisons reveal, among other things, that pair-wise forces between nucleons are insufficient to account for the size of certain transition probabilities in the three- and four-body nuclear systems.

New compound systems have been analyzed in order to provide cross sections and rates of interest to astrophysics and the nuclear weapons program. These include the helium-burning reactions in stars that made the carbon and oxygen on which terrestrial life is based and the reactions that created the lightest elements during the big bang. Some of the astrophysical rates obtained at temperatures where stars burn differ significantly from the previous values. Information developed recently from these analyses forms the most complete and accurate database for the thermonuclear reactions ever available in the weapons program.

Thus, the theoretical promise of R-matrix theory as a general framework for describing nuclear reactions, as proposed originally by Wigner and Eisenbud, and elaborated by Lane and Thomas, was realized in the numerical work begun in the Theoretical Division more than 30 years ago by Dodder and Witte. This program continues with the efforts of Hale and his collaborators, providing nuclear data that are used in a variety of applications and insight into the nature of nuclear forces and light nuclei that is derived from these sophisticated analyses of the experimental measurements.

Meson Nuclear Physics

Gibson

by Benjamin F. Gibson

At the core of each atom lies the atomic nucleus, a many-body system of neutrons and protons (nucleons) held together by the "strong interaction," the short-range nuclear force. Strong interaction nuclear physics is the branch of science devoted to the investigation of the properties of nuclei and the strong force that binds them together. Atomic nuclei play an essential role in the Laboratory's mission relating to fission and fusion. In the 1960s Congress authorized building the Los Alamos Meson Physics Facility (LAMPF) to explore nature's secrets in strong interaction physics, and Theoretical Division's effort was essential in that endeavor.

Subatomic mesons, which are exchanged among nucleons, constitute the glue that binds together the atomic nucleus. The pion, which is the lightest of these mesons, plays a key role throughout strong-interaction nuclear physics. Pions can be created, propagated, and absorbed within the nucleus and play a dominant role in the strong force. At LAMPF, pions were created by slamming energetic protons into copper targets, and beams of resulting pions were focused like flashlights to probe the structure of nuclei. The pion exists in three charge states (π^+, π^0, π^-), which make possible single-charge-exchange reactions (converting a proton into a neutron or vice versa) and double-charge-exchange reactions (converting two protons into two neutrons or vice versa). This versatile probe was the foundation of more than two decades of nuclear physics basic research at Los Alamos.

Just as a pion can be created, it can be absorbed or made to disappear. In particular, the pion can be absorbed radiatively, a process in which pion plus nucleon is converted into photon plus nucleon. Absorption of negatively charged pions (π^-s) by deuterium (the bound state of a neutron and proton) leading to a photon and two neutrons was proposed as a direct means of measuring the low-energy neutron-neutron (*nn*) interaction. The *nn* interaction cannot be investigated in an incident-beam-on-target experiment because constructing a neutron target is not possible. Based upon T-Division's theoretical analysis, experiments at PSI in Switzerland and at LAMPF yielded definitive values for low-energy *nn* scattering, which confirmed the existence of a sizable charge symmetry breaking—that is, the *nn* interaction is not equivalent to the proton-proton (*pp*) interaction—in the fundamental strong interaction.

Elastic scattering of electrons from nuclei is sensitive to the charge distribution and radius of the target nucleus. The charge radii difference among isotopes (e.g., carbon-12-carbon-13) and isotones (e.g., hydrogen-3-helium-4) has been extensively studied by means of elastic electron scattering. Because a π^+ interacts more strongly with protons than with neutrons, elastic scattering of π^+s was confirmed to provide a similar sensitivity to proton charge radii differences for light nuclei. As suggested by theory, elastic scattering of π^-s from nuclei was shown to be sensitive to neutron radii differences, which cannot be measured via electron scattering. Therefore, using elastic scattering of π^+s and π^-s from the same target, one obtains a precision measurement of the difference between the neutron radius and the proton radius for a given nucleus. Analysis of π^+ and π^- elastic scattering from hydrogen-3 and helium-3—1 proton and 2 neutrons versus 2 protons and 1 neutron—showed the proton radii difference to be in agreement with results of state-of-the-art trinucleon calculations that were based upon contemporary, realistic models of the nucleon-nucleon strong interaction and established a first measurement of the neutron radii difference, also in agreement with the T-Division calculations.

Understanding collective degrees of freedom in nuclei has been a goal of strong interaction nuclear physics since the discovery of the giant isovector dipole resonance, a vibration of the nucleus as pictured in Fig. 1a. The collective motion is isovector because protons move in a direction opposite to that of neutrons; it is dipole because the charged protons are separated to one side while the neutrons are to the opposite side. A negatively charged electron can interact with the positively charged protons in a nucleus pushing them to one side, and the separated proton and neutron fluids then oscillate back and forth as do the positive and negative charges in a dipole radio antenna while emitting a signal. It was only when the (π^-,π^0) single-charge-exchange reaction was promoted to the ranks of a reliable probe of nuclear structure with completion of the LAMPF π^0 spectrometer that one was successful in locating the giant isovector monopole resonance, a vibration of the nucleus as depicted in Fig. 1b. The collective motion is again isovector because protons and neutrons move in opposite directions as their spherical distributions alternately compress and expand; it is monopole because there are equal numbers of protons (neutrons) on any side of the nucleus at any point in time as the protons and neutrons oscillate against one another in a radial direction. As predicted by T-Division, the π^- single-charge-exchange reaction interacted more strongly with the neutrons that form the surface of heavier nuclei than with the protons on the interior, to excite the isovector monopole resonance.

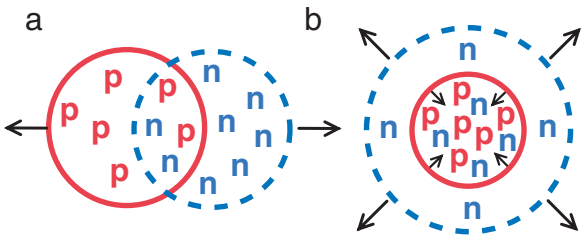


Fig. 1. Schematic description of (a) the isovector giant dipole resonance in which protons (*p*) and neutrons (*n*) oscillate collectively in opposite directions and (b) the isovector giant monopole resonance in which they alternately compress and expand radially.

Understanding nuclei that are not found in nature has important implications for nuclear astrophysics and the creation of the elements in our solar system. In particular, isotopes which contain more neutrons than do nuclei that lie along the "line of stability," defined by nuclei found in nature, provide constraints on our models of the strong interaction. The pion double-charge-exchange reaction was first proposed as a tool to search for nuclear isotopes off the line of stability, because of the clean signal provided by a pion projectile, with no excited states to cloud the interpretation of data resulting from scattering a beam of pions. An example was the search for the exotic helium-9 (containing 5 more neutrons than the naturally occurring helium-4) isotope using the (π^-, π^+) reaction. No isotope was found, which suggested that one had reached the "neutron drip line," the point at which no additional neutrons could be added to the nucleus such that the system would be bound. It was then realized that the pion double-charge-exchange reaction could be used to probe the size of the neutron halo (pure neutron matter lying outside the neutrons and protons comprising the core of a neutron rich nucleus) in a nucleus such as lithium-11, which has four more neutrons than the lithium-7 isotope found in nature. That the distance from the center-of-mass of the last two neutrons in lithium-11 is more than twice the size of the lithium-9 nuclear core was established through model calculations in a T-Division analysis of the pion double-charge-exchange reaction data.

During the LAMPF era, the scattering of pions from atomic nuclei proved to be a particularly useful tool with which to investigate the size and structure characteristics of many nuclei. T-Division played a major role.

Heavy Ions at Los Alamos National Laboratory

by Daniel D. Strottman

Amsden
Bethe
Carruthers
Clare
Copper
Goldhaber
Harlow
Kapusta
Nix
Ornik
Schlei
Strottman
E. Teller

The Theoretical Division developed a numerical technique for the calculation of high-energy nuclear physics processes. The results of the calculations agree remarkably well with the experimental results.

One of the new and exciting areas of nuclear physics research is the study of extreme states of nuclear matter. This is usually thought of as either studying nuclei that have a large excess of either neutrons or protons or nuclear matter that is many times the density of normal nuclear matter. Low energy and lighter mass projectiles have been used for decades to study properties of nuclei. In the last decade, however, accelerators have been built that can accelerate lead nuclei to near the velocity of light, thereby allowing one to explore regimes of nuclear matter hitherto impossible. The collisions of such nuclei produce densities and temperatures that are so great that the identities of individual nucleons are lost and it is only meaningful to describe the result in terms of subnucleonic degrees of freedom such as quarks and gluons. Such an entity is a new phase of matter—the quark-gluon plasma.

The initial theoretical work in the early 1970s by Bethe, Teller, and others described the collisions using fluid dynamics. Fluid dynamics provides an intuitively simple description of heavy ion collisions. After two nuclei smash into each other, they are rapidly compressed with the production of shock waves, before expanding and breaking up into bits of hadronic matter that ultimately reach the detectors. Fluid dynamics can summarize the conservation laws for energy and momentum into a few equations and all of the unknown physics could be summarized into an equation of state. The conditions for the validity of such an approach were reasonably well satisfied: nucleons in nuclear matter at the energies then possible to achieve in the Laboratory had an almost zero mean free path. The original work described results using concepts such as shock waves and the maximum densities for a given equation of state. (It is interesting to note that one of the standard references for shock waves then was an unpublished Los Alamos report written by Hans Bethe during the Manhattan Project while he was T-Division Leader and Deputy Laboratory Director.)

The original work dealt exclusively with nonrelativistic fluid dynamics. Anticipating the much higher energies to come, Ray Nix generalized these early results to a relativistic formalism. (The highest energies then available in the laboratory was 250 MeV per nucleon from the Bevalac or a velocity of approximately 0.27 c.) Harlow, Amsden, and Nix in T-Division published the first numerical method for the solution of relativistic fluid dynamics in 1976; it generalized the Particle-in-Cell (PIC) method introduced by Harlow in 1957. This first effort was for

two dimensions but a three-dimensional code was introduced within a year. The figure shows an example of three different offsets of the collision path. The result of these calculations when compared with experiment was extraordinary–so much so that several other codes were written elsewhere (Kurchatov, Frankfurt, and Marburg). Only the fluid dynamics method could explain the particle distributions from the collisions as well as the measured flow and other observables. In subsequent years the Laboratory's code method was used to investigate many conjectures, e.g., the effects on measured quantities of a Lee-Wick density isomer, the production of entropy with different equations of state, pion production, etc.

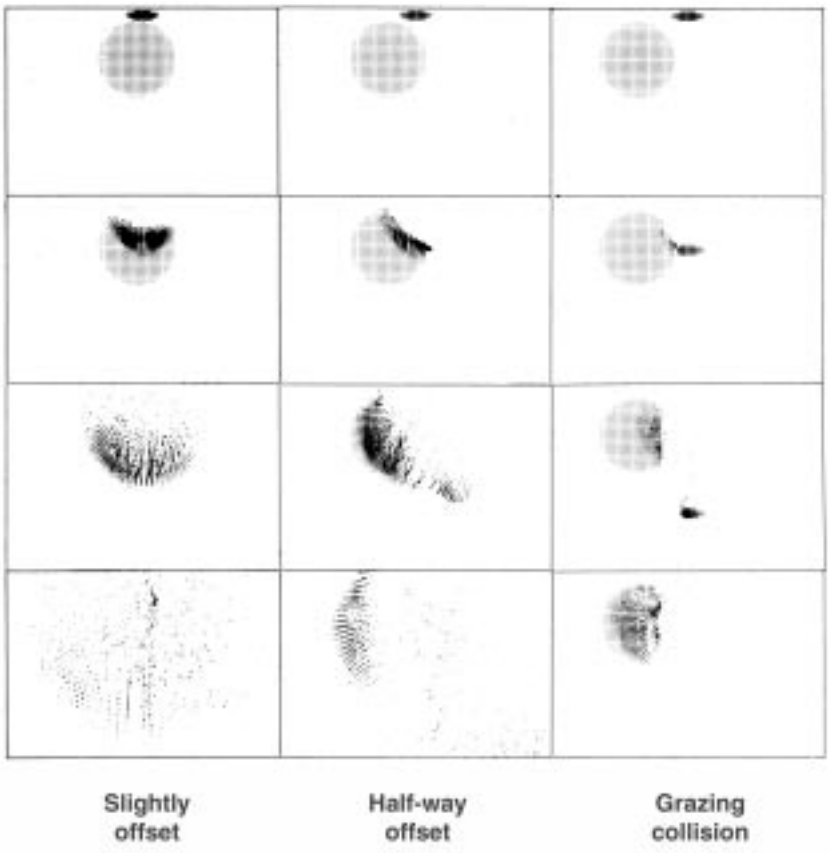


Fig. 1. Calculated time evolution of the matter distribution in the collision of a neon nucleus and a uranium nucleus for high enough impact velocity that the neon shows strong relativistic contraction.

The original work of Nix demonstrated that relativity imposes strict requirements on the form of the equation of state–at large densities, the compression energy cannot increase with density faster than linearly with density. If the compression energy increased faster than this, the speed of sound in the nuclear matter would ultimately exceed the speed of light; this led to the development of the Sierk-Nix equation of state. This and subsequent work at the Laboratory on the restrictions imposed on the thermal degrees of freedom showed that most fluid-dynamic calculations performed elsewhere up to that time violated causality.

Because some of the underlying principle assumptions of fluid dynamics, e.g., that of a zero mean free path and instant thermalisation, conflict with the presumed transparency of nuclear matter at high-bombarding energies, it has long been assumed that fluid dynamics would not be valid at higher bombarding energies. To include effects of transparency, a two-fluid model was developed in 1978 in T-Division by Amsden, Goldhaber, Harlow, and Nix. To obtain the equations that describe the two-fluid model, each nucleus is assumed to be a fluid that has the identical properties of the fluid representing the other nucleus. When the two fluids collide they were allowed to exchange energy and momentum. The most extensive calculations were performed in T-Division by Clare and Strottman and by Goldhaber and Strottman.

As already mentioned, it was always assumed by theorists that the hydrodynamic description of relativistic heavy ions would fail as the bombarding energies increased. Contrary to expectation, however, there continued to be agreement between the data and theoretical calculations, first at 2.1 GeV per nucleon measurements at Lawrence Berkeley Laboratory, 15 GeV per nucleon work at Brookhaven National Laboratory, and then at 60 and 200 GeV per nucleon measurements at CERN, the European Laboratory for Particle Physics. Importantly, some of the hydrodynamic results were later confirmed by experiment.

There are several applications of fluid models not already mentioned. Chief among these are the study of correlations by Ornik and Schlei, the effects of the equation of state on the hydrodynamic flow, comparisons between results from hydrodynamic models and microscopic models, and the role of hydrodynamic expansion in the low p_T anomaly. The most extensive calculations using hydrodynamic models of Bose-Einstein correlations for inclusive spectra have been performed over several years by Schlei. As the hot, nuclear matter that has resulted from a heavy ion collision expands, it ultimately becomes too dilute for a fluid dynamical description to be valid. At this point it is assumed the matter freezes out; a description invented by Cooper and Frye has been historically used, but recent work at Los Alamos and elsewhere has suggested improved ways to handle freeze-out.

The success of fluid dynamical models continues to confound skeptics. Many open questions remain including why they are as successful as they appear to be. T-Division Leader Peter Carruthers summarized speculations along this line already in 1974. But even the strongest adherents realize that at the energies available at the Relativistic Heavy Ion Collider, the hydrodynamic description cannot be valid. A theoretical model has been constructed in which the merits of both microscopic and hydrodynamic models are included. The initial conditions of a relativistic collision are described by an effective string-rope model that incorporates known features of partons in high-energy proton-proton collisions. The results show that quark-gluon plasma forms a tilted disk, such that the direction of the largest pressure gradient stays in the reaction plane but deviates from both the beam and the usual transverse flow directions. The produced initial state can be used as an initial condition for further hydrodynamical calculations. The resulting initial locally equilibrated state of matter in semicentral collisions takes a rather unusual form and the resulting flow velocity field is in agreement with the "firestreak" phenomenological picture. Such an initial state will manifest itself by the asymmetry of the produced collective flow, even after the subsequent hydrodynamical expansion.

Thus I was assigned to an entirely different sort of job; I was made a group leader in Hans Bethe's Theoretical Division and my group was given the problem of determining all of the effects of an atomic bomb which would take place after the nuclear fission had occurred.

—Joseph O. Hirschfelder [former group leader in T-Division], "The Scientific and Technological Miracle at Los Alamos," in *Reminiscences of Los Alamos: 1943–1945*, p. 72.

Gibson

Novel Aspects of Λ Hypernuclei

by Benjamin F. Gibson

Over the past 30 years, Theoretical Division has developed the computer codes to make definitive calculations of the properties of light atomic nuclei using contemporary, realistic models of the forces that govern the interactions of neutrons with neutrons, protons with protons, and neutrons with protons. Unanswered was the question of whether the truly sophisticated nuclear-force models that form the basis for these calculations could extrapolate beyond the realm of conventional nuclear physics where they had been developed, or if they were merely exquisite interpolation tools whose validity lies within the bounds of known neutron-proton physics. T-Division research is now focused on multihyperon systems, moving out along the third dimension (strangeness), exploring what may well be the "stepping stones" to the stars.

Conventional atomic nuclei, composed of protons and neutrons, account for most of the observable mass within the visible universe. Protons and neutrons provide two degrees of freedom along which nuclei can develop isotopes and isotones. Isotopes are nuclei consisting of a specific number of protons and differing numbers of neutrons, e.g., hydrogen-1, hydrogen-2, and hydrogen-3 (hydrogen, deuterium, and tritium), each have a single proton plus zero, one, or two neutrons, respectively. Isotones are nuclei consisting of a specific number of neutrons and differing numbers of protons, e.g., hydrogen-3 and helium-4 each have two neutrons but one or two protons, respectively. Over the past 30 years, Theoretical Division has developed the computer codes to make definitive calculations of the properties of light atomic nuclei using contemporary, realistic models of the forces that govern the interactions of neutrons with neutrons, protons with protons, and neutrons with protons. Properties of isotopes and isotones such as their binding energies and sizes plus characteristics like beta-decay (conversion of a neutron into a proton, an electron, and a neutrino) of tritium (hydrogen-3) into helium-3 can be calculated with unparalleled precision. Unanswered was the question of whether the truly sophisticated nuclear-force models that form the basis for these calculations could extrapolate beyond the realm of conventional nuclear physics where they had been developed or if they were merely exquisite interpolation tools whose validity lies within the bounds of known neutron-proton physics.

Hypernuclei contain, in addition to neutrons and protons, one or more "strange" baryons (Λ, Σ, or Ξ hyperons), where strangeness (S) is a means of distinguishing the neutron from the neutral Λ, Σ, or Ξ (each a different mass) just as charge distinguishes a proton from a neutron.

The baryon octet of nucleons and hyperons is pictured in the figure with baryons of a given strangeness (S = 0, –1, –2) aligned by row. By injecting a strange hyperon into a nucleus, one creates a new degree of freedom, a third dimension along which to explore the properties of nuclei. Just as a third dimension in our everyday world can lead to discoveries not possible in a flat world of two dimensions, exploring "strange" hypernuclei has revealed physics that is novel and, at times, puzzling. Our intuition has been stretched beyond the insight acquired from-three-quarters of a century of research involving conventional (nonstrange, or S = 0) nuclear and particle physics.

The energy required to separate a deuteron into its constituent nucleons (a neutron and a proton) has been measured to be ~2.2 MeV. The energy required to separate the additional neutron from a triton (hydrogen-3) leaving behind a deuteron is some 3 times that 2.2 MeV, or ~6.3 MeV. Likewise, the energy required to separate a neutron (or a proton) from the alpha particle, helium-4, leaving behind helium-3 (or hydrogen-3) is approximately 3 times that 6.3 MeV, or ~20 MeV. That is, the ratio of neutron separation energies for neighboring light nuclei is ~3. The Λ does not bind with a neutron or a proton, so that the lowest mass hypernucleus is the hypertriton (Λ-hydrogen-3)—a Λ bound to a deuteron nuclear core. If we examine the hypertriton (Λ-hydrogen-3), the mass number 4 pair (Λ-hydrogen-4, Λ-helium-4), and the alpha-like Λ-helium-5 system, the factor of 3 in comparing neutron and proton separation energies from neighboring nuclei becomes a factor of 20 when one compares the energy required to separate a Λ from Λ-hydrogen-4 (~2 MeV) with that required to separate a Λ from Λ-hydrogen-3 (~0.13 MeV). Moreover, the factor between the Λ separation energy for Λ-helium-5 (~3 MeV) and that for Λ-hydrogen-4 is only 1.5. Hypernuclei are not simple extensions of conventional S = 0 phenomena; the physics is different. The difference is further emphasized by the fact that Λ-helium-4 and Λ-hydrogen-4 have an excited state that lies only ~1 MeV above their ground state and de-excite by emitting a photon, whereas the S = 0 conventional nuclei hydrogen-3, helium-3, and helium-4, have no such excited states.

Although the data for hyperon scattering from protons is sparse, models of Λ-nucleon (ΛN) scattering can be constructed. Perhaps not surprisingly, T-Division calculations demonstrated that a model of the lightest hypernuclei based upon such interaction models yielded the same factor of ~3 in the ratio for Λ separation energies that is observed (and calculated) in the case of neutron separation energies in the light S = 0 nuclei. This is in complete contradiction of the experimental observations outlined above. That is, our sophisticated models of the S = 0 interaction must be modified in order to successfully describe the analogous S = –1 hyperon-nucleon (ΛN) interaction in light hypernuclei. Research in T-Division discovered the neglected aspect in the modeling of the ΛN interaction—the fact that the (S = –1) ΛN and ΣN systems are strongly mixed. That is, when a Λ scatters from a

nucleon, the ΛN pair spend a significant fraction of the interaction time as a "virtual" ΣN pair, even though it is not energetically possible for the Σ to separate from the nucleon and be observed as a "free" Σ. This important difference in the hyperon-nucleon interaction compared with nucleon-nucleon interactions gives rise to a number of physical observations not anticipated from our extensive experience in the $S = 0$ sector.

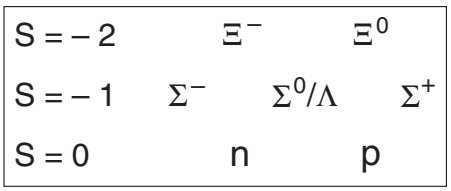


Fig. 1. The baryon octet of nucleons and hyperons aligned by row according to their strangeness $S = 0, -1, -2$.

T-Division has played a major role in developing an understanding of these novel hypernuclear phenomena, providing forefront 3-body, 4-body, and 5-body calculations. For example, ΛN-ΣN mixing was demonstrated to play a key role in explaining the unexpectedly small Λ-helium-5 binding, where the Λ separation energy is only 3 MeV rather than the 5–6 MeV predicted by models based on an unmixed ΛN interaction. Such $S = -1$ mixing was shown to produce a three-body force (an interaction involving the Λ, neutron, and proton) without which the hypertriton would be unbound. The unequal masses of the three charge states (Σ^+,Σ^0, Σ^-), in contrast to the single mass of the Λ, were shown to lead to charge symmetry breaking (the Λp-Σ^+n interaction differs from the Λn-Σ^+p interaction) and a significant difference (~0.35 MeV) between the Λ separation energy in Λ-helium-4 and Λ-hydrogen-4. Finally, the Λ is heavier than the proton (or neutron), and it decays into a proton plus π^- (or neutron plus π^0), in analogy with the beta decay of a neutron into a proton. Because the Λπ system mixes with the Σ^+n system, it is possible for a hypernucleus with more protons than neutrons in the nuclear core (e.g., Λ-helium-4 with a nuclear core of helium-3 containing 2 protons and 1 neutron) to emit a π^+ even in the absence of any such π^+ decay by a Λ in free space. T-Division research uncovered this unique explanation for the anomalous π^+ decay of Λ-helium-4.

Because of its novel aspects, hypernuclear physics has been the focus of experimental programs at both foreign and DOE accelerator facilities. More recently, nuclear astrophysicists have suggested that "strange" matter may play an important role in stellar phenomena such as quasars and neutron stars. T-Division research is now focused upon multihyperon systems, moving out along the third dimension (strangeness), exploring what may well be the "stepping stones" to the stars.

The Nucleon–Nucleon Interaction

by Leon Heller

Beery
Beyer
Brolley
Friar
Gammel
Gibson
Gursky
Heller
Matthews
Rich
Seagrave
Signell
Thaler
Wender
Yoder

Beginning before 1950 and extending right to the present, one of the most important questions in nuclear physics has been the determination of the force that acts between two nucleons. It was known to be of short range, of the order of 1 fermi (10^{-13} centimeters), but the details concerning the strength and shape of the potential energy function became the subject of a large number of experimental and theoretical investigations.

By the 1970s a change in the theoretical outlook arose from the realization that quarks are more fundamental particles than nucleons and mesons, and one should try to understand the nucleon-nucleon interaction in terms of quantum chromodynamics (QCD), which is the theory of quarks and the forces between them due to the exchange of particles called "gluons." Gluons play a similar role in QCD theory as photons do in the theory of electrodynamics.

Early experiments in which protons were scattered elastically off hydrogen targets were limited to the energy range below 10 MeV, but they were sufficiently accurate that in addition to the large effect of the Coulomb repulsion between the two protons an additional small effect predicted by theory, called vacuum polarization, had to be included in the analysis. [1] This was especially true of a Los Alamos experiment by J. E. Brolley, J. D. Seagrave, and J. G. Beery, that was analyzed together with M. L. Gursky. [2] The key point here is that the known Coulomb scattering amplitude, due to the repulsive electrical force between two protons, and the amplitude due to the nuclear force, which is predominantly attractive at low energy, cancel against each other to a very great extent at an energy of 0.38243 MeV, and this enables one to determine the nuclear amplitude with considerable accuracy.

Analysis of the combined set of all the low-energy experiments yielded two parameters of the proton-proton (p-p) interaction called the scattering length and effective range. These two numbers contain information about the effective strength and range of the force acting between the protons. Two related issues concerned the values of the corresponding parameters for the neutron-proton (n-p) and neutron-neutron (n-n) systems. The expectation that the n-n system should have the same nuclear interaction as the p-p system is referred to as "charge symmetry," and when the n-p system is included it is called "charge independence." Since the neutron has no electric charge, one does not expect the experimental data on n-p scattering and n-n scattering to be exactly the same as the data from p-p scattering.

In order to compare those data sets the theoretical analysis of data involving neutrons must correct for the fact that there is no coulomb force. Work in T-Division by M. Rich, W. A. Beyer, and me showed that a published procedure for doing this was incorrect, and gave the right method.

The fact that free neutrons cannot be made into a target means that the n-n interaction has to be obtained indirectly, and two experiments were reported in 1964 in which a negatively charged pi-meson is captured on a deuteron and produces two neutrons and a gamma ray. The interaction of the two outgoing neutrons influences the shape of their energy spectrum. The theoretical analysis together with P. Signell and N. R. Yoder [3] showed that the value of the n-n scattering length obtained from that data is consistent with charge symmetry; but the n-p scattering length differs from the n-n value, which represents an apparent violation of charge independence.

A number of groups produced nucleon-nucleon potential energy functions that were designed to fit the elastic scattering data all the way up to energies of 300-million electron volts. It was known since the work of Yukawa that the longest range part of the interaction is due to the exchange of a virtual pi-meson, which is a purely quantum-mechanical effect in which one nucleon emits a pi-meson for an extremely short period of time, and it is very rapidly absorbed by the other nucleon. This meson is said to be "virtual," and that process was incorporated into these potentials. But the next smaller range contribution comes from the exchange of two virtual pi-mesons, which is considerably more difficult to treat theoretically. Since two pi-mesons interact with each other very strongly at a particular energy, they effectively form a single composite particle at that energy; this resonance is called the rho-meson, and the exchange of a rho-meson was included in these potentials. Still shorter range contributions to the potential were treated purely phenomenologically, and fairly good fits to the elastic scattering data were obtained. Strong repulsion at very short distances was required, and work by J. Gammel and R. Thaler in T-Division was very important in this effort.

It became clear as a matter of principle that elastic scattering experiments alone, no matter how accurate and complete, cannot decide what the correct potential is. The reason is that elastic scattering only depends on the part of the wave function that is beyond the range of the nuclear force. It is possible, however, to make a class of changes in the potential at short distances, called unitary transformations, which do not change the wave function at large distances, and therefore the effect of such changes in the potential cannot be detected in elastic scattering experiments. Bremsstrahlung, in which the colliding nucleons lose some of their energy by radiating a single photon, was proposed as the simplest inelastic scattering experiment because the electromagnetic interaction of protons and neutrons is well known. Even though the neutron has no electric charge it does have a magnetic moment, which can contribute to bremsstrahlung.

Difficult bremsstrahlung experiments of limited accuracy of both the p-p and n-p type were carried out, and comprehensive theoretical calculations were performed in T-Division and elsewhere using a potential that agreed with the elastic scattering experiments; fair agreement with most of the bremsstrahlung data was obtained. Once again, however, some theoretical issues made the pursuit of the 'true' potential a difficult undertaking. A theorem by F. E. Low says that if the bremsstrahlung amplitude is expanded in powers of the photon momentum, then the first *two* terms in that expansion are completely determined by the elastic scattering amplitude and the charges and magnetic moments of the colliding particles. Consequently, unless the photon carries off a significant fraction of the available energy, a bremsstrahlung experiment will not contain any new information beyond what was already known from elastic scattering. I summarized this matter and other issues about bremsstrahlung, [4] and this led to experiments in which the photon carries off a large fraction of the energy, including a very recent one on the n-p system performed at the Los Alamos Neutron Science Center (LANSCE) by J. L. Matthews, S. A. Wender, and others.

A second difficulty in trying to obtain the nucleon-nucleon interaction from bremsstrahlung experiments is that knowing the electromagnetic interaction of individual nucleons is not sufficient because the virtual mesons that mediate the force can also have electromagnetic interaction. For example, in n-p bremsstrahlung a virtual charged pi-meson can be exchanged between the neutron and the proton, and the meson can radiate the photon. Such a contribution is known as an "exchange current." This contribution can be calculated for the well known pi-meson, but what should be done with the shorter range contributions to the force, which were largely phenomenological to begin with? Without a theoretical understanding of the medium- and short-range parts of the potential one could not be sure what additional exchange currents should be included. [4]

Another aspect of the nucleon-nucleon interaction that has been explored in T-Division has to do with systems consisting of three nucleons. There are the bound states, tritium (hydrogen-3) and helium-3, and also scattering experiments of a neutron or proton on a deuteron (hydrogen-2). The question here is whether these systems can be understood using just forces between two nucleons at a time, or whether there are true three-body forces when all three particles are close together. The consensus appears to be that three-body forces are needed, and work by J. L. Friar and B. F. Gibson in T-Division figured very prominently in the effort to understand the three-body system.

By the 1970s a change in the theoretical outlook arose from the realization that quarks are more fundamental particles than nucleons and mesons, and one should try to understand the nucleon-nucleon interaction in terms of quantum chromodynamics (QCD), which is the theory of quarks and the forces between them due to the exchange of particles called "gluons," which play a similar role in this theory as photons do in the theory of electrodynamics. Very complex mathematically, QCD has led to a number of phenomenological and semi-phenomenological approaches, some of which have been pursued in T-Division. But the direct attack on the problem, using lattice gauge theory, has primarily been pursued elsewhere and continues to this day.

The Los Alamos Model of Neutron Emission in Fission

by David G. Madland and J. Rayford Nix

Madland
Nix

Two of the most important quantities in initiating and sustaining a nuclear chain reaction in a reactor or nuclear weapon are the energy spectrum of the emitted prompt neutrons N(E), where E is the kinetic energy of the emitted neutrons, and the average number of prompt neutrons emitted per fission $\bar{\nu}_p$. In 1979, we decided to develop a theoretical model that treats both of these quantities simultaneously. Our approach is based upon classical nuclear physics theory and it emphasizes predictive capabilities while requiring minimal input.

The Los Alamos model (sometimes called the Madland-Nix model) has been used to calculate $N(E)$ and $\bar{\nu}_p$ with confidence for neutron-induced and spontaneous fission and for unmeasured as well as measured systems in both the actinide and transactinide regions. It is now used in many laboratories throughout the world.

Two of the most important quantities in initiating and sustaining a nuclear chain reaction in a reactor or nuclear weapon are the *energy spectrum of the emitted prompt neutrons N(E)*, where *E* is the kinetic energy of the emitted neutrons, and the *average number of prompt neutrons emitted per fission* $\bar{\nu}_p$. By "prompt" here is meant during the lifetimes of the two hot (excited) fission-fragment nuclei resulting from the (binary) fission event, which range from ~10^{-20}–10^{-14} seconds. The spectrum $N(E)$ and average multiplicity $\bar{\nu}_p$ depend upon many quantities. Chief among them are:

(1) the nuclear structure and excitation energy of the fissioning nucleus and the two (nascent) fission fragments,

(2) the distribution of fission-fragment excitation energy,

(3) the motion of the fission fragments emitting the prompt neutrons,

(4) the energy dependence of the inverse process—neutron absorption, and

(5) multiple-chance fission (neutrons emitted prior to fission) at high incident neutron energy.

Furthermore, $N(E)$ and $\bar{\nu}_p$ are related. Therefore, the measurement of one of them assists in the calculation of both of them. This fact is very important because whereas $N(E)$ is very difficult and expensive to measure accurately, $\bar{\nu}_p$ is less difficult and less expensive to measure. More importantly, the predictive power of a theoretical model for both quantities is clearly superior to that for only one of them.

However, early attempts to calculate the two quantities did not make use of the fact that they are related. Two representations of the spectrum $N(E)$ that were used are the Maxwellian spectrum, with a single (temperature) parameter T_M appearing, and the two-parameter Watt [1] spectrum (which is a center-of-mass Maxwellian spectrum transformed to the laboratory system for the motion of an average fission fragment [2]), again with a single temperature parameter T_W appearing, but also the average kinetic energy per nucleon E_f of the average fragment appearing. At the same time that these early representations were introduced for $N(E)$, the average prompt neutron multiplicity $\bar{\nu}_p$ was modeled [3] by a simple polynomial (usually linear) in the incident neutron energy E_n, for each fissioning system considered: $\bar{\nu}_p = \nu_0 + \alpha E_n \bar{\nu}_p$, and again, the two parameters appearing were adjusted to optimally reproduce the experimental values. The approach is completely empirical.

To summarize, none of the methods just described can be used to predict $N(E)$ and/or $\bar{\nu}_p$ for a different fissioning nucleus or for a different excitation energy from what has been experimentally measured. That is, there is very little, if any, predictive power here.

We therefore decided, in 1979, to develop a theoretical model that treats simultaneously the prompt fission neutron spectrum $N(E)$ and the average prompt neutron multiplicity $\bar{\nu}_p$. Our approach [4] is based upon standard nuclear evaporation theory [5] and it emphasizes predictive capabilities while requiring minimal input. The original Los Alamos model [4] addresses both spontaneous fission and neutron-induced single- and multiple-chance fission. The model explicitly accounts for all five of the dependencies (1) through (5) listed above.

The Los Alamos model prompt fission neutron spectrum in the laboratory system is given by

$$N(E) = \frac{1}{2}\left[N\left(E, E_f^L, \sigma_c^L\right) + N\left(E, E_f^H, \sigma_c^H\right)\right] ,$$

where

$$N\left(E, E_f, \sigma_c\right) = \frac{1}{2\sqrt{E_f}T_m^2} \int_{\left(\sqrt{E}-\sqrt{E_f}\right)^2}^{\left(\sqrt{E}+\sqrt{E_f}\right)^2} \sigma_c(\varepsilon)\sqrt{\varepsilon}\, d\varepsilon \times \int_0^{T_m} c(T)\, T \exp(-\varepsilon\, T)\, dT \ .$$

In these equations, E_f^L and E_f^H are the average kinetic energies per nucleon in the average light and heavy fragments, respectively, σ_c^L and σ_c^H are the corresponding probabilities for the inverse process of neutron absorption in these fragments (obtained using an optical potential with explicit isospin dependence), T is the fission-fragment temperature and T_m its maximum value, and ε in the center-of-mass kinetic energy of the emitted neutron. Finally, $c(T)$ is the normalization integral.

The numerical solution of this equation in the exact Los Alamos model is obtained using Gauss-Legendre and Gauss-Laguerre quadrature integration, whereas for the approximate Los Alamos model (σ_c = constant) one obtains a four-term closed-form expression involving the exponential integral function and the incomplete gamma function.

The Los Alamos model average prompt neutron multiplicity $\bar{\nu}_p$ is obtained from energy conservation and is given by

$$\bar{\nu}_p = \frac{aT_m^2 - \left\langle E_\gamma^{\mathrm{tot}} \right\rangle}{\left\langle S_n \right\rangle + \left\langle \varepsilon \right\rangle} ,$$

where a is the nuclear level-density parameter, $\left\langle E_\gamma^{\mathrm{tot}} \right\rangle$ is the total average prompt gamma-ray energy, $\left\langle S_n \right\rangle$ is the average fission-fragment neutron separation energy, and $\left\langle \varepsilon \right\rangle$ is the average center-of-mass energy of the emitted neutrons.

There are two important connections between $N(E)$ and $\bar{\nu}_p$ in the Los Alamos model [6]. The first is that the maximum temperature T_m appearing as one of three parameters in $N(E)$ also appears in the above expression for $\bar{\nu}_p$ as T_m^2. The second is that the average center-of-mass neutron energy $\left\langle \varepsilon \right\rangle$ appearing in the expression for $\bar{\nu}_p$ is also the first moment of the center-of-mass neutron spectrum $\Phi(\varepsilon)$ corresponding to the laboratory neutron spectrum $N(E)$. These two connections are very significant because they mean that if one has experimental information on either $N(E)$ or $\bar{\nu}_p$, then that information can be used as a constraint in the calculation of the other, unmeasured, observable. We note here that there are many more measurements of $\bar{\nu}_p$ than there are of $N(E)$.

The Los Alamos model has been used to calculate $N(E)$ and $\bar{\nu}_p$ with confidence for neutron-induced and spontaneous fission and for unmeasured as well as measured systems in both the actinide and transactinide regions [4,6]. It is now used in many laboratories throughout the world. By way of example, we show in Fig. 1 (energy units are MeV) the recently completed calculation of the prompt fission neutron spectrum matrix for the n + ^{239}Pu system.

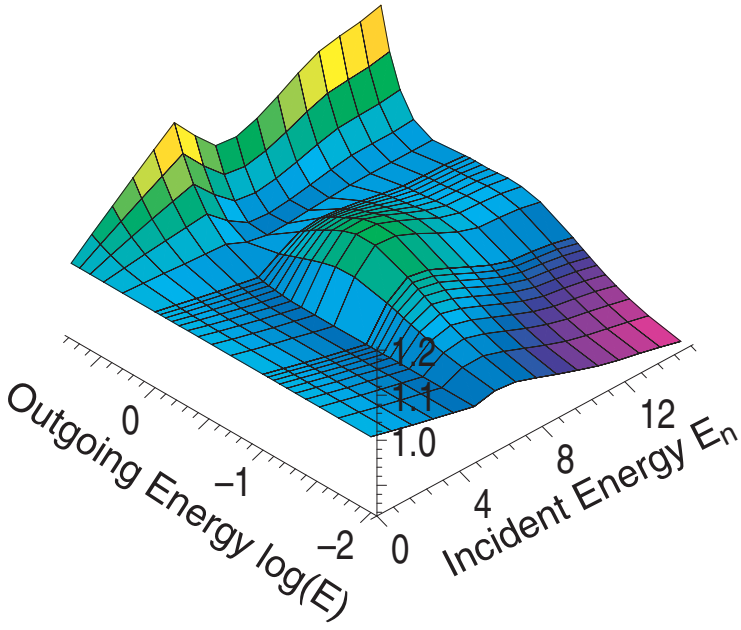


Fig. 1. Prompt Fission Neutron Spectrum Matrix for the n + ^{239}Pu system (divided by the thermal spectrum).

Dynamics of Atomic Nuclei

Gibbs
Ginocchio
Haxton
Kirson
Stephenson
Strottman

by Joseph N. Ginocchio

Although it has been known for 50 years that pairing correlations are important in atomic nuclei, not until 1975 has the importance of quadrupole pairing correlations been recognized in addition to the well-known monopole pairing correlations. Reconciliation of monopole and quadrupole pairing correlations with, on the one hand, the long-standing liquid drop model of collective motion for which the Nobel prize had been awarded in the same year to A. Bohr and Ben Mottelson, on the other hand, the nuclear shell model in which nucleons move in spherical orbits was the outcome of research in Theoretical Division in the late seventies and early eighties. Pairing correlations also explained puzzles observed in pion double exchange. Also, at this time there was renewed interest in the use of nuclear double beta decay as a test of lepton number violation. State-of-the-art shell model calculations with realistic correlations led to laboratory counting experiments which have resulted in the actual measurement of the longest-lived radioactive transmutations ever observed and to a major review of the double beta decay and its implications for lepton number conservation. More recently certain regularities of shell model orbits were shown to be due to a relativistic symmetry and research continues to determine if this symmetry has a fundamental rationale in the nucleon-nucleon interaction and the behavior of quarks in nuclei.

In 1975 a new model of collective motion in nuclei had begun to gain prominence. [1] The interacting boson model approximated pairs of nucleons as bosons and further assumed that those with angular momentum zero and two were the most dominant for the low-energy spectroscopy of nuclei.

Two questions naturally arose. First, how does one reconcile this model based on paired nucleons with the long-standing liquid-drop model of collective motion for which the Nobel prize had been awarded in the same year to A. Bohr and Ben Mottelson? Second, are there nuclear shell- model Hamiltonians that have eigenstates composed only of pairs of nucleons coupled to angular momentum zero and two?

M. Kirson, on sabbatical from the Weizmann Institute of Physics, Israel, and I resolved the answer to the first question. [2] We introduced a boson intrinsic coherent state that was a function of collective deformation variables which could be related to the deformation parameters of the liquid-drop model. As for the second question, I showed that there was a class of shell model Hamiltonians which had a subset of eigenstates that were composed entirely of pairs of nucleons coupled to angular momentum zero and two, and that the remaining eigenstates

could be classified by the number of pairs of nucleons coupled to angular momentum zero and two. [3] Both of these works had a significant impact.

In the late 1970s, there was renewed interest in the use of nuclear double beta decay as a test of lepton number violation. This was driven by developments in particle physics, particularly the seesaw mechanism for small but nonzero neutrino mass, which strongly suggested that the neutrino was a Majorana particle. The most direct evidence for this would be the existence of neutrinoless double beta decay. At that time the evidence was contained in studies of the relative abundance of possible daughter products and the analyses usually assumed that the unknown nuclear matrix elements could be taken as equal for the two processes and factored out. Given the difference in the structure of the operators, that assumption was dubious at best. With this in mind, Haxton, Stephenson, and Strottman undertook large shell model calculations for some cases of interest, particularly the decays of tellerium-128, -130. [4] These calculations strained the technical abilities of the time, requiring separate diagonalizations of the neutron and proton spaces followed by a weak coupling approximation. The results, while state of the art then, were not capable of predicting the two neutrino branch, and spawned a cottage industry in the theoretical study of double beta decay matrix elements which persists today. These calculations led to a strong collaboration with the experimental community which planned, and subsequently carried out, laboratory counting experiments that have resulted in the actual measurement of the longest-lived radioactive transmutations ever observed and to a major review of the field.

Experiments at Los Alamos Meson Physics Facility in which positive pions were scattered from nuclei and outcoming negative pions were detected leaving a residual nucleus with two more protons and two less neutrons showed some surprises. These reactions were expected to populate primarily the excited state in the neighboring nucleus, which had the same spatial properties as the target nucleus ground state. For high-energy pions this expectation was realized but low-energy pions populated the ground state of the neighboring nucleus with equal or greater intensity. W. R. Gibbs and I with N. Auerbach from Tel Aviv University, Israel, and W. B. Kaufmann from Arizona State University explained this unexpected result by showing the pion reaction was sensitive to spatial correlation between the nucleons at low-pion energies whereas the reaction was not sensitive to the correlations at high energy and it was this sensitivity which caused the transition to the ground state. [5]

A long-standing puzzle in nuclear spectroscopy has been the fact that certain states in nuclei, which have different orbital angular momentum and radial wave functions, have almost the same energy. About thirty years after the observation of these systematics, I discovered that these

quasi-degeneracies are due to a small breaking of a relativistic symmetry of the Dirac Hamiltonian. [6] This finding came as a surprise since the dynamics of nucleons in nuclei are dominantly nonrelativistic. This symmetry has led to predictions relating the nuclear eigenfunctions of these quasi-degenerate states and, in addition, to predictions of magnetic dipole and Gamow-Teller transitions between these quasi-degenerate states, which have proven to be approximately valid. Furthermore, nucleon-nucleus scattering approximately conserves this relativistic symmetry. Relativistic mean field approximations in nuclei and relativistic optical models of nucleon-nucleus scattering substantiate the fact that this symmetry is a relativistic symmetry of the Dirac Hamiltonian. Also investigation of nuclear matter at the quark level also indicates the existence of this symmetry, motivating the ongoing search for a more fundamental rationale for this symmetry. [7]

Macroscopic Dynamics of Nuclear Fission and Fusion

by Arnold J. Sierk

Cohen
Davies
Krappe
Möller
Nix
Sierk
Strottman
Swiatecki

For 20 years, beginning in 1968, Theoretical Division was the world leader in the development and applications of dynamical models of fission and fusion and investigations of models for nuclear dissipation. The papers on those subjects written at Los Alamos during that period have received more than 2,500 citations in the refereed journal literature. Several T-Division authors have contributed articles in this chapter that discuss the diverse research areas begun in the division by Ray Nix. This article focuses on the pioneering developments in the understanding of the dynamics of the fission process and what is thought to happen when two nuclei collide at lower energies.

Introduction

During the 30 years in which Ray Nix was a staff member of the Theoretical Division at Los Alamos Scientific Laboratory (later Los Alamos National Laboratory), he authored papers that have received more than 6,000 citations. The articles in this book written by Madland and Nix (See "The Los Alamos Model of Neutron Emission in Fission" in this chapter.), Möller and Sierk (See "A Unified Theory of Nuclear Masses, Fission Barriers, and Superheavy Nuclei" in this chapter.), and Strottman (See "Heavy Ions at Los Alamos National Laboratory" in this chapter.) outline developments in some of the disparate fields in which Nix worked: the neutron spectra of fissioning nuclei, the calculation of potential energy surfaces, fission barriers and nuclear ground-state masses and deformations, and the modeling of relativistic collisions of nuclei, respectively. In this article, I will discuss pioneering developments in the understanding of the dynamics of the fission process and what is thought to happen when two nuclei collide at lower energies.

As is discussed in the article by Möller and Sierk, fission is a process that involves large shape changes of a nucleus, from the spherical or slightly deformed nuclear ground state, through extreme deformations including the formation of a necked shape, finally to separated nuclei, which are referred to as fission fragments. The point at which the neck radius vanishes and the previously single nucleus first is separated into two pieces is called the scission point. The first step to understanding the fission process, as is discussed in that article, is to determine how the binding energy, or equivalently the mass of the nucleus, depends on its shape. The negative of this binding energy is referred to as the potential energy of the nucleus; the nucleus will try to deform in a direction that decreases its potential energy (increases binding). For any nucleus, the potential energy always rises as it is deformed away from the ground state.

One may understand the situation qualitatively by imagining being at the shore of a mountain lake. The land rises above the shore of the lake in all directions, but depending upon which direction you go, you may pass over a mountain, or find an outlet at a lower elevation, corresponding to a pass, after which the ground slopes downward. Such a geographical pass is often referred to as a saddle, because the curvature of the ground in its neighborhood resembles a saddle: curving upward in one direction, while curving downward in a perpendicular direction. The pass out of the lake basin with the lowest height is analogous to the fission saddle point. In this case, if the basin were to have large amounts of water added, the lake surface would rise until the water began to spill out of this pass. Thus the potential energy in this direction rises until the saddle point is reached, after which it again decreases. For the nucleus, the difference between the height of the saddle point and the ground-state energy (bottom of the lake) is called the fission-barrier height. The fission-barrier height determines the minimum energy that needs to be added to the nucleus to get it to fission in a short time. Of course, quantum-mechanical effects allow for fission to occur without any added energy through "barrier penetration." This occurs on a much longer time scale, on the order of 10^{14} years for uranium-238.

However, understanding the nature of fission or any other dynamical process does not depend only upon the potential energy. The forces tending to change the deformation of the nucleus depend on the slope of the potential energy surface, being proportional to the gradient vector. The motion of an object will depend on its mass (inertia), as well as the force applied. For example, one person pushing on a small car will make it move faster than he would a truck using the same force. So, clearly the time that it takes for a nucleus to fission will depend not only on the potential energy surface but also on the inertia. A simple analogy is an object hanging from and bouncing at the end of a spring. The spring and gravity provide forces on the object, but the frequency of the oscillations depends not only on the properties of the spring but also on the mass of the object. For the same spring, a heavier object will move more slowly than a lighter one.

Another important aspect of dynamical systems is dissipation. The object on a spring would behave differently if it were immersed in water or oil; the former probably leading to decaying oscillations, and the latter likely to damp movement so strongly that if the object is displaced from its equilibrium point, it will only relax to that point without oscillating. The word dissipation is used because the energy of the spring and mass disappears or is dissipated into heating the water or oil a tiny amount. So, to understand the motion of a fissioning nucleus, we need not only to understand the potential energy but also the inertia and the dissipation.

Dynamics of Fission and Low-Energy Heavy-Ion Collisions

Ray Nix's first major publication after joining T-Division in 1968 was a pioneering effort in studying the dynamics of the fission process in the framework of the liquid-drop-model (LDM). [1] In addition to introducing the most realistic nuclear shape parameterization yet devised for the description of nuclear shapes at the fission barriers and beyond and calculating for the first time with reasonable accuracy the fission barriers for nuclei with atomic numbers less than 40, this work also contained a careful dynamical study of distributions of kinetic energies and masses of fission fragments, starting from the assumption that the inertia of the deforming nucleus was that due to incompressible fluid flow.

A few years later in 1972, when Arnold Sierk joined T-Division as a postdoctoral fellow, this work was extended to address the formation of heavy nuclei by the fusion of medium-mass nuclei in a collision of heavy ions, which is the term applied to atomic nuclei that have some of their atomic electrons removed. These ions can be accelerated by electromagnetic fields in particle accelerators and then made to collide with other atomic nuclei. When given enough energy, these ions can overcome the electrostatic repulsion of the protons in the nuclei to the point where the nuclei can be brought close enough together so that the short-range nuclear forces begin to attract them to each other. At this energy, it becomes possible for nuclear reactions to occur with a reasonable probability. This energy acquired the name of interaction barrier, because at this energy the nuclei began to interact in such a way as to produce nuclear reactions. The interaction barriers scale in a very simple and well-understood way as the masses of the colliding nuclei increase.

At that time, the possibility of the existence of superheavy nuclei (See article by Möller and Sierk.) motivated the study of how to get lighter nuclei to join together or fuse to form such large atomic nuclei. For lighter nuclei, the cohesive nuclear force dominates the disruptive Coulomb force. However, as the size of the system increases, the Coulomb energy increases faster than the nuclear energy (See Eq. (1) in the article by Möller and Sierk). This not only makes fission of heavy nuclei possible but also makes it relatively harder for nuclei to fuse as their masses and atomic numbers increase. It happens that nuclei need to fuse to a more compact shape than the fission saddle point of the combined system in order to have a chance to lose enough energy through neutron and gamma-ray emission so that a stable combined nucleus might be formed. For lighter nuclei, for which the fusion barrier is essentially the same as the interaction barrier, the saddle point is more deformed than the contact point of the two nuclei.

As the mass of the total system increases, because of the more rapid increase of the Coulomb force than the nuclear force, the fission saddle point becomes more and more compact, until for nuclear mass numbers in the neighborhood of A = 220 to 230, the saddle point is more compact than the two colliding nuclei when they first make contact. If the colliding nuclei are brought together with just enough energy to make contact

with no relative motion, the strong Coulomb forces will drive the system to rapid fission, with no fusion occurring. This means that once one tries to make a nucleus heavier than this by heavy-ion fusion, an additional energy must be imparted to the colliding nuclei to drive them together to a more compact configuration than the saddle-point shape. This additional energy is termed a dynamical threshold energy. As discussed previously, the nature of the motion of the system after the two nuclei have been brought into contact will depend not just on the forces acting but also on the nature of the nuclear inertia and dissipation.

The first publication from the Nix-Sierk collaboration [2] contains the first dynamical calculation of fusion thresholds assuming hydrodynamical inertia and no dissipation. This early calculation showed that such thresholds occur even when no dissipative mechanisms are present, in this case solely as a result of the interaction between the potential energy and the inertia of the nucleus. This calculation of dynamical thresholds was a decade in advance of the rest of the world.

This paper also contained the first calculation of the self-consistent dynamics of a dissipative fissioning system. [2] For simplicity, this first calculation only was able to consider the effect of dissipation on spheroidal fission fragments as they separate during the fission process, showing that high dissipation increases the kinetic energy acquired by the fragments after scission. Soon after, in collaboration with Tom Davies, a visitor to T-Division from Oak Ridge National Laboratory, dissipation was introduced into the complete dynamical calculations of fission. [3] This work was unmatched for more than a decade, until other groups began addressing the problem of fission in more limited shape parametrizations, but with the addition of fluctuation effects in the dynamics, starting about 1990.

This early work on nuclear dissipation started with a very simple assumption: that the dissipation was due to ordinary viscosity, of the same sort affecting the motion of familiar fluids like water or honey. The fundamental mechanism of ordinary viscosity involves the energy transported by the collisions of the molecular constituents of the fluid. Because these collisions occur between two atoms or molecules, hydrodynamic viscosity is sometimes called two-body dissipation. The success of the nuclear shell model (See the article by Möller and Sierk.) implies nucleons inside the nucleus appear to move as though they interact mostly with an effective potential due to the average effect of all the remaining constituents of the nucleus, with only small residual interactions with individual nucleons. Such an effective potential is called a one-body potential, because the nuclear energy levels are approximately those that quantum mechanics predict for a single nucleon moving in the potential.

Such ideas led in the mid to late 1970s to the idea of one-body dissipation, which may be thought of as the interaction of individual nucleons with the moving boundary or “wall” of the nucleus. Interaction with a

moving potential wall leads both to a conservative spring-like force (See [6].) and to a dissipative force affecting the motions of the nuclear shape, as this motion couples energy into the motion of the individual nucleons inside the nucleus. With some simplifying assumptions, this idea leads to the one-body "wall" dissipation model. A consideration of the transport of momentum between collision partners or nascent fission fragments by transport of particles through the neck connecting the ions or fragments leads to the "window" dissipation models.

Nix and Sierk collaborated with Lawrence Berkeley Laboratory (LBL) scientists on calculating the consequences of the so-called "wall-and-window" model for nuclear fission. [4] This one-body dissipation model turns out to lead to very strongly damped motion. The dissipation is so strong that oscillations are not possible, which is called overdamped motion, for which the inertia is not very important. The interaction of this dissipation with the dynamics of fission leads to scission configurations that are more compact than those for no dissipation. [1,4] Two-body dissipation, on the other hand, leads to motion that is slower than nonviscous but with more elongated scission shapes. For one particular value of the viscosity coefficient (the two-body model requires one empirical parameter), both the two-body and the one-body dissipation models can predict the same values for the most important fission observable they affect, the experimental kinetic energy of the fragments. This occurs because they have different amounts of both pre-scission kinetic energy and deformation energy, which compensate each other.

Because of deficiencies in the simplifications and parametrization of the pure LDM, it is not possible to learn about dissipation from fission-fragment energies, as in this model the predicted kinetic energy is less than that measured, even for no dissipation. [1] During the 1970s Nix and Sierk, with a collaborator Hans Krappe from the Hahn-Meitner Institute in Berlin, developed an improved semiphenomenological model for the nuclear surface energy, which is in the spirit of the LDM, but also takes into account the diffuseness of the nuclear surface and the finite range of the nuclear force.

This model gives rise to the approximation of a surface energy in the limit of very large nuclei, [5] while giving an appropriate reduction of the surface energy for small nuclei and leading to a realistic nuclear attractive force between nuclei. This formulation for the shape-dependent nuclear energy is able to describe in a single simple framework with only three adjusted parameters (in comparison to two for the LDM) such previously disjointed phenomena as the optical-model potentials needed to describe the elastic scattering of heavy ions, the systematic behavior of nuclear interaction potentials, the heights of fission barriers, and the macroscopic contribution to the masses and shapes of nuclei throughout the periodic table (See the article by Möller and Sierk.).

With this more realistic model for the macroscopic nuclear energy, it is possible to solve a realistic dynamical model for fission and to compare the most important observable, the sum of kinetic energies of the fragments formed in fission, to systematic experimental measurements. It turns out that the extreme one-body dissipation discussed previously leads to too small fragment kinetic energy for heavy nuclei, while a good representation of the data for the entire periodic table can be found for a single value of the two-body viscosity coefficient. However, the two-body model is unsatisfactory for theoretical reasons.

In 1980, a simple physical model for giant isoscalar multipole resonances of nuclei was developed. [6] These resonances approximately correspond to high-energy shape oscillations of the nucleus as a whole with the volume being conserved. This simple model gives a parameter-free prediction of the energies of the quadrupole (spheroidal) and octupole (egg-shaped) resonances, which fits very well the average behavior of these energies for nuclei throughout the periodic table. In addition to their energies, these nuclear resonances also exhibit widths, which correspond to finite lifetimes of the resonant states.

When we applied the simple dissipation models discussed above to calculating these decay lifetimes, we found neither of these dissipation pictures gave reasonable results. This situation prompted careful thought about a more realistic dissipation model. Because of the success of the nuclear shell model, it is clear that one-body effects are important in nuclei. There are also good theoretical reasons to expect that dissipative two-body interactions occur in dynamical systems. Qualitative considerations and the success of the nuclear optical model with absorptive potentials leads to the conclusion that both one- and two-body effects should be concentrated near the nuclear surface.

Because there is not yet a convincing model for the details of the relative sizes of the two effects, in the spirit of the liquid-drop model, we proposed a one-parameter model, called Surface-plus-Window dissipation. The one parameter was chosen to fit the widths of the isoscalar giant quadrupole resonances. [7] It leads to a prediction of the octupole widths that slightly underestimates the actual ones, by about 15% for spherical nuclei, while exhibiting the correct mass dependence for both. This is to be compared to the pure two-body dissipation, which underestimates the widths significantly, and has the wrong dependence with nuclear mass, and the extreme one-body "wall" dissipation, which gives widths a factor of 4 too high.

The Surface-plus-Window dissipation model was then applied to fission-fragment kinetic energy measurements. [8] When the model, with no readjustment of parameters, is compared to experimental data, we find that the average uncertainty-weighted mean-square deviation of the measured energies from the calculated ones is within 15% of the deviation from a semiempirical 2-parameter fit which has seen much use in the community. Just as the finite-range nuclear energy model [5] brought together several diverse phenomena into a simple model with only three empirically determined parameters, so this dissipation model explains fission dynamics with its one parameter determined

from consideration of giant multipole widths. This model has also been applied to the calculation of dynamical thresholds to heavy-ion fusion, giving results that are qualitatively more in keeping with experimental results than are other mechanisms. Because of political changes in the Lab and T-Division occurring in the late 1980s, it was not possible to continue this work until some effort was resumed in 2001. In that time, no one has developed an alternative dynamical and dissipation model that is as consistent with the observed fragment-kinetic-energy and other data.

Fission Barriers and Rotating Nuclei

As we have discussed previously, for the LDM and for the finite-range nuclear energy model (FRLDM), the ground state of nuclei lighter than about A = 300–320 is a sphere. The increasing effect of the repulsive Coulomb energy for nuclei with large atomic numbers leads to an instability that causes the sphere to no longer be stable for systems heavier than this. In a similar manner, for lighter nuclei, adding angular momentum (increasing the rate of spin) will also cause the nucleus to lose stability. For angular momentum the forces, unlike Coulomb forces are not equal in all directions, because the axis about which the nucleus rotates adds an asymmetry to the problem. The study of systems with large amounts of angular momentum, which had begun in LBL in the 1950s, began taking off in the 1970s as many heavy-ion accelerator facilities were constructed around the world. With these facilities, nuclei with high spins were produced by colliding two lighter nuclei. Understanding how the shapes of nuclei change with angular momentum and the effects on the fission barrier and the rotational moments of inertia became of great interest.

The problem of determining the stable shapes of rotating systems has been around for a long time. A simple (not too realistic) early model for stars and planets was a self-gravitating liquid drop of constant density. Because gravity obeys the same $1/r^2$ force law as the Coulomb force, this problem is identical to the nuclear LDM with no surface tension, except the sign of the force is attractive, instead of repulsive. In the 18th century, Maclaurin showed that rotating constant-density gravitating liquid drops were oblate spheroids, with the axis of symmetry coinciding with the rotational axis. In the 19th century, Jacobi was able to prove that above a critical angular velocity or angular momentum, the spheroid became unstable to a triaxial deformation, leading to an ellipsoid whose longest axis is perpendicular to the rotation axis. As the rotation is increased, the ellipsoid becomes close to a prolate spheroid but is slightly flattened along the rotational axis. As the rotation is increased still further, Poincaré showed around the turn of the 20^{th} century that another instability occurs, in which the Jacobi ellipsoid tended to form into lobes in which one side had more mass than the other (octupole deformation). For this gravitational problem, there is no analog to the fission barrier, because the gravitational forces try to keep the system together, whereas in the nuclear LDM, the surface energy tries to form a compact shape, and the Coulomb energy tries to spread the charge over a more extended volume.

The corresponding problem in the nuclear LDM began to be addressed in the 1950s and 60s. In this case, the equilibrium shapes, although similar to the Maclaurin spheroids, are not exactly spheroidal. This led to various approximate treatments that tried to determine the onset of the analog of the Jacobi triaxial instability, with only approximate success. The nuclear problem is much more interesting than the corresponding gravitational one in that the nuclear saddle-point shapes and energies also change with rotation. Eventually, there is a limiting angular momentum above which there is no fission barrier, and both the ground state and the saddle point no longer exist. A system with more angular momentum than this has no stable equilibrium point and flies apart immediately upon being formed.

It is interesting to determine this maximum angular momentum that a particular nucleus can sustain. In the LDM, two qualitatively different situations occur. For light nuclei, the ground state evolves with increasing angular momentum through oblate axially symmetric but nonspheroidal shapes analogous to the Maclaurin spheroids through a triaxial instability to prolate triaxial shapes which may even form a neck in the middle for high spins, before becoming coincident with the saddle shape and disappearing. The saddle-point shape, on the other hand, starts out as a shape symmetric about an axis, with a relatively small neck region in the center. As the angular momentum increases from zero, the shape flattens slightly, becoming triaxial immediately, with the rotation axis perpendicular to the axis of (almost) symmetry.

As the spin increases, the neck radius grows, and the extension of the shape shortens, with the triaxiality very slowly increasing, until the ground state and the saddle are identical, with still not very much triaxiality. For heavy nuclei, the saddle starts with a large neck radius, or no neck at all. As the angular momentum is increased, the flattening along the axis of rotation happens more strongly than for light systems, and the shape shortens until the two axes perpendicular to the rotational axis become identical just as the fission barrier disappears. The ground state remains oblate and symmetric about the axis of rotation for all values of spin. There exists a critical value of nuclear mass at which a transition occurs in the shape of the ground state having the maximum possible spin for a particular nucleus. For heavier nuclei, this shape is oblate and axially symmetric; for lighter ones it is triaxial. This transition point in mass number was uncertain by about 50 nucleons prior to our work. [8]

There is a separate phenomenon and transition that also occurs. For heavy nuclei, the fission saddle point is stable to mass-asymmetric distortions. For very light nuclei, the saddle-point shape in unstable to this mode. The result of this instability is that symmetric fission is favored for heavier nuclei, unless one is looking at actinide nuclei at low excitation energy and low spin. (See the article by Möller and Sierk.) For very light nuclei, extremely asymmetric fission (the emission of protons, neutrons, alpha particles, etc.) is favored. This transition thus has observable consequences, unlike that discussed in the

previous paragraph. This transition was calculated quantitatively by Cohen and Swiatecki at LBL in the early 1960s.

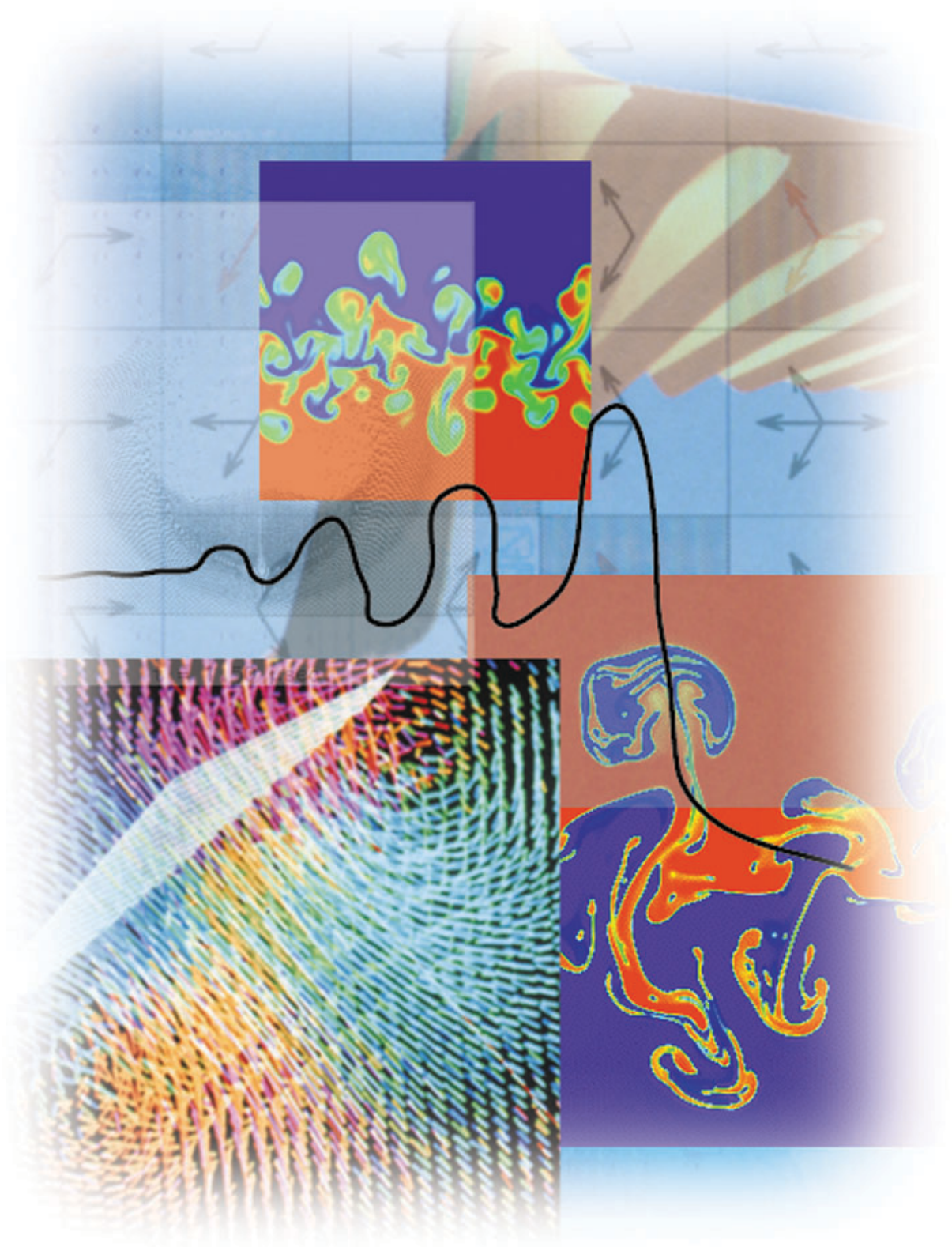
In addition to the intellectual interest of these problems, large amounts of data on fission barrier heights of lighter nuclei at high angular momenta began to be measured starting in the 1960s, continuing into the current century. After a while, it became clear that the LDM, with parameters appropriate to the masses of nuclei and to the fission barrier heights of heavy nuclei (See the article by Möller and Sierk.), was totally unable to predict the barrier heights of nuclei with mass number A < ~ 210. In addition to the finite-range effects on the nuclear energy, [5] it is known that the nuclear matter and charge is spread out over space, unlike in the sharp-surfaced LDM. Such extended matter distributions lead to a reduced rotational energy for a specified angular momentum. These considerations led to the Rotating-Finite-Range Model (RFRM), [8] which took the parameters determined in the 1981 nuclear mass model of Möller and Nix (See the article by Möller and Sierk.) and the empirical diffuseness of the nuclear charge and matter and found an excellent agreement with the measured fission-barrier heights at all angular momenta for systems for which a macroscopic model is appropriate. These include almost all systems made in heavy-ion collisions, because of the high temperatures and angular momenta occurring in such reactions. The shapes of ground states and saddle points behave qualitatively just like those of the LDM as described previously, with the saddle points in this model behave qualitatively for a specific nucleus being somewhat more compact in the RFRM.

This fission-barrier model [8] is used in all the statistical models employed in labs throughout the world studying fission and fusion. In the process of developing this model, we also determined precisely the location of all the instabilities in the LDM (as well as in the RFRM), [8] solving numerically the analogs of the classical Maclaurin/Jacobi/Poincare' problems.

In an outstanding series of experiments begun in the 1980s and continuing to the present day, researchers at LBL have measured fission barriers for very light nuclei. They find that the calculated barriers in the original RFRM [8] are slightly (1 to 4 MeV out of 30 to 40) too low for systems with A = 70 to 120. We have recently discovered that by considering the previously ignored shape dependence of an already existing term in the macroscopic part of the nuclear mass model (See the article by Möller and Sierk.), we can slightly modify the RFRM to remove this discrepancy, along with improving the reproduction of the nuclear masses and the fission barriers of nonrotating nuclei (See the article by Möller and Sierk.).

Summary

For 20 years, beginning in 1968, group T-9 was the world leader in the development and applications of dynamical models of fission and fusion and investigations of models for nuclear dissipation. The papers on those subjects written at Los Alamos during that period have received more than 2,500 citations in the refereed journal literature.



Numerical Analysis and Algorithms

In science and technology and in related fields, such as economics, sociology, and even anthropology, the ability to solve problems numerically by high-speed computers has become essential to the accomplishment of most research projects. The scope of applications is so broad, even within the Los Alamos National Laboratory's Theoretical Division, that a proper description of the major aspects would require many volumes. Numerous professional journals are dedicated to the description of new numerical techniques for solving problems. As described in the chapter, "Computers and Computing," the computers themselves have evolved in an absolutely astonishing fashion, beginning in the late 1940s and expanding to a proliferation of both supercomputers and desktop computers all around the world.

The early days of computer development were greatly influenced by the activities within T-Division at Los Alamos. The incentive was to be able to solve problems of great national importance, many of which were intractable by the classic techniques of analysis. But having the computers alone would be almost useless. For the attainment of meaningful results, techniques had to be developed by which the laws of nature could be expressed in ways that the computer can understand and use.

Nowhere does the adage, "garbage in, garbage out," apply more appropriately than to the field of computing. With computations, the input of garbage means the specification of faulty data and/or improper techniques for data manipulation. Even when those problems have been seemingly avoided, numerical instabilities can develop that reduce the results to nonsense—in a manner that resembles experimental instabilities (as, for example, with a loose bearing on a wind-tunnel fan that vibrates so hard as to destroy the apparatus). Numerical inaccuracies are more insidious; they may not result in reduction of the answers to obvious nonsense, and therefore are capable of leading an investigator to false conclusions that result in severe loss in the understanding of what really occurs in nature.

During the 1950s, the people in T-Division were sometimes criticized for playing games with the toys only a rich laboratory could afford. The results of their calculations were said to produce pretty pictures but to furnish little or no insight into basic scientific understanding, whereas the attainment of analytical solutions can reveal correlations, scalings, and universalities that are hidden in the mass of numerical output from a computer run.

Indeed, for some years, the Soviet scientists took much pride in their ability to push analysis to astonishing lengths, as a means for getting answers to hard problems in the absence of computing power. In more recent years, the scientific world has come to accept the complementary combination of analysis and computing in the pursuit of meaningful answers to an inexhaustible supply of fascinating questions.

Almost every discipline described in this book has relied, for much of its progress, on numerical solutions to hard problems. Each application has had to face the challenge of developing or extending numerical techniques for extracting meaningful results. In this chapter, we emphasize the T-Division role in developing the world-recognized foundations for some important types of numerical problem solving.

The Birth of Computer Fluid Dynamics

by Francis H. Harlow

Amsden
Butler
Daly
M. Evans
Fromm
Gentry
Harlow
Hirt
R. Martin
Meixner
Richtmyer
Shannon
von Neumann
Welch
Young

In September 1959, Carson Mark, leader of the Theoretical Division recognized the enormous growth in potential for fluid-dynamics calculations with the newly arriving supercomputers. As a result, he formed a new group to develop numerical techniques that could exploit the power of the big machines in ways that never before had been dreamed possible.

Already, members of the new group had developed an innovative idea called the Particle-in-Cell (PIC) method for calculating the highly contorting dynamics of materials interacting with each other. That original small group soon grew to include additional people who shared the vision of discovery, from which Los Alamos Scientific Laboratory gave birth to the new science of Computer Fluid Dynamics (CFD).

There had, of course, been earlier numerical-methods developments in T-Division and elsewhere, enabling the solution of some types of material dynamics problems. Members of the new group, however, realized that major problems still existed, which computer scientists had no capability to solve. Here we describe some of the highlights of discovery accomplished by that group.

Introduction

The concept of a fluid includes any material that flows, for example a liquid or a gas responding to pressure gradients that produce large contortions of the material. More generally, the behavior of any substance can be fluid—if the driving forces are sufficiently large.

With fine enough resolution, the dynamics can be seen to consist of the motion of discrete entities: atoms, molecules, or small-scale inhomogeneities such as particles of dust in the wind or crystal grains in a metal. For many investigations, however, describing the processes in such detail is not necessary.

We are concerned instead with the behavior of a continuous material with constitutive relations that describe the bulk-material response to appropriate averages of the driving forces. For example, the driving pressures within a gas can usually be calculated from the density and temperature by means of an equation of state (EOS). Likewise, the heat energy can be related to temperature and density by means of a thermal EOS.

The material properties are treated as though the variations of velocity, pressure, density, temperature, and other descriptive quantities are completely smooth. A computer, however, has no way to calculate the behavior of a continuous material. The computer can only keep in its memory a finite set of numbers. The analogy can be drawn to the half-tone pictures in a newspaper, which are constructed of enough dots to give a discrete but meaningful representation of the image.

Not only spatial representation needs division into discrete points, so also does the representation of time advancement require time steps of finite duration (like the frames of a motion picture). Some techniques are based on calculating the evolution of coefficients of a mathematical expansion in terms of continuous fields, in which case the process is constrained to the use of only a finite number of terms. In any case, the subdivision of space and time into finite segments is necessary, but purely fictitious. An ideal numerical technique for describing the behavior must give results that are essentially independent of the spatial and temporal discretization.

Many techniques are based on finite-difference approximations to the partial-differential equations that express the basic laws of nature. Closely related are finite-volume and finite-element approaches.

In addition to the constitutive relations, which vary from one material to another, the formulation of the driving equations is based on the universal laws of mass, momentum, and energy conservation, which are not dependent on the specific material properties.

Two very common approaches to formulating expressions that drive the dynamics are called Lagrangian and Eulerian. The Lagrangian formulation is based on the resolution of parcels of material whose dynamics are followed by the calculations in such a way as to ensure that the same elements of material are always present in the parcel. The Eulerian formulation divides space into small compartments (cells) that remain stationary while the fluid flows into and out of each region. Each approach has computational advantages and disadvantages. In flows that are highly contorting, the careful tracking of each element of material with Lagrangian techniques becomes quite difficult. While this tracking is not required in the Eulerian approach, the principal difficulty is that of following the motions of interfaces between materials, for example, the wavy motion of the surface between water and air in a lake.

A New Challenge

The core members of the fluid dynamics staff in the early 1960s included Tony Amsden, Dan Butler, Bart Daly, Martha Evans, Jake Fromm, Dick Gentry, Tony Hirt, Bob Martin, Bill Meixner, John Shannon, Eddie Welch, Emma Lou Young, and me.

Carson Mark gave the new group considerable freedom to pursue its dreams; and the results have, indeed, formed the foundation on which

much of the material-dynamics computing capability throughout the world has been based.

Particle-in-Cell Method

One of the earliest CFD accomplishments in T-Division was to develop a way to exploit the best properties of combined Lagrangian and Eulerian formulations so as to avoid the difficulties of each separately. Marker particles were constructed to follow the individual parcels of fluid moving through an Eulerian mesh of small rectangular cells.

Figure 1 shows a representative calculation that depicts the high-speed impact of a projectile onto a deformable target. The Eulerian mesh of cells is not shown in the computer-generated illustration of Lagrangian marker particles. Most of the calculation takes place on the Eulerian mesh, while the Lagrangian markers describe the trajectories of the individual elements of material.

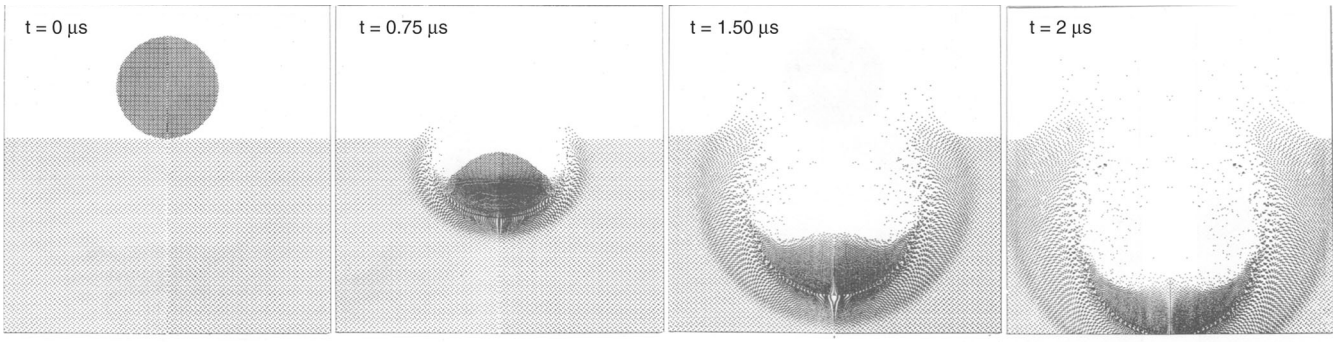
The PIC method continues to be used extensively for many investigations at Los Alamos and throughout the world. Many variations have been developed to extend the technique in a variety of different directions.

Fromm's Method for Incompressible Flows

The PIC method worked well for high-speed flows, but what about circumstances in which the fluid is incompressible?

The concepts of vorticity and stream functions had been known for many years. Building on these concepts, Jacob Fromm of T-Division developed an Eulerian finite-different technique for adapting the ideas to calculate the contorting flows of a single incompressible material. The vorticity for each mesh point was updated at each computational cycle, in terms of the values of vorticity and velocities at the previous time cycle. The first spectacular success was an application of the technique to the fluid-dynamics instability known as the von Karman vortex street. In addition to the production of beautiful pictures like those achieved in experiments, the shedding frequency for the vortices agreed very well with laboratory observations. (See Fig. 2.)

Fig. 1. A PIC-method calculation showing the Lagranian particles.



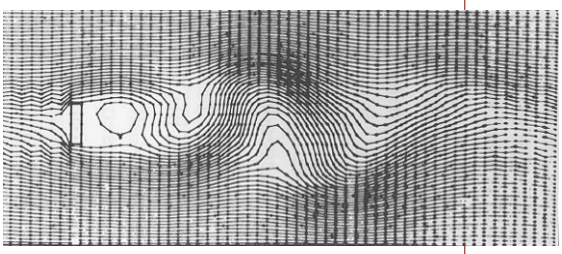


Fig. 2. Calculation of the von Karman vortex street using Fromm's vorticity and stream-function technique.

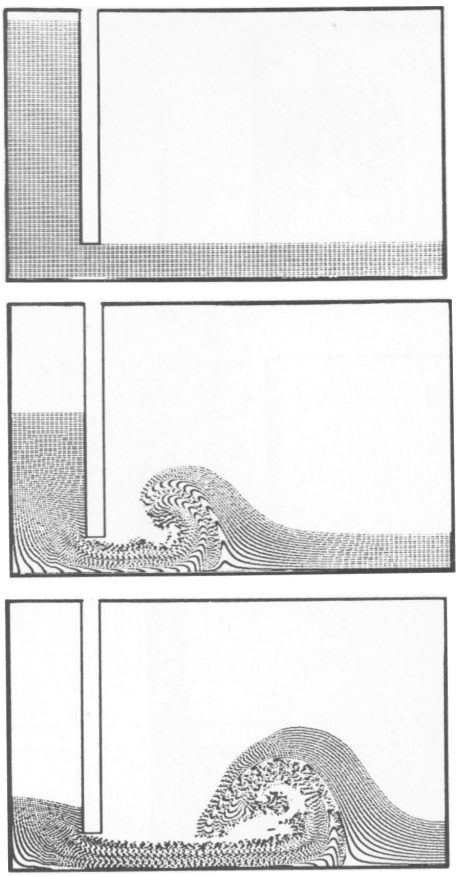


Fig. 3. MAC-method calculations of a water wave from a sluice gate.

Marker-and-Cell Method

One restriction on the use of a vorticity-and-stream-function approach is the difficulty in the incorporation of free-surface boundary conditions. To remove that problem, T-Division members developed the Marker-and-Cell (MAC) technique, in which the primitive variables, pressure and velocity, are used instead of the derived variables, vorticity and stream function. The concept of a staggered mesh of Eulerian cells was introduced to provide a convenient way to represent the vanishing divergence of the velocity field in the incompressible fluid. With pressure adopted as a primary variable, it became straightforward to prescribe free-surface boundary conditions for that variable.

The application of this technique to calculate the dynamics of water waves and splashing drops generated international interest, and the technique continues to be used in numerous computer laboratories for an astonishing variety of investigations. At last count, the original paper has been cited over 900 times. (See Fig. 3.)

Implicit Continuous-Field Eulerian Method

The next capability that was required for a variety of fluid-flow problems was a technique for calculating fluid flows that combine regions of low Mach number (e.g., incompressible) flow with regions in which the Mach number is high enough for the presence of compressibility to be important.

As with any technique attempting to represent regions of far-subsonic flow, the method requires an implicit formulation of the equations, in which the updated flow variables at each time cycle depend on the updated values of that and other variables in the same or adjacent computational cells. To solve these coupled equations requires the inversion of a complicated matrix, which can be accomplished by either an iterative procedure or by one of the powerful new solving techniques that have been developed in recent years.

The development of various modifications to the technique has enabled its extension to numerous applications that continue to be investigated to the present day.

Two-Field Flow

Gas bubbles in a liquid and raindrops in the atmosphere are two examples of two-field flow (often called two-phase flow, even though the two materials can both be of the same phase, e.g., oil droplets in water). The interpenetration of two materials occurs in numerous industrial circumstances, as, for example, in a petroleum processing plant where catalytic particles are mixed with crude oil in order to enhance the rate of chemical reactions.

Of particular significance to the Laboratory's activities is the mixing of materials across an unstable interface. Caused by bulk shearing motion or by the presence of pressure gradients, the growth of instability amplitude can form a mixing zone in which the dynamics occur as an imbalance between the dominant driving forces and the deceleration from drag between the interpenetrating entities. Figure 4 shows a detailed calculation of the complex mixing that the two-field equations were constructed to describe.

The equations for two-field flow were very controversial for years, the principal problem being their "ill-posedness." T-Division scientists, investigating the steam-water flow that might occur in a pressurized-water reactor, demonstrated that a very small amount of dissipation cures the difficulty and enables the development of a fine numerical technique for solving a wide range of problems. A further description of two-field flow is given in the "Fluid Dynamics" chapter.

Truncation Error Analysis

An important T-Division contribution to CFD is the systematic investigation of the differences between the continuum equation and their discretized approximations. A fundamental paper by C. W. Hirt shows the origins of the effects of various differencing schemes on the numerical stability and accuracy of the discretized approximations.

Arbitrary Eulerian-Lagrangian Techniques

In Lagrangian calculations, the discrete mesh moves with the fluid; in Eulerian approaches the fluid moves through a stationary mesh of cells. As an alternative to the PIC method, the T-Division Arbitrary Eulerian-Lagrangian (ALE) technique combines the best properties of each approach by a technique that requires only one mesh of computational cells. These cells can move with the fluid or be stationary, or be anything in between these limiting cases. The Lagrangian advantage of interface resolution is combined with the allowance for mass, momentum, and energy flux across the cell edges away from the interface, thus avoiding mesh tangling that would occur in a purely Lagrangian calculation. Techniques for front tracking and interface reconstruction have also been developed in T-Division as a means for ensuring the sharp definition of material interfaces in an otherwise purely Eulerian technique.

Artificial Viscosity

In the absence of terms representing the effects of viscosity, the equations of fluid dynamics suffer from a mathematical difficulty called "ill-posedness." Nature cures this difficulty by the incorporation of one or another dissipative process, associated with molecular interactions and/or fluctuations at the microscopic scales.

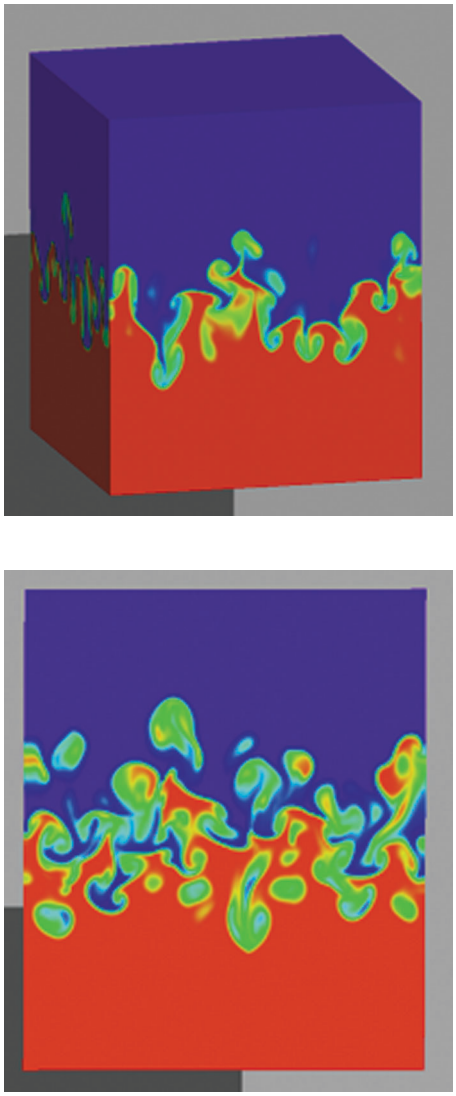


Fig. 4. Two-field interpretation of fluids driven by pressure gradients. The heavy field lies above the lighter one. Detailed calculations were performed in T-Division by Tim Clark, using a Lattice-Boltzmann technique.

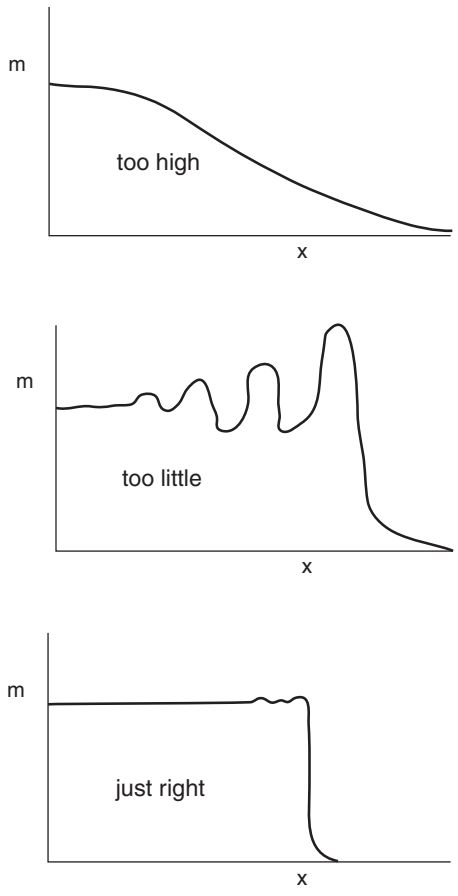


Fig. 5. Shock profiles of material velocity for various levels of artificial viscosity.

The most common manifestations of these small-scale effects in a gas are viscosity and heat conduction. In two-field flow the addition of a small amount of viscosity enables meaningful calculations for numerous applications. In single-field flows, however, the shocks that arise from nearby explosions (or other drivers) have an effective thickness of a few molecular mean-free paths. In ordinary air, for example, the shock width is typically about 10^{-5} cm, far too small for finite-difference numerical techniques to resolve in practical circumstances.

It was recognized, even in the early days of T-Division, that an artificial process had to be introduced, by which the shock thickness could be increased to a large enough scale for resolution. But this process had to enable the representation of the known proper jump conditions across the shock. The problem was solved by von Neumann and Richtmyer of T-Division, who, in a very famous paper, introduced a viscous stress proportional to the square of the velocity gradient in a direction normal to the shock. (See Fig. 5.)

This important T-Division contribution has been used in its original or modified form for virtually every high-speed-flow calculation performed anywhere in the world.

The Beginning of the Monte Carlo Method

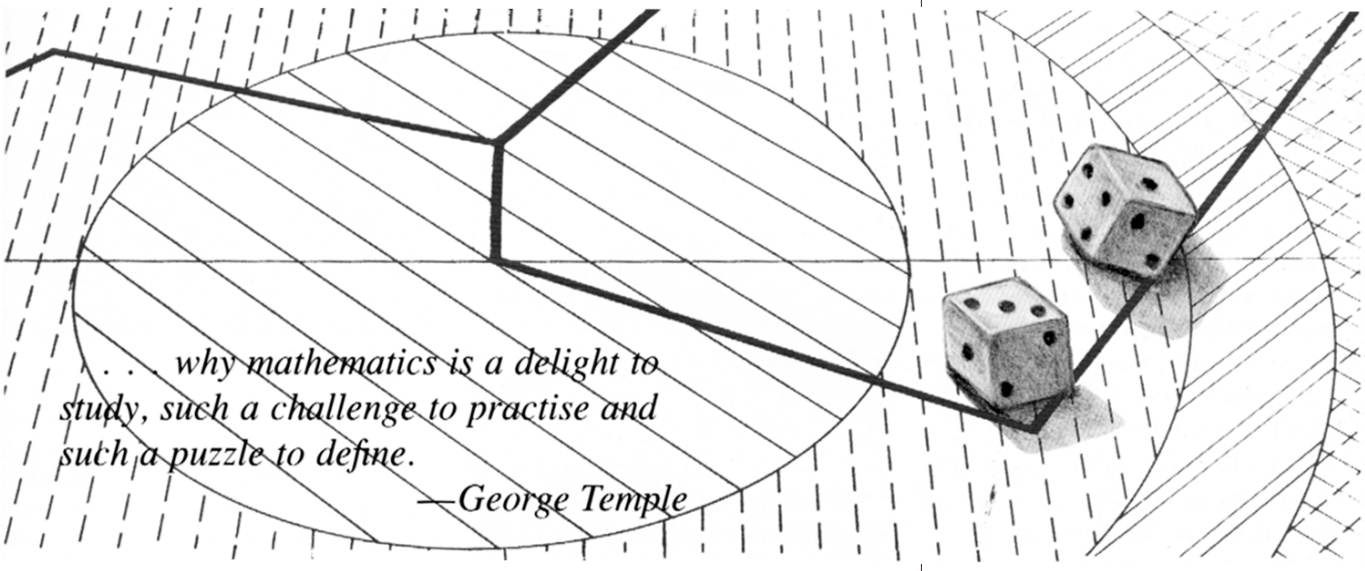
by Nicholas C. Metropolis

This article is reprinted with permission from *Los Alamos Science, Special Issue: Stanislaw Ulam 1909–1984* (Los Alamos National Laboratory, 1987), pp. 125–130.

The year was 1945. Two earthshaking events took place: the successful test at Alamogordo and the building of the first electronic computer. Their combined impact was to modify qualitatively the nature of global interactions between Russia and the West. No less perturbative were the changes wrought in all of academic research and in applied science. On a less grand scale these events brought about a renascence of a mathematical technique known to the old guard as statistical sampling; in its new surroundings and owing to its nature, there was no denying its new name of the Monte Carlo method.

This essay attempts to describe the details that led to this renascence and the roles played by the various actors. It is appropriate that it appears in an issue dedicated to Stan Ulam.

Bradbury
Calkin
Cashwell
Clippinger
Eckert
Edward
C. Evans
F. Evans
Everett
Fermi
Frankel
Gray
King
Lehmer
Mauchly
Merwin
Metropolis
Richtmyer
A. Rosenbluth
M. Rosenbluth
Segrè
E. Teller
A. Teller
Turkevich



Ulam
J. von Neumann
K. von Neumann

Some Background

Most of us have grown so blasé about computer developments and capabilities—even some that are spectacular—that it is difficult to believe or imagine there was a time when we suffered the noisy, painstakingly slow, electromechanical devices that chomped away on punched cards. Their saving grace was that they continued working around the clock, except for maintenance and occasional repair (such as removing a dust particle from a relay gap). But these machines helped enormously with the routine, relatively simple calculations that led to Hiroshima.

The ENIAC. During this wartime period, a team of scientists, engineers, and technicians was working furiously on the first electronic computer—the ENIAC—at the University of Pennsylvania in Philadelphia. Their mentors were Physicist First Class John Mauchly and Brilliant Engineer Presper Eckert. Mauchly, familiar with Geiger counters in physics laboratories, had realized that if electronic circuits could count, then they could do arithmetic and hence solve, *inter alia*, difference equations—at almost incredible speeds! When he'd seen a seemingly limitless array of women cranking out firing tables with desk calculators, he'd been inspired to propose to the Ballistics Research Laboratory at Aberdeen that an electronic computer be built to deal with these calculations.

John von Neumann, Professor of Mathematics at the Institute for Advanced Study, was a consultant to Aberdeen and to Los Alamos. For a whole host of reasons, he had become seriously interested in the thermonuclear problem being spawned at that time in Los Alamos by a friendly fellow-Hungarian scientist, Edward Teller, and his group. Johnny (as he was affectionately called) let it be known that construction of the ENIAC was nearing completion, and he wondered whether Stan Frankel and I would be interested in preparing a preliminary computational model of a thermonuclear reaction for the ENIAC. He felt he could convince the authorities at Aberdeen that our problem could provide a more exhaustive test of the computer than mere firing-table computations. (The designers of the ENIAC had wisely provided for the capability of much more ambitious versions of firing tables than were being arduously computed by hand, not to mention other quite different applications.) Our response to von Neumann's suggestion was enthusiastic, and his heuristic arguments were accepted by the authorities at Aberdeen.

In March, 1945, Johnny, Frankel, and I visited the Moore School of Electrical Engineering at the University of Pennsylvania for an advance glimpse of the ENIAC. We were impressed. Its physical size was overwhelming—some 18,000 double triode vacuum tubes in a system with 500,000 solder joints. No one ever had such a wonderful toy!

The staff was dedicated and enthusiastic; the friendly cooperation is still remembered. The prevailing spirit was akin to that in Los Alamos. What a pity that a war seems necessary to launch such revolutionary scientific endeavors. The components used in the ENIAC were joint-army-navy (JAN) rejects. This fact not only emphasizes the genius of Eckert and Mauchly and their staff, but also suggests that the ENIAC was technically realizable even before we entered the war in December, 1941.

After becoming saturated with indoctrination about the general and detailed structure of the ENIAC, Frankel and I retumed to Los Alamos to work on a model that was realistically calculable. (There was a small interlude at Alamogordo!) The war ended before we completed our set of problems, but it was agreed that we continue working. Anthony Turkevich joined the team and contributed substantially to all aspects of the work. Moreover, the uncertainty of the first phase of the postwar Los Alamos period prompted Edward Teller to urge us not only to complete the thermonuclear computations but to document and provide a critical review of the results.

The Spark. The review of the ENIAC results was held in the spring of 1946 at Los Alamos. In addition to Edward Teller, the principals included Enrico Fermi, John von Neumann, and the Director, Norris Bradbury. Stanley Frankel, Anthony Turkevich, and I described the ENIAC, the calculations, and the conclusions. Although the model was relatively simple, the simplifications were taken into account and the extrapolated results were cause for guarded optimism about the feasibility of a thermonuclear weapon.

Among the attendees was Stan Ulam, who had rejoined the Laboratory after a brief time on the mathematics faculty at the University of Southern California. Ulam's personality would stand out in any community, even where "characters" abounded. His was an informal nature; he would drop in casually, without the usual amenities. He preferred to chat, more or less at leisure, rather than to dissertate. Topics would range over mathematics, physics, world events, local news, games of chance, quotes from the classics—all treated somewhat episodically but always with a meaningful point. His was a mind ready to provide a critical link.

During his wartime stint at the Laboratory, Stan had become aware of the electromechanical computers used for implosion studies, so he was duly impressed, along with many other scientists, by the speed and versatility of the ENIAC. In addition, however, Stan's extensive mathematical background made him aware that statistical sampling techniques had fallen into desuetude because of the length and tediousness of the calculations. But with this miraculous development of the ENIAC—along with the applications Stan must have been pondering—it occurred to him that statistical techniques should be resuscitated, and he discussed this idea with von Neumann. Thus was triggered the spark that led to the Monte Carlo method.

Stanislaw Ulam

John von Neumann

The Method

The spirit of this method was consistent with Stan's interest in random processes—from the simple to the sublime. He relaxed playing solitaire; he was stimulated by playing poker; he would cite the times he drove into a filled parking lot at the same moment someone was accommodatingly leaving. More seriously, he created the concept of "lucky numbers," whose distribution was much like that of prime numbers; he was intrigued by the theory of branching processes and contributed much to its development, including its application during the war to neutron multiplication in fission devices. For a long time his collection of research interests included pattern development in two-dimensional games played according to very simple rules. Such work has lately emerged as a cottage industry known as cellular automata.

John von Neumann saw the relevance of Ulam's suggestion and, on March 11, 1947, sent a handwritten letter to Robert Richtmyer, the Theoretical Division leader. (see "Stan Ulam, John von Neumann, and the Monte Carlo Method"). His letter included a detailed outline of a possible statistical approach to solving the problem of neutron diffusion in fissionable material.

Johnny's interest in the method was contagious and inspiring. His seemingly relaxed attitude belied an intense interest and a well-disguised impatient drive. His talents were so obvious and his cooperative spirit so stimulating that he garnered the interest of many of us. It was at that time that I suggested an obvious name for the statistical method—a suggestion not unrelated to the fact that Stan had an uncle who would borrow money from relatives because he "just had to go to Monte Carlo." The name seems to have endured.

The spirit of Monte Carlo is best conveyed by the example discussed in von Neumann's letter to Richtmyer. Consider a spherical core of fissionable material surrounded by a shell of tamper material. Assume some initial distribution of neutrons in space and in velocity but ignore radiative and hydrodynamic effects. The idea is to now follow the development of a large number of individual neutron chains as a consequence of scattering, absorption, fission, and escape.

At each stage a sequence of decisions has to be made based on statistical probabilities appropriate to the physical and geometric factors. The first two decisions occur at time $t = 0$, when a neutron is selected to have a certain velocity and a certain spatial position. The next decisions are the position of the first collision and the nature of that collision. If it is determined that a fission occurs, the number of emerging neutrons must be decided upon, and each of these neutrons is eventually followed in the same fashion as the first. If the collision is decreed to be a scattering, appropriate statistics are invoked to determine the new momentum of the neutron. When the neutron crosses a material boundary, the parameters and characteristics of the new medium are taken

into account. Thus, a genealogical history of an individual neutron is developed. The process is repeated for other neutrons until a statistically valid picture is generated.

Random Numbers. How are the various decisions made? To start with, the computer must have a source of uniformly distributed *psuedo*-random numbers. A much used algorithm for generating such numbers is the so-called von Neumann "middle-square digits." Here, an arbitrary n-digit integer is squared, creating a $2n$-digit product. A new integer is formed by extracting the middle n-digits from the product. This process is iterated over and over, forming a chain of integers whose properties have been extensively studied. Clearly, this chain of numbers repeats after some point. H. Lehmer has suggested a scheme based on the Kronecker-Weyl theorem that generates all possible numbers of n digits before it repeats. (See "Random-Number Generators," for a discussion of various approaches to the generation of random numbers.)

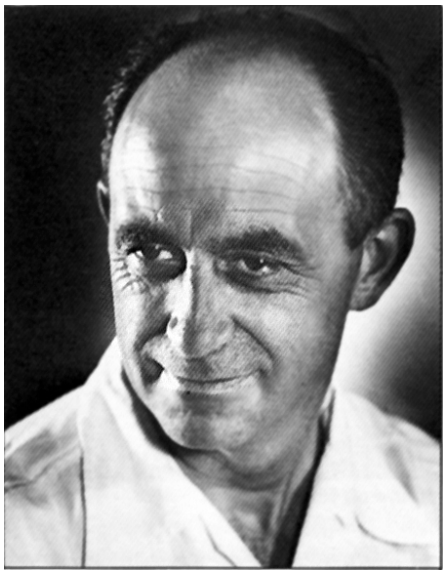
Once one has an algorithm for generating a uniformly distributed set of random numbers, these numbers must be transformed into the nonuniform distribution g desired for the property of interest. It can be shown that the function f needed to achieve this transformation is just the inverse of the nonuniform distribution function, that is, $f = g^{-1}$. For example, neutron physics shows us that the distribution of free paths—that is, how far neutrons of a given energy in a given material go before colliding with a nucleus—decreases exponentially in the interval $(0,\infty)$. If x is uniformly distributed in the open interval (0, 1), then $f = -1nx$ will give us a nonuniform distribution g with just those properties.

The reader will appreciate many of the advantages of the Monte Carlo method compared to the methods of differential equations. For example, a neutron-velocity spectrum with various peaks and valleys is difficult to handle mathematically. For Monte Carlo one needs only to mirror the velocity spectrum in the probability distribution. Also, the Monte Carlo method is sufficiently flexible to account for hydrodynamic effects in a self-consistent way. In an even more elaborate code, radiation effects can be dealt with by following the photons and their interactions (see "Monte Carlo at Work").

Clearly, applications of the Monte Carlo method are much broader than so far outlined. (Although I emphasize the use of Monte Carlo in the study of physical systems, random sampling is also an efficient way to evaluate complicated and many-dimensional integrals.) Since its inception, many international conferences have been held on the various applications of the method. …Recently, these range from the conference, "Monte Carlo Methods and Applications in Neutronics, Photonics, and Statistical Physics," at Cadarache Castle, France, in the spring of 1985 to the latest at Los Alamos, "Frontiers of Quantum Monte Carlo," in September 1985.

Then there was von Neumann, the great mathematician. We used to go for walks on Sunday. We'd walk in the canyons, and we'd often walk with Bethe, and Von Neumann, and Bacher. It was a great pleasure. And Von Neumann gave me an interesting idea; that you don't have to be responsible for the world that you're in. So I have developed a very powerful sense of social irresponsibility as a result of Von Neumann's advice. It's made me a very happy man ever since. But it was Von Neumann who put the seed in that grew into my active irresponsibility!

—Richard P. Feynman [Nobel Laureate and former T-Division scientist], "Los Alamos from Below," in Reminiscences of Los Alamos: 1943–1945, p. 129.



Enrico Fermi

Putting the Method into Practice

Let me return to the historical account. In late 1947, the ENIAC was to be moved to its permanent home at the Ballistics Research Laboratory in Maryland. What a gargantuan task! Few observers were of the opinion that it would ever do another multiplication or even an addition. It is a tribute to the patience and skill of Josh Gray and Richard Merwin, two fearless uninitiates, that the move was a success. One salutary effect of the interruption for Monte Carlo was that another distinguished physicist took this occasion to resume his interest in statistical studies.

Enrico Fermi helped create modern physics. Here, we focus on his interest in neutron diffusion during those exciting times in Rome in the early thirties. According to Emilio Segrè, Fermi's student and collaborator, "Fermi had invented, but of course not named, the present Monte Carlo method when he was studying the moderation of neutrons in Rome. He did not publish anything on the subject, but he used the method to solve many problems with whatever calculating facilities he had, chiefly a small mechanical adding machine."*

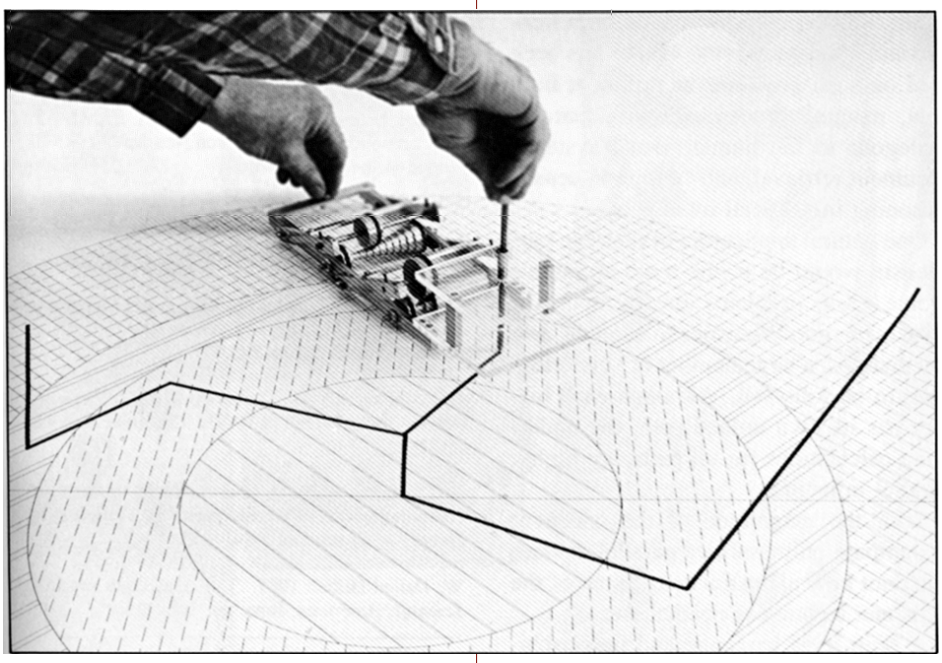
In a recent conversation with Segrè, I learned that Fermi took great delight in astonishing his Roman colleagues with his remarkably accurate, "too-good-to-believe" predictions of experimental results. After indulging himself, he revealed that his "guesses" were really derived from the statistical sampling techniques that he used to calculate with whenever insomnia struck in the wee morning hours! And so it was that nearly fifteen years earlier, Fermi had independently developed the Monte Carlo method.

It was then natural for Fermi, during the hiatus in the ENIAC operation, to dream up a simple but ingenious *analog* device to implement studies in neutron transport. He persuaded his friend and collaborator Percy King, while on a hike one Sunday morning in the mountains surrounding Los Alamos, to build such an instrument—later affectionately called the FERMIAC.

The FERMIAC developed neutron genealogies in two dimensions, that is, in a plane, by generating the site of the "next collision." Each generation was based on a choice of parameters that characterized the particular material being traversed. When a material boundary was crossed, another choice was made appropriate to the new material. The device could accommodate two neutron energies, referred to as "slow" and "fast." Once again, the master had just the right feel for what was meaningful and relevant to do in the pursuit of science.

*Quoted with permission of W. H. Freeman and Company from *From X-Rays to Quarks* by Emilio Segrè.

The First Ambitious Test. Much to the amazement of many "experts," the ENIAC survived the vicissitudes of its 200-mile journey. In the meantime, Richard Clippinger, a staff member at Aberdeen, had suggested that the ENIAC had sufficient flexibility to permit its controls to be reorganized into a more convenient (albeit static) stored-program mode of operation. This mode would have a capacity of 1800 instructions from a vocabulary of about 60 arithmetical and logical operations. The previous method of programming might be likened to a giant plugboard, that is to say, to a can of worms. Although implementing the new approach is an interesting story, suffice it to say that Johnny's wife, Klari, and I designed the new controls in about two months and completed the implementation in a fortnight. We then had the opportunity of using the ENIAC for the first ambitious test of the Monte Carlo method—a variety of problems in neutron transport done in collaboration with Johnny.



THE FERMIAC.
The Monte Carlo trolley, or FERMIAC, was invented by Enrico Fermi and constructed by Percy King. The drums on the trolley were set according to the material being traversed and a random choice between fast and slow neutrons. Another random digit was used to determine the direction of motion, and a third was selected to give the distance to the next collision. The trolley was then operated by moving it across a two-dimensional scale drawing of the nuclear device or reactor assembly being studied. The trolley drew a path as it rolled, stopping for changes in drum settings whenever a material boundary was crossed. This infant computer was used for about two years to determine, among other things, the change in neutron population with time in numerous types of nuclear systems.

Nine problems were computed corresponding to various configurations of materials, initial distributions of neutrons, and running times. These problems, as yet, did not include hydrodynamic or radiative effects; but complex geometries and realistic neutron-velocity spectra were handled easily. The neutron histories were subjected to a variety of statistical analyses and comparisons with other approaches. Conclusions about the efficacy of the method were quite favorable. It seemed as though Monte Carlo was here to stay.

Not long afterward, other Laboratory staff members made their pilgrimages to ENIAC to run Monte Carlo problems. These included J. Calkin, C. Evans, and F. Evans, who studied a thermonuclear problem using a cylindrical model as well as the simpler spherical one. B. Suydam and R. Stark tested the concept of artificial viscosity on time-dependent shocks. They also, for the first time, tested and found satisfactory an approach to hydrodynamics using a realistic equation of state in spherical geometry. Also, the distinguished (and mysterious) mathematician C. J. Everett was taking an interest in Monte Carlo that would culminate in a series of outstanding publications in collaboration with E. Cashwell. Meanwhile, Richtmyer was very actively running Monte Carlo problems on the so called SSEC during its brief existence at IBM in New York.

In many ways, as one looks back, it was among the best of times.

Rapid Growth. Applications discussed in the literature were many and varied and spread quickly. By midyear 1949 a symposium on the Monte Carlo method, sponsored by the Rand Corporation, the National Bureau of Standards, and the Oak Ridge Laboratory, was held in Los Angeles. Later, a second symposium was organized by members of the Statistical Laboratory at the University of' Florida in Gainesville.

In early 1952 a new computer, the MANIAC, became operational at Los Alamos. Soon after Anthony Turkevich led a study of the nuclear cascades that result when an accelerated particle collides with a nucleus. The incoming particle strikes a nucleon, experiencing either an elastic or an ineleastic scattering, with the latter event producing a pion. In this study, particles and their subsequent collisions were followed until all particles either escaped from the nucleus or their energy dropped below some threshold value. The "experiment" was repeated until sufficient statistics were accumulated. A whole series of target nuclei and incoming particle energies was examined.

Another computational problem run on the MANIAC was a study of equations of state based on the two-dimensional motion of hard spheres. The work was a collaborative effort with the Tellers, Edward and Mici, and the Rosenbluths, Marshall and Arianna (see "Monte Carlo at Work"). During this study a strategy was developed that led to greater computing efficiency for equilibrium systems that obey the Boltzmann distribution function. According to this strategy, if a statistical of a particle in the system resulted in a decrease in the energy of the system, the new configuration was accepted. On the other hand, if there was an increase in energy, the next configuration was accepted only if it survived a game of chance biased by a Boltzmann factor. Otherwise, the old configuration became a new statistic.

It is interesting to look back over two-score years and note the emergence, rather early on, of experimental mathematics, a natural consequence of the electronic computer. The role of the Monte Carlo method in reinforcing such mathematics seems self-evident. When display units were introduced, the temptation to experiment became almost irresistible, at least for the fortunate few who enjoyed the luxury of a hands-on policy. When shared-time operations became realistic, experimental mathematics came of age. At long last, mathematics achieved a certain parity—the twofold aspect of experiment and theory—that all other sciences enjoy.

It is, in fact, the coupling of the subtleties of the human brain with rapid and reliable calculations, both arithmetical and logical, by the modem computer that has stimulated the development of experimental mathematics. This development will enable us to achieve Olympian heights.

The Future

So far I have summarized the rebirth of statistical sampling under the rubric of Monte Carlo. What of the future—perhaps even a not too distant future?

The miracle of the chip, like most miracles, is almost unbelievable. Yet the fantastic performances achieved to date have not quieted all users. At the same time, we are reaching upper limits on the computing power of a single processor.

One bright facet of the miracle is the lack of macroscopic moving parts, which makes the chip a very reliable bit of hardware. Such reliability suggests parallel processing. The thought here is not a simple extension to two, or even four or eight, processing systems. Such extensions are adiabatic transitions that, to be sure, should be part of the immediate, short-term game plan. Rather, the thought is massively parallel operations with thousands of interacting processors—even millions!

Already commercially available is one computer, the Connection Machine, with 65,536 simple processors working in parallel. The processors are linked in such a way that no processor in the array is more than twelve wires away from another; and the processors are pairwise connected by a number of equally efficient routes, making communication both flexible and efficient. The computer has been used on such problems as turbulent fluid flow, imaging processing (with features analogous to the human visual system), document retrieval, and "common-sense" reasoning in artificial intelligence.

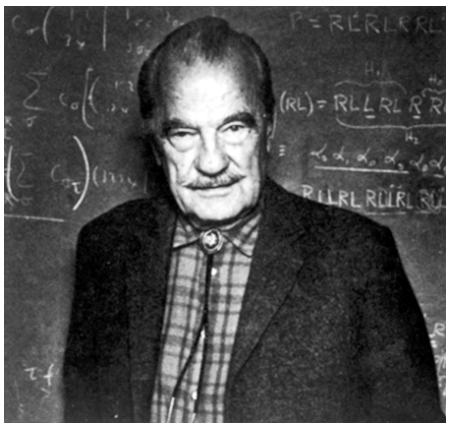
One natural application of massive parallelism would be to the more ambitious Monte Carlo problems already upon us. To achieve good statistics in Monte Carlo calculations, a large number of "histories" need to be followed. Although each history has its own unique path, the underlying calculations for all paths are highly parallel in nature.

Still, the magnitude of the endeavor to compute on massively parallel devices must not be underestimated. Some of the tools and techniques needed are

- A high-level language and new architecture able to deal with the demands of such a sophisticated language (to the relief of the user);
- Highly efficient operating systems and compilers;
- Use of modem combinatorial theory, perhaps even new principles of logic, in the development of elegant, comprehensive architectures;
- A fresh look at numerical analysis and the preparation of new algorithms. (we have been mesmerized by serial computation and purblind to the sophistication and artistry of parallelism).

Where will all this lead? If one were to wax enthusiastic, perhaps—just perhaps—a simplified model of the brain might be studied. These studies, in turn, might provide feedback to computer architects designing the new parallel structures.

Such matters fascinated Stan Ulam. He often mused about the nature of memory and how it was implemented in the brain. Most important, though, his own brain possessed the fertile imagination needed to make substantive contributions to the very important pursuit of understanding intelligence.



Nicholas C. Metropolis

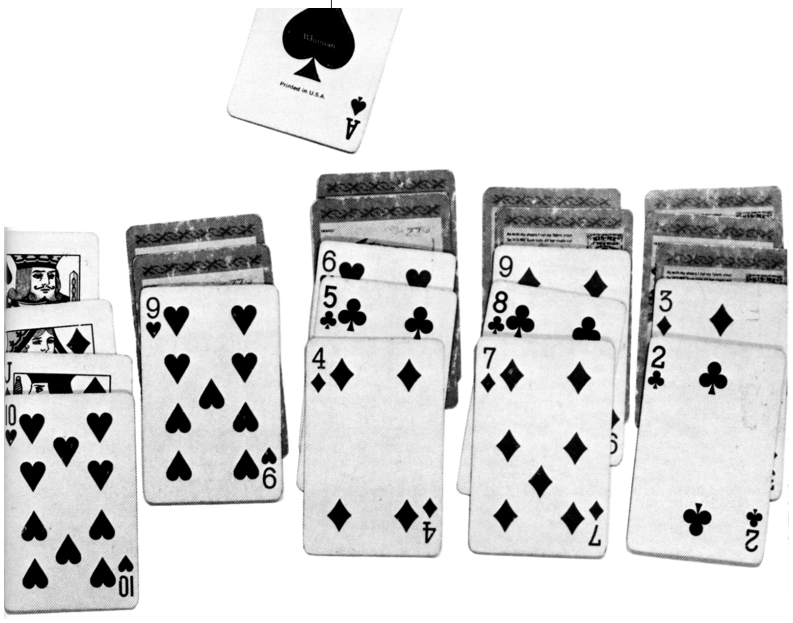
Nicholas Metropolis received his B.S. (1937) and his Ph.D. (1941) in physics at the University of Chicago. He arrived in Los Alamos, April 1943, as a member of the original staff of fifty scientists. After the war, he returned to the faculty of the University of Chicago as Assistant Professor. He came back to Los Alamos in 1948 to form the group that designed and built MANIAC I and II. (He chose the name MANIAC in the hope of stopping the rash of such acronyms for machine names, but may have, instead, only further stimulated such use.) From 1957 to 1965, he was professor of physics at the University of Chicago and was the founding director of its Institute for Computer Research. In 1965, he returned to Los Alamos, where he was made a Laboratory Senior Fellow in 1980.

Stan Ulam, John von Neumann, and the Monte Carlo Method

by Roger Eckhardt

This article is reprinted with permission from *Los Alamos Science, Special Issue: Stanislaw Ulam 1909–1984* (Los Alamos National Laboratory, 1987), pp. 131–137.

The Monte Carlo method is a statistical sampling technique that over the years has been applied successfully to a vast number of scientific problems. Although the computer codes that implement Monte Carlo have grown ever more sophisticated, the essence of the method is captured in some unpublished remarks Stan [Ulam] made in 1983 about solitaire.

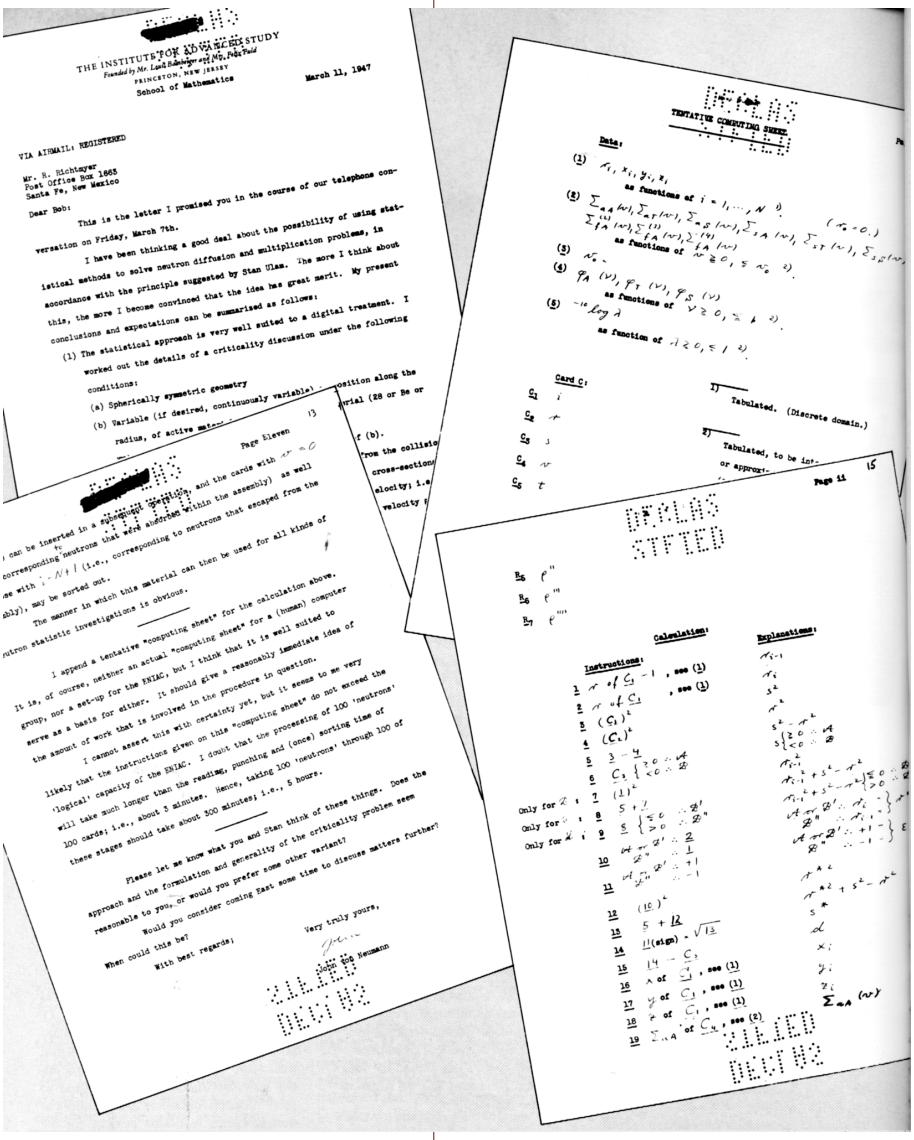


"The first thoughts and attempts I made to practice [the Monte Carlo method] were suggested by a question which occurred to me in 1946 as I was convalescing from an illness and playing solitaire. The question was what are the chances that a Canfield solitaire laid out with 52 cards will come out successfully? After spending a lot of time trying to estimate them by pure combinatorial calculations, I wondered whether a more practical method than "abstract thinking" might not be to lay it out say one hundred times and simply observe and count the number of successful plays. This was already possible to envisage with the beginning of the new era of fast computers, and I immediately thought of problems of neutron diffusion and other questions of mathematical physics, and more generally how to change processes described by certain differential equations into an equivalent form interpretable as a succession of random operations. Later...[in 1946, I] described the idea to John von Neumann and we began to plan actual calculations."

Von Neumann was intrigued. Statistical sampling was already well known in mathematics, but he was taken by the idea of doing such sampling using the newly developed electronic computing techniques. The approach seemed especially suitable for exploring the behavior of neutron chain reactions in fission devices. In particular, neutron multiplication rates could be estimated and used to predict the explosive behavior of the various fission weapons then being designed.

In March of 1947, he wrote to Robert Richtmyer, at that time the Theoretical Division leader at Los Alamos (Fig. 1). He had concluded that "the statistical approach is very well suited to a digital treatment," and he outlined in some detail how this method could be used to solve neutron diffusion and multiplication problems in fission devices for the case "of 'inert' criticality" (that is, approximated as momentarily static configurations). This outline was the first formulation of a Monte Carlo computation for an electronic computing machine.

In his formulation, von Neumann used a spherically symmetric geometry in which the various materials of interest varied only with the radius. He assumed that the neutrons were generated isotropically and had a known velocity spectrum and that the absorption, scattering, and fission cross-sections in the fissionable material and any surrounding materials (such as neutron moderators or reflectors) could be described as a function of neutron velocity. Finally, he assumed an appropriate accounting of the statistical character of the number of fission neutrons with probabilities specified for the generation of two, three, or four neutrons in each fission process.


THE INSTITUTE FOR ADVANCED STUDY
Founded by Mr. Louis Bamberger and Mrs. Felix Fuld
PRINCETON, NEW JERSEY
School of Mathematics

March 11, 1947

VIA AIRMAIL; REGISTERED

Mr. R. Richtmyer
Post Office Box 1663
Santa Fe, New Mexico

Dear Bob:

This is the letter I promised you in the course of our telephone conversation on Friday, March 7th.

I have been thinking a good deal about the possibility of using statistical methods to solve neutron diffusion and multiplication problems, in accordance with the principle suggested by Stan Ulam. The more I think about this, the more I become convinced that the idea has great merit. My present conclusions and expectations can be summarized as follows:

(1) The statistical approach is very well suited to a digital treatment. I worked out the details of a criticality discussion under the following conditions:

(a) Spherically symmetric geometry

(b) Variable (if desired, continuously variable) position along the radius, of active mate...

Page Eleven

...can be inserted in a subsequent operation, and the cards with ... corresponding to neutrons that were scattered within the assembly as well as with ... (i.e., corresponding to neutrons that escaped from the ...ably), may be sorted out.

The manner in which this material can then be used for all kinds of neutron statistic investigations is obvious.

I append a tentative "computing sheet" for the calculation above. It is, of course, neither an actual "computing sheet" for a (human) computer group, nor a set-up for the ENIAC, but I think that it is well suited to serve as a basis for either. It should give a reasonably immediate idea of the amount of work that is involved in the procedure in question.

I cannot assert this with certainty yet, but it seems to me very likely that the instructions given on this "computing sheet" do not exceed the 'logical' capacity of the ENIAC. I doubt that the processing of 100 'neutrons' will take such longer than the reading, punching and (once) sorting time of 100 cards; i.e., about 5 minutes. Hence, taking 100 'neutrons' through 100 of these stages should take about 500 minutes; i.e., 8 hours.

Please let me know what you and Stan think of these things. Does the approach and the formulation and generality of the criticality problem seem reasonable to you, or would you prefer some other variant?

Would you consider coming East some time to discuss matters further? When could this be?

With best regards;

Very truly yours,

John von Neumann

TENTATIVE COMPUTING SHEET

Date:

Card C:

Tabulated. (Discrete domain.)

Tabulated, to be int... or approx...

Page 11

Instructions:

Explanations:

Only for ... ; see (1)

Only for ... ; see (1)

Only for ... ; see (2)

Fig.1. The first and last pages of von Neumann's remarkable letter to Robert Richtmyer are shown here, as well as a portion of his tentative computing sheet. That sheet illustrates how extensively von Neumann had applied himself to the details of a neutron-diffusion calculation.

The idea then was to trace out the history of a given neutron, using random digits to select the outcomes of the various interactions along the way. For example, von Neumann suggested that in the computation "each neutron is represented by [an 80-entry punched computer] card. . . which carries its characteristics," that is, such things as the zone of material the neutron was in, its radial position, whether it was moving inward or outward, its velocity, and the time. The card also carried "the necessary random values" that were used to determine at the next step in the history, such things as path length and direction, type of collision, velocity after scattering—up to seven variables in all. A "new" neutron was started (by assigning values to a new card) whenever the neutron under consideration was scattered or whenever it passed into another shell; cards

were started for several neutrons if the original neutron initiated a fission. One of the main quantities of interest, of course, was the neutron multiplication rate—for each of the 100 neutrons started, how many would be present after, say, 10^{-8} second?

At the end of the letter, von Neumann attached a tentative "computing sheet" that he felt would serve as a basis for setting up this calculation on the ENIAC. He went on to say that "it seems to me very likely that the instructions given on this 'computing sheet' do not exceed the 'logical' capacity of the ENIAC." He estimated that if a problem of the type he had just outlined required "following 100 primary neutrons through 100 collisions [each]...of the primary neutron or its descendants," then the calculations would "take about 5 hours." He further stated, somewhat optimistically, that "in changing over from one problem of this category to another one, only a few numerical constants will have to be set anew on one of the 'function table' organs of the ENIAC."

His treatment did not allow "for the displacements, and hence changes of material distribution, caused by hydrodynamics," which, of course, would have to be taken into account for an explosive device. But he stated that "I think that I know how to set up this problem, too: One has to follow, say 100 neutrons through a short time interval Δt ; get their momentum and energy transfer and generation in the ambient matter; calculate from this the displacement of matter; recalculate the history of the 100 neutrons by assuming that matter is in the middle position between its original (unperturbed) state and the above displaced (perturbed) state; ...iterating in this manner until a "self-consistent" system of neutron history and displacement of matter is reached. This is the treatment of the first time interval Δt. When it is completed, it will serve as a basis for a similar treatment of the second time interval...etc., etc."

Von Neumann also discussed the treatment of the radiation that is generated during fission. "The photons, too, may have to be treated 'individually' and statistically, on the same footing as the neutrons. This is, of course. a non-trivial complication, but it can hardly consume much more time and instructions than the corresponding neutronic part. It seems to me, therefore, that this approach will gradually lead to a completely satisfactory theory of efficiency and ultimately permit prediction of the behavior of all possible arrangements, the simple ones as well as the sophisticated ones."

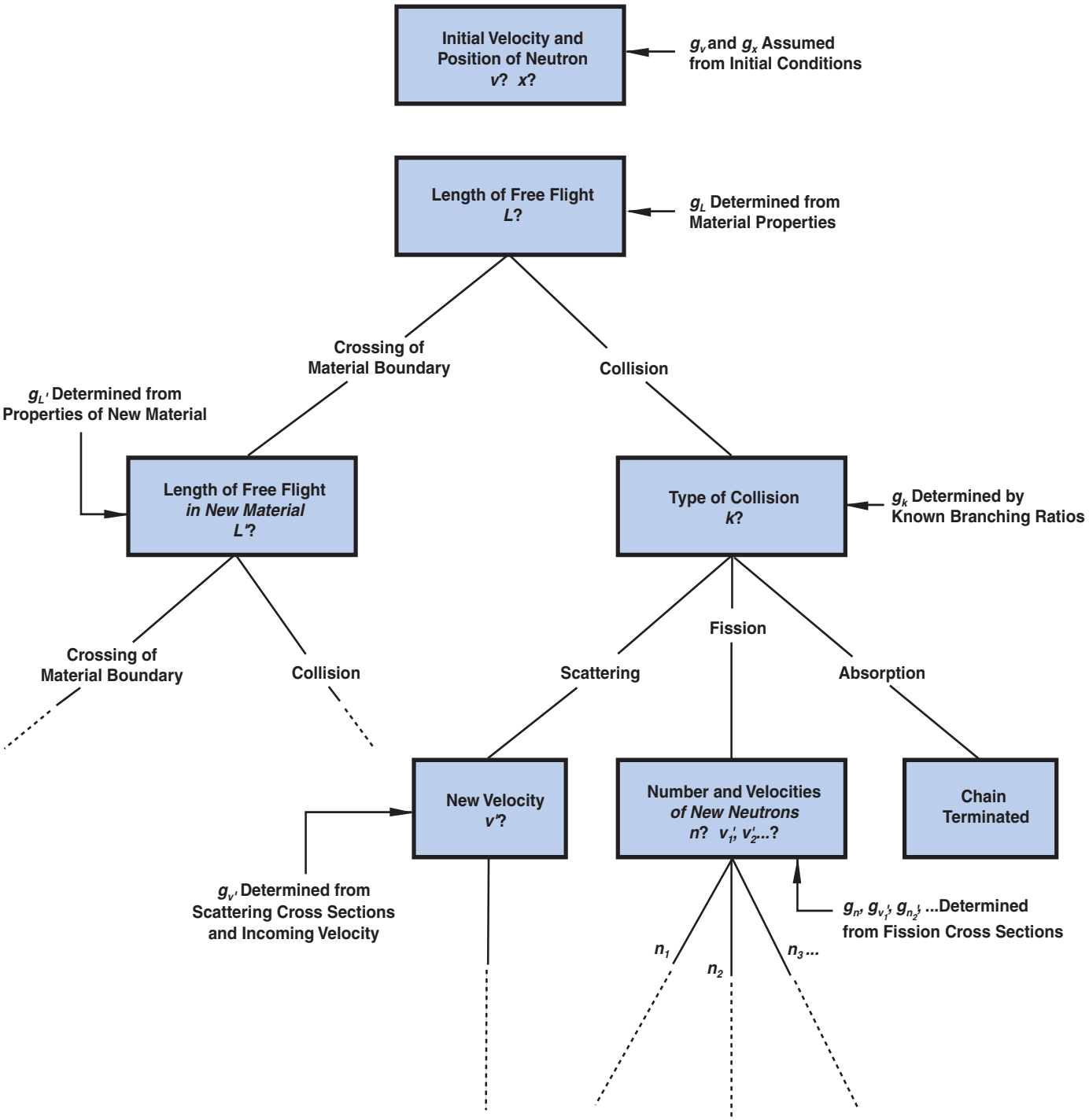
And so it has. At Los Alamos in 1947, the method was quickly brought to bear on problems pertaining to thermonuclear, as well as fission, devices; and, in 1948, Stan was able to report to the Atomic Energy Commission about the applicability of the method for such things as cosmic ray showers and the study of the Hamilton Jacobi partial differential equation. Essentially, all the ensuing work on Monte Carlo neutron-transport codes for weapons development and other applications has been directed at implementing the details of what von Neumann outlined so presciently in his 1947 letter. (see "Monte Carlo at Work").

In von Neumann's formulation of the neutron diffusion problem, each neutron history is analogous to a single game of solitaire, and the use of random numbers to make the choices along the way is analogous to the random turn of the card. Thus, to carry out a Monte Carlo calculation, one needs a source of random numbers, and many techniques have been developed that pick random numbers that are *uniformly* distributed on the unit interval (see "Random-Number Generators"). What is really needed, however, are *nonuniform* distributions that simulate probability distribution functions specific to each particular type of decision. In other words, how does one ensure that in random flights of a neutron, on the average, a fraction $e^{-x/\lambda}$ travels a distance x/λ mean free paths or farther without colliding?

The history of each neutron is generated by making various decisions about the physical events that occur as the neutron goes along (Fig. 2). Associated with each of these decision points is a known, and usually nonuniform, distribution of random numbers g that mirrors the probabilities for the outcomes possible for the event in question. For instance, returning to the example above, the distribution of random numbers g_L used to determine the distance that a neutron travels before interacting with a nucleus is exponentially decreasing, making the selection of shorter distances more probable than longer distances. Such a distribution simulates the observed exponential falloff of neutron path lengths. Similarly, the distribution of random numbers g_k used to select between a scattering, a fission, and an absorption must reflect the known probabilities for these different outcomes. The idea is to divide the unit interval (0, 1) into three subintervals in such a way that the probability of a uniform random number being in a given subinterval equals the probability of the outcome assigned to that set.

In another 1947 letter, this time to Stan Ulam, von Neumann discussed two techniques for using uniform distributions of random numbers to generate the desired nonuniform distributions g (Fig. 3). The first technique, which had already been proposed by Stan, uses the inverse of the desired function $f = g^{-1}$. For example, to get the exponentially decreasing distribution of random numbers on the interval $(0,\infty)$ needed for path lengths, one applies the inverse function $f(x) = -1nx$ to a uniform distribution of random numbers on the open interval (0, 1).

What if it is difficult or computationally expensive to form the inverse function, which is frequently true when the desired function is empirical? The rest of von Neumann's letter describes an alternative technique that will work for such cases. In this approach two uniform and independent distributions (x^i) and (y^i) are used. A value x^i from the first set is accepted when a value y^i from the second set satisfies the condition $y^i \le f(x^i)$, where $f(\xi^i)d\xi$ is the density of the desired distribution function (that is, $g(x) = \int f(x)dx$).



DECISION POINTS IN MONTE CARLO

Fig. 2. A schematic of some of the decisions that are made to generate the "history" of an individual neutron in a Monte Carlo calculation. The nonuniform random-number distributions *g* used in those decisions are determined from a variety of data.

THE INSTITUTE FOR ADVANCED STUDY
SCHOOL OF MATHEMATICS
PRINCETON, NEW JERSEY

May 21, 1947

Mr. Stan Ulam
Post Office Box 1663
Santa Fe
New Mexico

Dear Stan:

Thanks for your letter of the 19th. I need not tell you that Klari and I are looking forward to the trip and visit at Los Alamos this Summer. I have already received the necessary papers from Carson Mark. I filled out and returned mine yesterday; Klari's will follow today.

I am very glad that preparations for the random numbers work are to begin soon. In this connection, I would like to mention this: Assume that you have several random number distributions, each equidistributed in $0,1$: $(x^i), (y^i), (z^i), \ldots$. Assume that you want one with the distribution function (density) $f(\xi)d\xi$: (ξ^i). One way to form it is to form the cumulative distribution function: $g(\xi) = \int f(\xi)d\xi$, to invert it $h(x) = \xi \rightleftarrows x = g(\xi)$, and to form $\xi^i = h(x^i)$ with this $h(x)$, or some approximant polynomial. This is, as I see, the method that you have in mind.

An alternative, which works if ξ and all values of $f(\xi)$ lie in $0, 1$, is this: Scan pairs x^i, y^i and use or reject x^i, y^i according to whether $y^i \leqq f(x^i)$ or not. In the first case, put $\xi^j = x^i$ in the second case form no ξ^j at that step.

The second method may occasionally be better than the first one. In some cases combinations of both may be best; e.g., form random pairs $\xi = \sin x$, $\eta = \cos x$ with x equidistributed between 0° and 360°. The obvious way consists of using the sin - cos - tables (with interpolation). This is clearly closely related to the first method. This is an alternative procedure:

Put $\xi = \frac{2t}{1+t^2}$, $\eta = \frac{1-t^2}{1+t^2}$, $t = tg\, y$,

with y (which is $\frac{x}{2}$) equidistributed between 0° and 180°. Restrict y to 0° to 45°. Then the ξ, η will have to be replaced randomly by η, ξ and again by $\pm\xi, \pm\eta$. This can be done by using random digits $0, \ldots, 7$. It is also feasible with

dom digits 0, . . . , 9:
0 Replace ξ, η by ξ, η
1 " $-\xi, \eta$
2 " $\xi, -\eta$
3 " $-\xi, -\eta$
4 " η, ξ
5 " $\eta, -\xi$
6 " $-\eta, \xi$
7 " $-\eta, -\xi$
8 Reject this digit
9 " " "

Now $t = tg\, y$, $0^\circ \leqq y \leqq 45^\circ$, lies between 0 and 1, and its distribution function is $\frac{dt}{1+t^2}$. Hence one may pick pairs of numbers t, s both (independently) equidistributed between 0 and 1, and then

use t } for $(1+t^2)\, s \leqq 1$
reject t, s and form no t at this step } for $(1+t^2)\, s > 1$

Of course, the first part requires a divider, but the method may still be worth keeping in mind, especially when the ENIAC is available.

* * *

With best regards from house to house.

Yours, as ever,

John

John von Neumann

JvN:MW

ANOTHER VON NEUMANN LETTER

Fig. 3. In this 1947 letter to Stan Ulam, von Neumann discusses two methods for generating the nonuniform distributions of random numbers needed in the Monte Carlo method. The second paragraph summarizes the inverse-function approach in which (x^i) represents the uniform distribution and (ξ^i) the desired nonuniform distribution. The rest of the letter describes an alternative approach based on two uniform and independent distributions: (x^i) and (y^i). In this latter approach, a value x^i from the first set is accepted when a value y^i from the second set satisfies the condition $y^i \leq f(x^i)$, where $f(\xi^i)d\xi$ is the density of the desired distribution function. (It should be noted that in von Neumann's example for forming the random pairs $\xi = \sin x$ and $\eta = \cos x$, he probably meant to say that x is equidistributed between 0 and 360, (rather than "300"). Also, his notation for the tangent function is "tg," so that the second set of equations for ξ and η are just half-angle ($y = x/2$) trigonometric identities.)

This acceptance-rejection technique of von Neumann's can best be illustrated graphically (Fig. 4). If the two numbers x^i and y^i are selected randomly from the domain and range, respectively, of the function f, then each pair of numbers represents a point in the function's coordinate plane (x^i, y^i). When $y^i > f(x^i)$, the point lies above the curve for $f(x)$, and x^i is rejected; when $y^i \leq f(x^i)$ the point lies on or below the curve, and x^i is accepted. Thus, the fraction of accepted points is equal to the fraction of the area below the curve. In fact, the proportion of points selected that fall in a small interval along the x axis will be proportional to the average height of the curve in that interval, ensuring generation of random numbers that mirror the desired distribution.

After a series of "games" have been played, how does one extract meaningful information? For each of thousands of neutrons, the variables describing the chain of events are stored, and this collection constitutes a numerical model of the process being studied. The collection of variables is analyzed using statistical methods identical to those used to analyze experimental observations of physical processes. One can thus extract information about any variable that was accounted for in the process. For example, the average energy of the neutrons at a particular time is calculated by simply taking the average of all the values generated by the chains

at that time. This value has an uncertainty proportional to $\sqrt{V/(N-1)}$, where V is the variance of, in this case, the energy and N is the number of trials, or chains, followed.

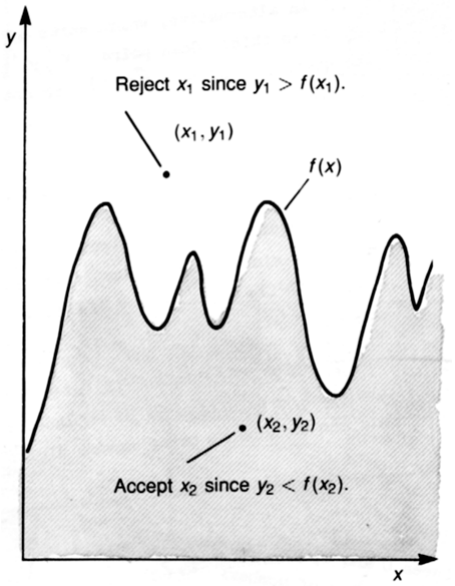
It is, of course, desirable to reduce statistical uncertainty. *Any* modification to the stochastic calculational process that generates the same expectation values but smaller variances is called a variance reduction technique. Such techniques frequently reflect the addition of known physics to the problem, and they reduce the variance by effectively increasing the number of data points pertinent to the variable of interest.

An example is dealing with neutron absorption by weighted sampling. In this technique, each neutron is assigned a unit "weight" at the start of its path. The weight is then decreased, bit by bit at each collision, in proportion to the absorption cross section divided by the total collision cross section. After each collision, an outcome *other* than absorption is selected by random sampling and the path is continued. This technique reduces the variance by replacing the sudden, one-time process of neutron absorption by a gradual elimination of the neutron.

Another example of variance reduction is a technique that deals with outcomes that terminate a chain. Say that at each collision *one* of the alternative outcomes terminates the chain and associated with this outcome is a particular value x_t for the variable of interest (an example is x_t being a path length long enough for the neutron to escape). Instead of collecting these values *only* when the chain terminates, one can generate considerably more data about this particular outcome by making an extra calculation at *each* decision point. In this calculation, the value x_t for termination is multiplied by the probability that that outcome will occur. Then *random* values are selected to continue the chain in the usual manner. By the end of the calculation, the "weighted values" for the terminating outcome have been summed over all decision points. This variance-reduction technique is especially useful if the probability of the alternative in question is low. For example, shielding calculations typically predict that only one in many thousands of neutrons actually get through the shielding. Instead of accumulating those rare paths, the small probabilities that a neutron will get through the shield on its very next free flight are accumulated after each collision.

The Monte Carlo method has proven to be a powerful and useful tool. In fact, "solitaire games" now range from the neutron- and photon-transport codes through the evaluation of multidimensional integrals, the exploration of the properties of high-temperature plasmas, and into the quantum mechanics of systems too complex for other methods.

Quite a handful.



THE ACCEPTANCE-REJECTION METHOD

Fig. 4. If two independent sets of random numbers are used, one of which (x^i) extends uniformly over the range of the distribution function *f* and the other (y^i) extends over the domain of *f*, then an acceptance-rejection technique based on whether or not $y^i \le f(x^i)$ will generate a distribution for (x^i) whose density is $f(x^i)\,dx^i$.

Random-Number Generators

Warnock

by Tony Warnock

This article is reprinted with permission from *Los Alamos Science, Special Issue: Stanislaw Ulam 1909–1984* (Los Alamos National Laboratory, 1987), pp. 137–141.

Random numbers have applications in many areas: simulation, game-playing, cryptography, statistical sampling, evaluation of multiple integrals, particle transport calculations, and computations in statistical physics, to name a few. Since each application involves slightly different criteria for judging the "worthiness" of the random numbers generated, a variety of generators have been developed each with its own set of advantages and disadvantages.

Depending on the application, three types of number sequences might prove adequate as the "random numbers." From a purist point of view, of course, a series of numbers generated by a truly random process is most desirable. This type of sequence is called a *random-number sequence*, and one of the key problems is deciding whether or not the generating process is, in fact, random. A more practical sequence is the *pseudo-random sequence*, a series of numbers generated by a deterministic process that is intended merely to imitate a random sequence but which, of course, does not rigorously obey such things as the laws of large numbers. Finally, a *quasi-random sequence* is a series of numbers that makes no pretense at being random but that has important predefined statistical properties shared with random sequences.

Physical Random-Number Generators

Games of chance are the classic examples of random processes, and the first inclination would be to use traditional gambling devices as random-number generators. Unfortunately, these devices are rather slow, especially since the typical computer application may require millions of numbers per second. Also, the numbers obtained from such devices are not always truly random: cards may be imperfectly shuffled, dice may not be true, wheels may not be balanced, and so forth. However, in the early 1950s the Rand Corporation constructed a million-digit table of random numbers using an electrical "roulette wheel." (The device had 32 slots, of which 12 were ignored; the others were numbered from 0 to 9 twice.)

Classical gambling devices appear random only because of our ignorance of initial conditions; in principle, these devices follow deterministic Newtonian physics. Another possibility for generating truly random numbers is to take advantage of the Heisenberg uncertainty principle and quantum effects, say by counting decays of a radioactive source or by tapping into electrical noise. Both of these methods have been used to generate random numbers for computers, but both suffer the defects of slowness and ill-defined distributions (however, on a different but better order of magnitude than gambling devices).

For instance, although each decay in a radioactive source may occur randomly and independently of other decays, it is not necessarily true that successive counts in the detector are independent of each other. The time it takes to reset the counter, for example, might depend on the previous count. Furthermore, the source itself constantly changes in time as the number of remaining radioactive particles decreases exponentially. Also, voltage drifts can introduce bias into the noise of electrical devices.

There are, of course, various tricks to overcome some of these disadvantages. One can partially compensate for the counter-reset problem by replacing the string of bits that represents a given count with a new number in which all of the original 1-1 and 0-0 pairs have been discarded and all of the original 0-1 and 1-0 pairs have been changed to 0 and 1, respectively. This trick reduces the bias caused when the probability of a 0 is different from that of a 1 but does not completely eliminate nonindependence of successive counts.

A shortcoming of *any* physical generator is the lack of reproducibility. Reproducibility is needed for debugging codes that use the random numbers and for making correlated or anticorrelated computations. Of course, if one wants random numbers for a cryptographic one-time pad, reproducibility is the last attribute desired, and time can be traded for security. A radioactive source used with the described bias-removal technique is probably sufficient.

Arithmetical Pseudo-Random Generators

The most common method of generating pseudo-random numbers on the computer uses a recursive technique called the linear-congruential, or Lehmer, generator. The sequence is defined on the set of integers by the recursion formula

$$x_{n+1} = Ax_n + C \pmod{M},$$

where x_n is the nth member of the sequence; and A, C, and M are parameters that can be adjusted for convenience and to ensure the

pseudorandom nature of the sequence. For example, M, the modulus, is frequently taken to be the word size on the computer; and A, the multiplier, is chosen to yield both a long period for the sequence and good statistical properties.

When M is a power of 2, it has been shown that a suitable sequence can be generated if, among other things, C is odd and A satisfies $A = 5 \pmod 8$ (that is, $A - 5$ is a multiple of 8). A simple example of the generation of a 5-bit number sequence using these conditions would be to set $M = 32$ (5 bits), $A = 21$, $C = 1$, and $x_0 = 13$. This yields the sequence

13, 18, 27, 24, 25, 14, 7, 20, 5, 10...

or, in binary,

01101, 10010, 11011, 11000, 11001, 01110, 00111, 10100, 00101, 01010,.... (1)

This type of generator has the interesting (or useful, or disastrous) property, illustrated by Seq. 1, that the least significant bit always has the alternating pattern 101010.… Further, the next bit has a pattern with period 4 (0110 above), the third bit has period 8, and so forth. Ultimately, the most significant bit has period M, which becomes the period of the sequence itself. Our example uses a short 5-bit word, which generates a sequence with a period of only 32. It is not unusual in many computer applications, however, to use many more bits (for example, to use a 32-bit word to generate a sequence with period $M = 2^{32}$).

One must be careful not to use such sequences in a problem with structures having powers of 2 in their dimensions. For example, a sequence with period 2^{32} would be a poor choice if the problem involved, say, a 3-dimensional lattice with sides of 128 ($= 2^7$) because the structure of the sequence can then interact unfavorably with the structure of the problem. Furthermore, there would be only $2^{32}/(2^7)^3 = 2048$ possible states. The usual assumption in Monte Carlo computations is that one has used a "representative" sample of the total number of possible computations—a condition that is certainly not true for this example.

One method of improving a pseudo-random-number generator is to combine two or more unrelated generators. The length of the hybrid will be the least common multiple of the lengths of the constituent sequences. For example, we can use the theory of *normal numbers* to construct a sequence that has all the statistical features of a "truly random" sequence and then combine it with a linear-congruential sequence. This technique yields a hybrid that possesses the strengths of both sequences—for example, one that retains the statistical features of the normal-number sequence.

We first construct a normal number, that is, a number in base b for which each block of K digits has limiting frequency $(1/b)^K$. A simple example in base 2 can be constructed by concatenating the sequence of integers

1, 10, 11, 100, 101, 110, 111, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111,...

to form the normal number

11011100101110111100010011010101111001101111011111....

If the number is blocked into 5-digit sets,

11011, 10010, 11101, 11100, 01001, 10101, 01111, 00110, 11110, 1111... (2)

it becomes a sequence of numbers in base 2 that satisfy all linear statistical conditions for randomness. For example, the frequency of a specific 5-bit number is $(1/2)^5$.

Sequences of this type do not "appear" random when examined; it is easy for a person to guess the rule of formation. However, we can further disguise the sequence by combining it with the linear-congruence sequence generated earlier (Seq. 1). We do this by performing an *exclusive-or* (XOR) operation on the two sequences:

01101, 10010, 11011, 11000, 11001, 01110, 00111, 10100, 00101, 01010,... (1)

and

11011, 10010, 11101, 11100, 01001, 10101, 01111, 00110, 11110, 11111,... (2)

yield

10110, 00000, 00110, 00100, 10000, 11011, 01000, 10010, 11011, 10101,....(3)

Of course, if Seq. 3 is carried out to many places, a pattern in it will also become apparent. To eliminate the new pattern, the sequence can be XOR'ed with a third pseudo-random sequence of another type, and so on.

This type of hybrid sequence is easy to generate on a binary computer. Although for most computations one does not have to go to such pains, the technique is especially attractive for constructing "canonical" generators of apparently random numbers.

A key idea here is to take the notion of randomness to mean simply that the sequence can pass a given set of statistical tests. In a sequence based on normal numbers, each term will depend nonlinearly on the previous terms. As a result, there are nonlinear statistical tests that can show the sequence not to be random. In particular, a test based on the transformations used to construct the sequence itself will fail. But, the sequence will pass all *linear* statistical tests; and, on that level, it can be considered to be random.

What types of linear statistical tests are applied to pseudo-random numbers? Traditionally, sequences are tested for uniformity of distribution of single elements, pairs, triples, and so forth. Other tests may be performed, depending on the type of problem for which the sequence will be used. For example, just as the correlation between two sequences can be tested, the autocorrelation of a single sequence can be tested after displacing the original sequence by various amounts. Or the number of different types of "runs" can be checked against the known statistics for runs. An increasing run, for example, consists of a sequential string of increasing numbers from the generator (such as, 0.08, 0.21, 0.55, 0.58, 0.73,...). The waiting times, for various events (such as the generation of a number in each of the five intervals (0.0.2), (0.2, 0.4),... .,(0.8, 1)) may be tallied and, again, checked against the known statistics for random-number sequences.

If a generator of pseudo-random numbers passes these tests, it is deemed to be a "good" generator, otherwise it is "bad." Calling these criteria "tests of randomness" is misleading because one is testing a hypothesis known to be false. The usefulness of the tests lies in their similarity to the problems that need to be solved using the stream of pseudorandom numbers. If the generator fails one of the simple tests, it will surely not perform reliably for the real problem. (Passing all such tests may not, however, be enough to make a generator work for a given problem, but it makes the programmers setting up the generator feel better.

Quasi-Random Numbers

For some applications, such as evaluating integrals numerically, the use of quasi-random sequences is much more efficient than the use of either random or pseudorandom sequences. Although quasi-random sequences do not necessarily mimic a random sequence, they can be tailored to satisfy the equi-distribution criteria needed for the integration. By this I mean, roughly speaking, that the numbers are spread throughout the region of interest in a much more uniform manner than a random or pseudo-random sequence.

For example, say one needs to find the average of the quantity $f(x)$ over the set of coordinates x, knowing the distribution of coordinate values $\rho(x)$ for the system being considered. Ordinarily, the average is given by the expression

$$(f) = \frac{\int \rho(x) f(x) dx}{\int \rho(x) dx}.$$

Rather than evaluating this integral, however, one can evaluate $f(x)$ at a series of random points. If the probability of picking a particular point x is proportional to the statistical weight $\rho(x)$, then (f) is given by the expression

$$\langle f \rangle = \sum_{i=1}^{N} f(x_i)/N,$$

where N is the total number of points chosen. This idea is the basis of the Metropolis technique of evaluating integrals by the Monte Carlo method.

Now if the points are taken from a random or a pseudorandom sequence, the statistical uncertainty will be proportional to $1/\sqrt{N}$. However, if a quasi-random sequence is used, the points will occupy the coordinate space with the correct distribution but in a more uniform manner, and the statistical uncertainty will be proportional to $1/N$. In other words, the uncertainty will decrease much faster with a quasi-random sequence than with a random or pseudo-random sequence.

How are quasi-random sequences generated? One type of sequence with a very uniform distribution is based on the radical-inverse function. The radical-inverse function $\phi(N, b)$ of a number N with base b is constructed by

1. writing the number in base b (for example, 14 in base 3 is 112);

2. reversing the digits (112 becomes 211), and

3. writing the result as a fraction less than 1 in base b (211 becomes 211/1000 in base 3 and, thus, ϕ (14. 3) = .211).

A sequence based on the radical-inverse function is generated by choosing a prime number as the base b and finding ϕ (l, b), ϕ (2, b), ϕ (3, b), ϕ (4, b),.... For a problem with k dimensions, the first k primes are used, and (ϕ (N, b_l), ϕ (N, b_2), ... ϕ (N, b_k)) becomes the Nth point of the k-dimensional sequence. This sequence has a very uniform distribution and is useful in multiple integration or multi-dimensional sampling.

There are many other types of random, pseudo-random, or quasi-random sequences than the ones I have discussed here, and there is much research aimed at generating sequences with the properties appropriate to the desired application. However, the examples I have discussed should illustrate both the approaches being taken and the obstacles that must be overcome in the quest of suitable “random” numbers.

Monte Carlo at Work

Doolen
Hendricks
Metropolis
A. Rosenbluth
M. Rosenbluth
A. Teller
E. Teller
Ulam

by Gary D. Doolen and John S. Hendricks

This article is reprinted with permission from *Los Alamos Science, Special Issue: Stanislaw Ulam 1909–1984* (Los Alamos National Laboratory, 1987), pp. 142–143.

Every second nearly 10,000,000,000 "random" numbers are generated on computers around the world for Monte Carlo solutions to problems that Stan Ulam first dreamed of solving forty years ago. A major industry now exists that has spawned hundreds of full-time careers invested in the fine art of generating Monte Carlo solutions—a livelihood that often consists of extracting an answer out of a noisy background. Here we focus on two of the extensively used Monte Carlo solvers: MCNP, an internationally used neutron and photon transport code maintained at Los Alamos; and the "Metropolis" method, a popular and efficient procedure for computing equilibrium properties of solids, liquids, gases, and plasmas.

MCNP

In the 1950s, shortly after the work on the Monte Carlo method by Ulam, von Neumann, Fermi, Metropolis, Richtmyer, and others, a series of Monte Carlo transport codes began emerging from Los Alamos. The concepts on which these codes were based were those outlined by von Neumann (see "Stan Ulam, John von Neumann, and the Monte Carlo Method"), but a great deal of detailed work was needed to incorporate the appropriate physics and to develop shorter routes to statistically valid solutions.

From the beginning, the neutron transport codes used a general treatment of the geometry; but successive versions added such features as cross-section libraries, variance reduction techniques (essentially clever ways to bias the random numbers so that the guesses will cluster around the correct solution), and a free-gas model treating thermalization of the energetic fission neutrons. Also, various photon transport codes were developed that dealt with photon energies from as low as 1 kilo-electron-volt to the high energies of gamma rays. Then, in 1973, the neutron transport and the photon transport codes were merged into one. In 1977, the first version of MCNP appeared in which photon cross sections were added to account for production of gamma rays by neutron interactions. Since then, the code has been distributed to more than 200 institutions worldwide.*

*The MCNP code and manual can be obtained from the Radiation Shielding Information Center (RSIC), P.O. Box X, Oak Ridge, TN 37831.

The Monte Carlo techniques and data now in the MCNP code represent more than 300 person-years of effort and have been used to calculate many tens of thousands of practical problems by scientists throughout the world. The types of problems include the design of nuclear reactors and nuclear safeguard systems, criticality analyses, oil well logging, health-physics problems, determinations of radiological doses, spacecraft radiation modeling, and radiation damage studies. Research on magnetic fusion has used MCNP heavily.

The MCNP code features a general three-dimensional geometry, continuous energy or multigroup physics packages, and sophisticated variance reduction techniques. Even very complex geometry and particle transport can be modeled almost exactly. In fact, the complexity of the geometry that can be represented is limited only by the dedication of the user.

The Metropolis Method

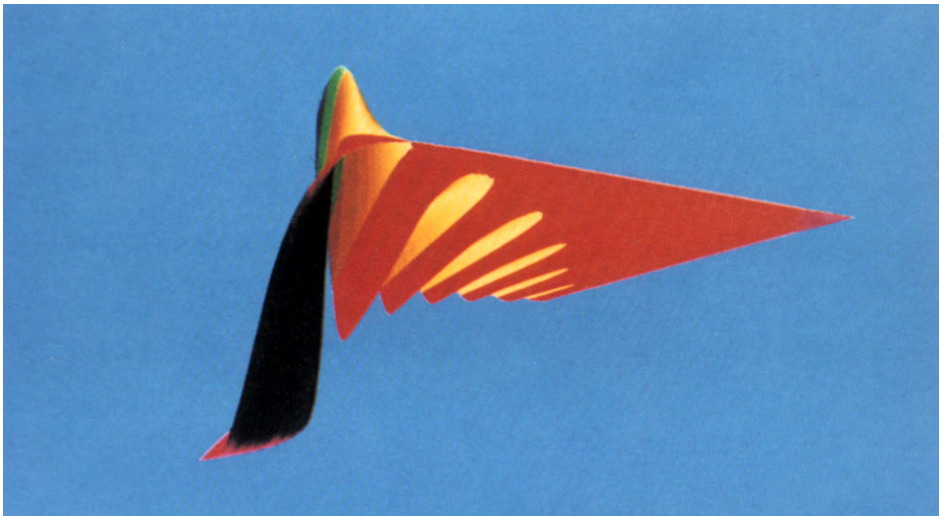
The problem of finding the energy and configuration of the lowest energy state of a system of many particles is conceptually simple. One calculates the energy of the system, randomly moves each particle a small distance, and recalculates the energy. If the energy has decreased, the new position is accepted; and the procedure continues until the energy no longer changes.

The question of how to calculate equilibrium properties of a finite system at a given temperature is more difficult, but it was answered in a 1953 *Journal of Chemical Physics* article by Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller, who decided that the calculation should follow the same steps for finding the minimum energy but with one important change. When a move results in an *increased* energy, one accepts the new position with probability $e^{-\Delta E/T}$, where ΔE is the change in energy and T is the temperature. This procedure gives the equilibrium solution for any physical system. In fact, a system with many particles can be solved with only a few lines of code and a fast computer.

Although calculations for short-range forces are much easier than for long-range forces (such as the Coulomb force), the Metropolis technique has been used for most systems in which the forces between particles are known. Wayne Slattery, Hugh DeWitt, and one of the authors (GD) applied the technique to a neutral Coulomb plasma consisting of thousands of particles in a periodic box. The purpose was to calculate such physical properties as the temperature at which this type of plasma

freezes and the pair distribution function which is the probability of finding one particle at a given distance from another. Because the uncertainty in a Monte Carlo result is proportional to $1/\sqrt{N}$, where N is the number of moves of a single particle, several million moves requiring several hundred Cray hours were needed to obtain accurate results for the plasma at many temperatures.

As computers are becoming faster and their memories increase, larger and more complicated systems are being calculated far more accurately than even Stan Ulam probably expected.



PAIR-DISTRIBUTION FUNCTION
This plot gives the probability of pairs of charged particles in a plasma being separated by a certain distance. The probabilities are plotted as a function of the distance between the pair of particles (increasing from left to right) and temperature (decreasing from front to back). At the left edge, both the distance and the probability are zero; at the right edge, the probability has become constant in value. Red indicates probabilities below this constant value; yellow and green indicate probabilities above it. As the temperature of the plasma decreases, lattice-like peaks begin to form in the pair-distribution function. The probabilities, generated with the Metropolis method described in the text, have been used for precise tests of many theoretical approximations for plasma models.

The CFDLib Project

by Bryan A. Kashiwa

Amsden
Baumgardner
Brackbill
Dukowicz
Edgerton
Harlow
N. Johnson
Kashiwa
Kothe
Padial-Collins
Rauenzahn
Zemach

Theoretical Division has a long history of using computer simulations to probe the nature of physical processes such as shock waves and free-surface motions for mission-dependent needs as well as for a fundamental understanding of fluid dynamic instabilities and their consequences—fluid turbulence. Over the years, the group created numerical methods for solving model equations, called numerical recipes. These efforts have produced a wide range of recipes whose acronyms are now part of a "language" of computational fluid dynamics in common use by practitioners at Los Alamos National Laboratory and in academia and industry worldwide.

In the late 1980s T-Division devised a repository called the Computational Fluid Dynamics Library (CFDLib) to keep the most valuable parts of each historical recipe up-to-date and available for use on new and ongoing studies. This article includes an example of a fluid dynamics study—viewing the dynamics of a single drop of milk falling onto a thin film of milk on a kitchen counter—that is aided by computer simulation. The physical process of interest in the simulation is the conversion of the initial kinetic energy in the falling drop into energy associated with the surface tension of the small droplets caused by the splash, and energy lost to viscous dissipation.

Theoretical Division has a long history of using computer simulations as a means of probing the nature of physical processes such as shock waves and free-surface motions. These studies have been driven by both problem-dependent (mission) needs, as well as the search for a fundamental understanding in areas such as fluid dynamic instabilities and their consequences—fluid turbulence.

In order to conduct a computer simulation, one needs three basic elements: (1) a set of model equations, (2) a numerical method for solving the equations, and (3) a computer program to carry out the numerical method. The problem of devising the model equations is one of physics; it is the process of setting forward a theory that is to be tested by the computer simulation and its comparison to a laboratory experiment. The problem of creating a numerical method for solving the model equations is an issue in applied mathematics. The numerical method is often called a "numerical recipe." Finally, the problem of creating a computer program to carry out the recipe faithfully and efficiently is the arena of computer science. The program is often called a code and is a set of instructions that is sent to the computer; the response from the computer is said to be a solution to the problem.

Over the years T-Division has found it necessary to focus squarely on the creation of innovative numerical recipes for very specific studies using computer simulations. These efforts have produced a wide range of recipes whose acronyms are now part of a "language" of Computational Fluid Dynamics (CFD) in common use by practitioners. In no particular order, some of the more important recipes have the names MAC (Marker-And-Cell); ICE (Implicit-Continuous-fluid-Eulerian); ALE (Arbitrary-Lagrangian-Eulerian); PIC (Particle-In-the-Cell); SOLA (SOLution-Algorithm); VOF (Volume-Of-Fluid); and FLIP (FLuid-Implicit-Particle).

For each recipe a new computer code was written to carry out the recipe and furnish solutions. An equally wide range of colorful names were given to the various codes, such as YAQUI, KACHINA, K-FIX, COCHINA, KIVA, and many, many others.

In the late 1980s, certain developments in the division's numerical recipes made it apparent that many (if not all) of the historical recipes could be cast in such a way that a single "library" of computer codes could be devised to serve as a repository for the MAC, ICE, ALE, PIC, VOF, and FLIP recipes. This Computational Fluid Dynamics Library is called CFDLib. The CFDLib project keeps the most valuable parts of each historical recipe up-to-date, and available for use on new and ongoing studies by Laboratory researchers as well as academic and industrial users worldwide.

CFDLib is an active research code, which means it is constantly undergoing extension and evolution. This is in strong contrast to a commercial CFD code, which is by necessity frozen in time so that proper documentation can be sold to users. (Hence CFDLib is a useful tool for sophisticated users who are capable of navigating the source code and tailoring it to their own particular needs.)

The CFDLib project can be thought of as an open-source project wherein those who use it, and who make extensions or improvements, can submit those extensions for permanent inclusion in the library. This means that the contributor list is very long and includes academic and industrial users as well as the Laboratory custodians. Principal among the Laboratory contributors are Nely T. Padial-Collins, Rick M. Rauenzahn, Norman L. Johnson, John R. Baumgardner, John K. Dukowicz, Charles Zemach, and I. Significant contributions have come from other Laboratory groups as well as Sandia National Laboratories, American Oil Company, Rocket Research Company, University of New Mexico, and University of Utah.

An Example of a Fluid Dynamics Study

As an example of a fluid dynamics study that is aided by computer simulation, consider the simple problem of a single drop of milk falling onto a thin film of milk on the kitchen counter. Viewing the dynamics of the ensuing splash was made famous by Harold E. Edgerton, using flash photography. [1] Edgerton called the beautiful pattern of the early splash a "Milk Drop Coronet" because of its strong resemblance to a crown. Harlow and Amsden performed the first computer simulations of this drop splash in two dimensions using the MAC method. [2] Since that time the 3-D Milk Drop Coronet has been a highly prized goal of many CFD practitioners. Figure 1 has a sequence of six frames from a modern CFDLib computer simulation of a milk-drop splash. The physical process of interest in the simulation is the conversion of the initial kinetic energy in the falling drop, into energy associated with the surface tension of the small droplets caused by the splash, and energy lost to viscous dissipation. In the simulation shown, the conditions are chosen to be very close to the point where splashing occurs; at a slightly lower-impact velocity, no splash happens at all. Therefore in the simulation, the droplets are large and relatively few. For a larger-impact velocity there are many more droplets, of a smaller size, produced by the splash.

In order to carry out the foregoing simulation, the CFDLib code for incompressible, multifield flow is used. This is a MAC method for multiple fluids (air and milk, in this case). The physical model being tested is the Continuum Surface Force (CSF) formulation due to Brackbill, Kothe, and Zemach. [3] The result shows that for the conditions tested, the CSF physical model is reliable.

Summary

In summary, the CFDLib project endeavors to keep many parts of the division's numerical recipe legacy in a modern and accessible form, using an open-source model for development. This has furnished a substantial springboard for the sophisticated CFD analyst for use in a variety of problems in academia, industry, and in mission-related pursuits at the Laboratory.

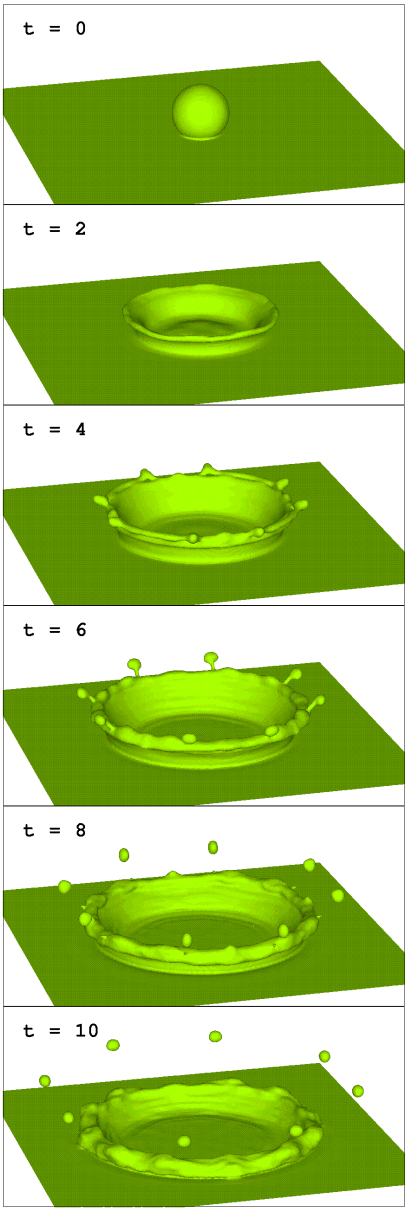


Fig. 1. A sequence of six frames from a modern CFDLib computer simulation of a milk-drop splash. In the simulation shown, the conditions are chosen to be very close to the point where splashing occurs; at a slightly lower-impact velocity, no splash happens at all. Hence, in the simulation the droplets are large and relatively few. For a larger-impact velocity there are many more droplets, of a smaller size, produced by the splash.

A Method for the Numerical Calculation of Hydrodynamic Shocks

Richtmyer
von Neumann

by John von Neumann and Robert Richtmyer

This article is reprinted with permission from John von Neumann and Robert D. Richtmyer, "A Method for the Numerical Calculation of Hydrodynamic Shocks," *J. Appl. Phys.* **21** (3), 232.

Von Neumann throughout the war was a consultant to Los Alamos. He spent, toward the end, I think, as much as half of his time there. When he was there he was a member of Theoretical Physics Division.

—George B. Kistiakowsky [chief of Explosives Division in World War II Los Alamos], "Reminiscences of Wartime Los Alamos," in Reminiscences of Los Alamos: 1943–1945, pp. 61-62.

The equations of hydrodynamics are modified by the inclusion of additional terms that greatly simplify the procedures needed for stepwise numerical solution of the equations in problems involving shocks. The quantitative influence of these terms can be made as small as one wishes by choosing a sufficiently fine mesh for the numerical integrations. A set of difference equations suitable for the numerical work is given, and the condition that must be satisfied to ensure their stability is derived.

Introduction

In the investigation of phenomena arising in the flow of a compressible fluid, it is frequently desirable to solve the equations of fluid motion by stepwise numerical procedures, but the work is usually severely complicated by the presence of shocks. The shocks manifest themselves mathematically as surfaces on which density, fluid velocity, temperature, entropy and the like have discontinuities; and clearly the partial differential equations governing the motion require boundary conditions connecting the values of these quantities on the two sides of each such surface. The necessary boundary conditions are, of course, supplied by the Rankine-Hugoniot equations, but their application is complicated because the shock surfaces are in motion relative to the network of points in space-time used for the numerical work, and the differential equations and boundary conditions are non-linear. Furthermore, the motion of the surfaces is not known in advance but is governed by the differential equations and boundary condition themselves. In consequence, the treatment of shocks requires lengthy computations (usually by trial and error) at each step, in time, of the calculation.

We describe here a method for automatic treatment of shocks which avoids the necessity for application of any such boundary conditions. The approximations in it can be rendered as accurate as one wishes, by suitable choice of interval sizes and other parameters occurring in the method. It treats all shocks, correctly and automatically, whenever and wherever they may arise.

The method utilizes the well-known effect on shocks of dissipative mechanisms, such as viscosity and heat conduction. When viscosity is taken into account, for example, the shocks are seen to be smeared out, so that the mathematical surfaces of discontinuity are replaced by thin layers in which pressure, density, temperature, etc. vary rapidly but continuously. Our idea is to introduce (artificial) dissipative terms into the equations so as to give the shocks a thickness comparable to (but preferably somewhat larger than) the spacing of the points of the network. Then the differential equations (more accurately, the corresponding difference equations) may be used for the entire calculation, just as though there were no shocks at all. In the numerical results obtained, the shocks are immediately evident as near-discontinuities that move through the fluid with very nearly the correct speed and across which pressure, temperature, etc. have very nearly the correct jumps.

It will be seen that for the assumed form of dissipation (and, indeed, for many others as well), the Rankine-Hugoniot equations are satisfied, provided the thickness of the shock layers is small in comparison with other physically relevant dimensions of the system. We then consider one way in which the transition from differential to finite-difference equations can be made and we discuss the mathematical stability of these equations. It will be seen that the dissipative terms have the effect of making the stability condition more stringent than the familiar one of Courant, Friedrichs, and Levy, but not seriously so if the amount of dissipation introduced is only enough to produce a shock thickness comparable with the spatial interval length of the network used.

The validity of our methods has been tested empirically on various computational applications.

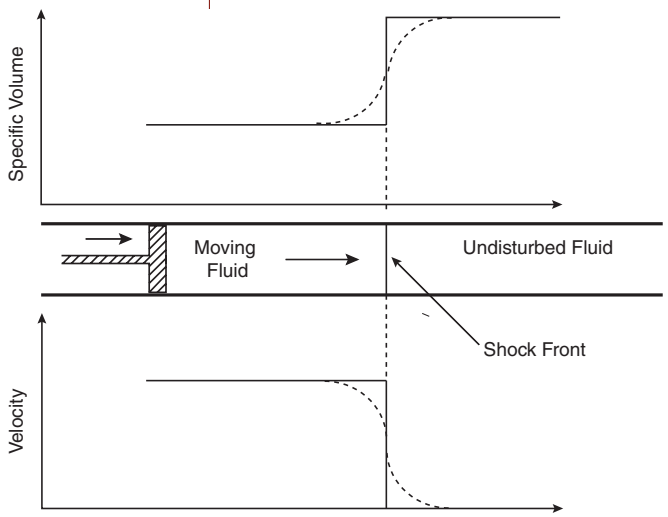


Fig. 1. Steady-State Plane Shock.

Imagine a long pipe containing a fluid initially in equilibrium (thermally and mechanically), into which a piston is pushing from one end with constant speed, as shown in Fig. 1. In the absence of dissipation the specific volume, V, and the fluid velocity, U, are, at a given instant, as shown by the solid curves in the graphs, whereas in the presence of dissipation they are as shown by the dashed curves. In either case, the shock is steady, at least approximately, after it has gone to a sufficiently great distance from the initiating piston.

Analysis of Engineering Flow Applications

Amsden
Brolley
Butler
Cline
Cort
Emigh
Farmer
Gentry
Hyde
Jenson
D. Martin
O'Rourke
Rivard
Robinson
Sahota
Thode
Wallace
Wantuck
Wilmoth

by Michael C. Cline

The VNAP series of codes were developed to support a wide variety of engineering projects at Los Alamos involving nuclear energy, chemical lasers, chemical-vapor deposition, isotope separation, neutral particle beams, aerodynamics, and air-breathing, rocket, and nuclear propulsion. In addition, these codes have enjoyed widespread use at universities, government laboratories, and aerospace companies around the world.

The VNAP series of codes includes the NAP, VNAP, and VNAP2 computer programs. The NAP code [1] was begun while I was a National Academy of Sciences postdoctoral research associate at the U.S. National Aeronautics and Space Administration (NASA)-Langley. After being hired into T-Division by Group Leader Frank Harlow, the NAP code was completed at Los Alamos. The follow-on VNAP [2] and VNAP2 [3] codes were developed at Los Alamos.

Code Descriptions

NAP is a computer program for calculating inviscid, unsteady flow in two-dimensional (2-D) and axisymmetric nozzles. NAP solves the time-dependent inviscid flow equations. Interior mesh points are computed using the MacCormack finite-difference scheme, but a characteristic scheme is used to calculate the boundary mesh points. An explicit artificial viscosity term is included for shock computations. The fluid is assumed to be a perfect gas.

VNAP is a computer program for calculating viscous, unsteady flow in 2-D and axisymmetric nozzles. VNAP solves the time-dependent viscous flow equations. VNAP also includes a mixing-length turbulence model for shear-layer flows.

VNAP2 is a computer program for calculating turbulent, unsteady flow in 2-D and axisymmetric geometries. VNAP2 also solves the viscous flow equations and includes algebraic mixing-length and partial-differential-equation turbulence models. The flow boundaries may be a combination of arbitrary curved solid walls, inflow/outflow boundaries, and free-jet envelopes. Over 125 copies of the VNAP2 code have been distributed to U.S. and foreign universities, government laboratories, and aerospace companies.

Applications

One of the milestones in the Controlled Thermonuclear Research (CTR) program was a thorough knowledge of the effects of high-level neutron irradiation on materials to be used in the CTR program. Bob Emigh and John Brolley of Physics (P) Division proposed that such knowledge could be obtained by a high-fluence 14-MeV neutron source with a supersonic-gas target. This gas target consisted of an annular expansion nozzle in which deuterium gas was accelerated to supersonic velocities and a central region where the deuterium interacted with a tritium-ion beam producing the neutrons. The NAP code was used as part of a theoretical and experimental program at Los Alamos to design the gas target. As a follow-on, the NAP code was used to compare the original supersonic design with a subsonic configuration developed by Jack Hyde of P- and later T-Division. While the facility wasn't built due to the less-than anticipated progress of the U.S. fusion energy program, this research helped solidify the design criterion for an intense neutron source.

One of the crucial areas for the Air Force Weapons Laboratory hydrogen-fluorine (HF) chemical laser program was the amount of mixing of the hydrogen and fluorine inflow streams. To achieve the necessary level of mixing, many small nozzles were employed. These small nozzles, combined with the low density required for efficient laser operation, resulted in fully viscous flows. The VNAP code was used to calculate the fully viscous flow for both the hydrogen and fluorine nozzles in order to define the inflow conditions for the downstream laser region calculations. These laser region calculations used the RICE code developed by Bill Rivard, Otis Farmer, and Dan Butler of T-Division.

The Air Force Rocket Propulsion Laboratory sponsored a program at Los Alamos, led by Terry Wallace of Materials Science and Technology Division, to study a chemical-vapor-deposition furnace used to prepare pyrographite/silicon carbide codeposited material for rocket nozzle applications. The VNAP code was used to compute the flow and heat transfer inside the chemical-vapor-deposition furnace. The gas flow inside the axisymmetric furnace was characterized by inviscid, low-speed flow with a large separated region. The results were used to help interpret the experimental measurements made by Ed Cort of Design Engineering Division.

The Los Alamos Jumper program was developed to separate isotopes of uranium and plutonium using photochemistry. In this process, proposed by Paul Robinson and Reed Jensen of Laser Research and Technology Division, a tuned laser was used to selectively excite a particular isotope in a highly cooled supersonic stream of uranium or plutonium hexafluoride vapor followed by condensation and collection

of the desired isotope. The VNAP code was used to analyze the supersonic nozzle and downstream laser interaction region including the effect of laser heat addition on the supersonic flow. These results were used to help guide the experimental program of Harry Watanabe and coworkers of Applied Photochemistry Division.

A competing process to laser isotope separation is the gas centrifuge. A collaboration between Oak Ridge National Laboratory and Los Alamos, led by Dick Gentry of T-Division, studied the use of an advanced gas centrifuge to separate isotopes of uranium. The VNAP2 code was used to analyze and design the low Reynolds number supersonic scoops used to collect the desired isotopes. These results, along with experiments at Oak Ridge and Jack Hyde's experiments at Los Alamos, were used to improve the efficiency of gas centrifuges for producing nuclear reactor fuel.

The Neutral Particle Beam Program, part of the Strategic Defense Initiative, included the Ground Test Accelerator (GTA) at Los Alamos. The GTA was a ground-based engineering prototype facility that included a gas neutralizer to convert the negative hydrogen ion beam to a neutral particle beam for which targeting is unaffected by the Earth's as well as local magnetic fields. Dick Martin, of Engineering Sciences and Applications Division, used the VNAP2 code to compute the flow inside the radial-injection gas neutralizer. These results were compared with experimental data to benchmark the code and verify its use in the evaluation of new design concepts.

NASA-Langley's Propulsion Aerodynamics Branch pursued an experimental/theoretical program to study the interaction of inlet and outlet flow of a jet engine with the external flow over the airframe. The object of this research was to improve the installed performance prediction capability for jet aircraft engines. The VNAP2 code was used to calculate many configurations with both internal flow only and internal/external flow interactions. These results were compared to another theoretical effort developed by Dick Wilmoth of NASA-Langley and experimental data from NASA-Langley's Sixteen Foot Transonic Wind Tunnel. These comparisons are presented in [3]. After this benchmarking, VNAP2 was added to the Propulsion Aerodynamics Branch's performance prediction toolbox of computational fluid dynamics (CFD) codes.

An unusually large over-pressure on the first Space Shuttle launch (STS-1) exceeded preflight predictions by as much as 5 to 1 causing minor damage to the Orbiter. Los Alamos was part of a joint study with Dick Wilmoth of NASA-Langley to determine the cause of this over-pressure problem. The VNAP2 code was used to calculate the unsteady, turbulent flow during the start-up of the Solid Rocket Booster. (See Fig. 1.)

The National Aero-Space Plane (NASP) Program was to develop a single-stage, horizontal take-off vehicle for low Earth orbit. The main propulsion system was to be a hydrogen fueled supersonic ramjet or scramjet engine. One of the NASP-related programs at Los Alamos was to experimentally study chemically-reacting shear layer flows. The VNAP2 code was used to analyze the fuel and oxidizer nozzles as well as the shear layer region for nonreacting flow. In addition, the nozzle flow solutions were used to provide the upstream conditions for chemically reacting flow calculations using the KIVA-II code developed by Tony Amsden, Peter O'Rourke, and Dan Butler of T-Division. These results were used to help interpret the experiments of Paul Wantuck and coworkers of Chemistry and Laser Sciences Division. One important result from these comparisons was the necessity of measuring the turbulence levels upstream of the fuel and oxidizer nozzles in all future experiments.

One propulsion system under consideration for deep space travel is the open-cycle gas-core nuclear rocket (OCGCNR). In an OCGCNR, the energy liberated from a central mass of near-critical fissioning uranium fuel is radiatively coupled to the hydrogen working fluid as it flows around the uranium and then is expanded out a propulsion nozzle. Les Thode of Applied Physics Division used the VNAP2 code to study an axisymmetric OCGCNR in collaboration with NASA-Marshall. As part of this study, the VNAP2 code was benchmarked against two commercial CFD codes as well as the KIVA-II and CHAD codes with excellent results. The CHAD code was developed by Manjit Sahota and Peter O'Rourke of T-Division.

The VNAP series of codes also was used in five other Los Alamos projects involving rocket exhaust plumes, aerodynamics, and the characterization of flows in several laboratory experiments. This latter case involves experiments where the time-of-flight of a flowing material is required in order to interpret an experimental result but for some reason could not be measured. The VNAP series of codes also was used in 22 external collaborations with universities, government laboratories, and aerospace companies.

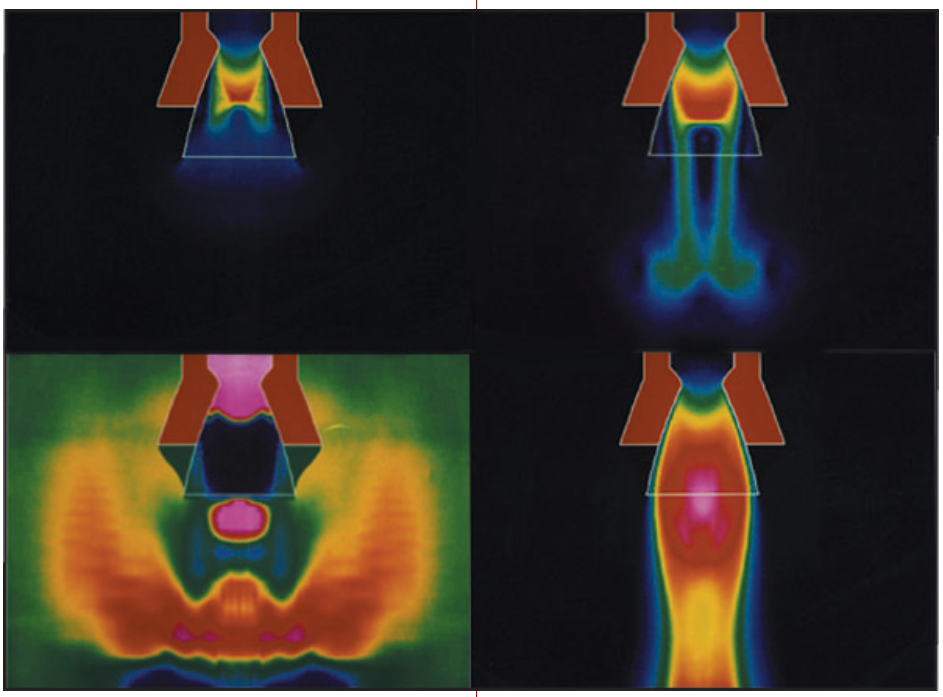


Fig. 1. The flow of combustion gas from the Solid Rocket Booster (SRB) propulsion nozzle at various times during the first 200 milliseconds after SRB ignition. The first three frames, clockwise starting with the top-left, show combustion-gas speed contours at increasing time after ignition (blue-low, red-high). The bottom-left frame shows pressure contours at the time of maximum over-pressure. These results along with those obtained by NASA were used to develop modifications to the launch platform that successfully alleviated the over-pressure problem.

Simulation of Mixing Layers

Clark

by Timothy T. Clark

Researchers in Theoretical Division used a lattice Boltzmann method to perform a series of direct numerical simulations of two-dimensional Rayleigh-Taylor mixing layers to support modeling efforts at Los Alamos.

The development of accurate, computationally tractable mathematical models of turbulent mixing layers requires detailed information regarding the evolution of these layers. To support the modeling efforts at Los Alamos National Laboratory, Theoretical Division researchers used a lattice Boltzmann method to perform a series of direct numerical simulations of two-dimensional Rayleigh-Taylor mixing layers.

The series consists of 104 statistically independent realizations of the evolving mixing layer and represents a comprehensive database for studying the adequacy of various assumptions embedded in the turbulent mix models. In addition, the results of these simulations are shedding new light on the nature of the self-similarity that emerges within these layers, thus providing crucial guidance to the model-developing efforts.

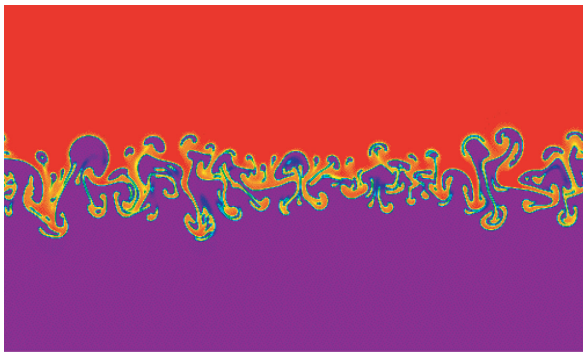
We anticipate the development of a database of three-dimensional mixing-layer simulations and extraction of further information to support modeling efforts at Los Alamos.

Figures 1a to 1f show the evolution of the mixing layer as it evolves from the initial fine-scale perturbations to a late stage, which is dominated by much larger scales.

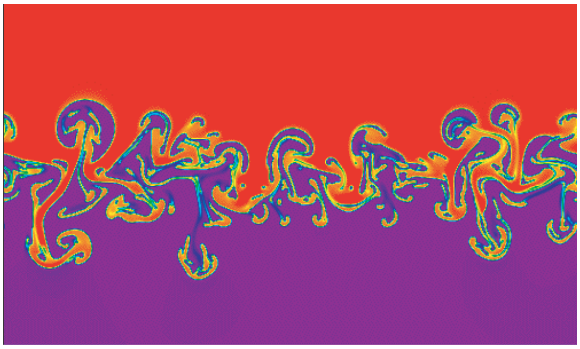
(1a)

(1b)

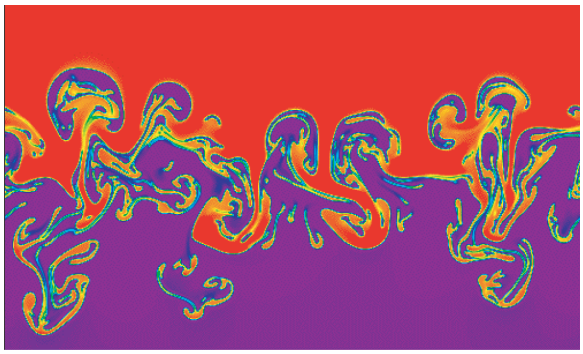
(1c)



(1d)



(1e)



(1f)

Fig. 1a–1f: Evolution of the mixing layer.

Numerical Solutions for Solving the Neutron Transport Equation

B. Carlson
Chandrasekhar
Davison
Feynman
Lathrop
Mark
Miller
Morel
Placzek
Serber
Sessions
Wick
Wing

by Denise Sessions (from interview with Bengt Carlson)

When Bengt Carlson arrived in Los Alamos in early June 1945 as part of the British Mission from Canada, he found it too late to be involved in the beehive of activity surrounding the preparation for the Trinity test. "It was a very exciting time because there were no distractions; everyone was so dedicated to the project that they worked constantly," he recalls. Carlson and his colleagues George Placzek and Carson Mark had been members of a British-Canadian atomic energy project in Montreal for the purpose of designing a heavy water pilot plant for the production of plutonium.

Nuclear Reactor Theory in the Montreal Laboratory

The Montreal Laboratory project was not directly part of the Manhattan Engineering District (MED) operation at Los Alamos, but it maintained a close liaison with the MED. Though rather simple approximations to the solutions of the neutron transport equation were sufficient for most problems met in the design of a moderated reactor, Carlson's colleague Placzek was well aware that much more accurate solutions to that equation would be necessary for fast-neutron problems of the sort arising at Los Alamos.

By sometime in 1944 Placzek's group had succeeded in outlining an approach, which, though quite cumbersome to apply, would be capable of providing numerically accurate solutions to some types of problems of interest. Word of this was welcomed at Los Alamos and some application of the method was made there. Thus, Placzek, Mark, and Carlson came to Project Y to work on more advanced weapon designs to use precious materials more efficiently.

Early Days in T-Division

Carlson says, "Living in Los Alamos in the early days was like living on a frontier in several ways: a frontier town, with people from everywhere mixing with local Hispanics and Native Americans, and a frontier in the fields of physics and computing." The tremendous spirit that Carlson encountered at Los Alamos continued for the next twenty years. "I enjoyed working with the Nobel Prize winners and so many people with technical experience." Carlson headed the Computing Center at Los Alamos Scientific Laboratory from 1951 to 1968.

The S-N Method

[Sessions] "I understand that you were the inventor of the S-N technique for calculating transport of radiation. Can you tell me a little bit about this method (in layman's terms) what sorts of problems could be solved with it?

[Carlson] "The S-N method is a numerical method for approximate solutions to the linear Boltzmann equation, a familiar integro-differential equation in physics, governing the distribution of particles in a given assembly.

The particles may, for example, be neutrons, photons or electrons, and the assembly a nuclear reactor, atomic bomb configuration, or an experimental setup. In the general case, the distributions depend on the following variables: the space coordinates for the assembly, the velocity and direction of the particles, and time. They also depend on the nature of the materials in the assembly, i.e., on what is known about the interactions between the particles and the nuclei in the materials.

Then, having found the distributions, one can calculate the parameters of interest for the reactor, bomb, or setup in question.

The S-N method is built on various older mathematical ideas but made more general and translated to finite difference formulas and then to computer language. The "N" in S-N relates to the number of discrete directions chosen. The base ideas were simplifications using step functions (discrete ordinates), connected line segments (early S-N), expansions in series (spherical harmonics), and transformation to integral equations. These are associated with names such as S. Chandrasekhar (astrophysics applications), C. Mark and B. Davison (spherical harmonics) and G. Wick, G. Placzek, and R. Serber (special applications). Discrete S-N depends on specific difference schemes such as the Diamond and Weighted Diamond schemes, and, recently, an Exponential scheme.

The first S-N formulation and code were based on connected line segments and stimulated by the expected arrival in 1953 of the IBM Type 701 high-speed electronic computer. The first codes were for simple geometries (planes and spheres) with no time dependence and limited number of velocity groups. My assistant in this early effort was Janet Bendt Wing, still a Los Alamos resident. In reactor work a simpler method akin to S-2, called the Diffusion Theory method, was accurate enough for most applications.

The first large computer setup at Los Alamos was that done by R. Feynman in 1944 to solve certain hydrodynamics problems (1944–50). It consisted of a group of IBM accounting machines, including an advanced multiplier. Next came three IBM card-programmed computers (CPCs) in 1949–54. The MANIAC was ready for use in 1952. Thereafter, until 1968, it was mostly IBM, 701s in the period 1953–56, 704s in the period 1955–62, and in the period 1961-71, MANIAC II, STRETCH, and two 7094s.

[Sessions] "I understand that it has received widespread usage around the world. Can you tell me a little bit about the other places where the method is used?"

[Carlson] "A paper on S-N was given at the second Peaceful Uses Conference in Geneva 1958, and on an improved formulation (discrete S-N with appropriate conservation laws built) in 1964 at the third conference. Beginning in the 1960s, the method was used at many laboratories, Argonne and Oak Ridge in particular, and places in Europe. In 1966, the Atomic Institute in Bucharest published a bibliographical booklet on S-N.

The method has its problems, but efforts continue to this day to improve its accuracy and reliability and to expand its usefulness. At Los Alamos, Pete Miller and Jim Morel and his group are involved, I believe, and in Colorado retirement Kaye Lathrop, working with Jim Morel, is making progress. Me, too, sporadically."

Hasslacher
Ulam
von Neumann

Discrete Fluids

by Brosl Hasslacher

This article is excerpted with permission from *Los Alamos Science, Special Issue: Stanislaw Ulam 1909–1984* (Los Alamos National Laboratory, 1987), pp. 175–219.

Background for Lattice-Gas Automata

The invention of a totally discrete model for natural phenomena was made by Ulam and von Neumann in the early fifties and was developed to the extent possible at the time. A few years earlier, von Neumann had designed the architecture for the first serial digital computers containing stored programs and capable of making internal decisions.

These machines are built of electronic logic devices that understand only packets of binary bits. Hierarchies of stored translators arrange them into virtual devices that can do ordinary or radix arithmetic at high speed. By transcribing continuum equations into discrete form, using finite difference techniques and their variants, serial digital computers can solve complex mathematical systems such as partial differential equations. Since most physical systems with large numbers of degrees of freedom can be described by such equations, serial digital machines equipped with large memories have become the standard way to simulate such phenomena.

As the architecture of serial machines developed, it became clear to both Ulam and von Neumann that such machines were not the most natural

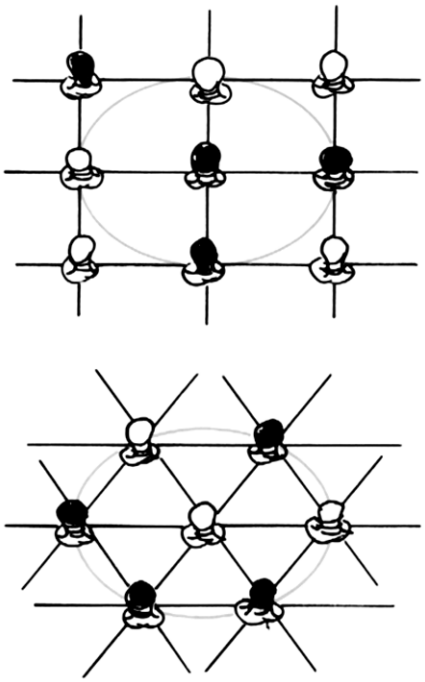
or powerful way to solve many problems. The scientists were especially influenced by biological examples. Biological systems appear to perform computational tasks by using methods that avoid both arithmetical operations and discrete approximations to continuous systems.

Though motivated by the complex information processing of biological systems, Ulam and von Neumann did not study how such systems actually solve tasks. Biological processes have been operating in hostile environments for a long time, finding the most efficient and often devious way to do something, a way that is also resistant to disturbance by noise. The crucial principles of their operation are hidden by the evolutionary process. Instead, von Neumann chose the task of simulating on a computer the least complex discrete system capable of self-reproduction. Ulam suggested an abstract setting for this problem and many other totally discrete models, namely, the idea of cellular spaces. The reasoning went roughly like this.

The question is simple: find a minimal logic structure and devise a dynamics for it that is powerful enough to simulate complex systems. Break this up into a series of sharper and more elementary pictures. We begin by setting up a collection of very simple finite-state machines with, for simplicity, binary values. Connect them so that given a state for each of them, the next state of each machine depends only on its immediate environment. In other words, the state of any machine will depend only on the states of machines in some small neighborhood around it. This builds in the constraint that we only want to consider local dynamics.

We will need rules to define how states combine in a neighborhood to uniquely fix the state of every machine, but these can be quite simple. The natural space on which to put all this is a lattice, with elementary, few-bit, finite-state machines placed at the vertices. The rules for updating this array of small machines can be done concurrently in one clock step, that is, in parallel.

One can imagine such an abstract machine in operation by thinking of a fishnet made of wires. The fishnet has some regular connection geometry, and lights are at the nodes of the net. Each light can be on or off. Draw a disk around each node of the fishnet, and let it have a 1-node radius. On a square net, four lights are on the edge of each disk; on a triangular net, six lights (Fig. 1). The next state of the light at the center of the disk depends on the state of the lights on the edge of the disk and on nothing else. Imagine all the disks in the fishnet asking their neighbors for their state at the same time and switching states according to a definite rule. At the next tick of an abstract clock, the pattern of lights on the fishnet would in general look different. This is what Ulam and von Neumann called a nearest-neighbor-connected cellular space. It is the simplest case of a parallel computing space. We can also see that this case can be imaged directly in hardware, so it is also the architecture for a physical parallel computing machine.



CELLULAR SPACES
Fig. 1. Two examples of fishnets made of wires with lights at the nodes. The lights are either on or off. In each example, a disk with a radius of 1 node is drawn around one of the lights. The next state of the light at the center depends on the states of the lights on the edge of the disk and on nothing else. Thus, these are examples of nearest-neighbor-connected cellular spaces.

We have not shown that such a device can compute. At worst, it is an elaborate light display. Whether or not such a cellular space can compute depends on the definition of computation. The short answer is that special cases of fishnets are probably universal computers in the standard Turing machine sense; that is, they can simulate the architecture of any other sequential machine.

But there are other interpretations of computation that lie closer to the idea of simulation. For any given mathematical simulation, we want to find the minimum cellular space that can do a simulation of it: At what degree of complexity does repeated iteration of the space, on which are coded both data and a solution algorithm, possess the power to come close to the solution of a complex problem? The answer depends on the complexity or degrees of freedom present in the problem.

An extreme case of complexity is physical systems with many degrees of freedom. These systems are ordinarily described by field theories in a continuum for which the equations of motion are highly nonlinear partial differential equations. Fluid dynamics is an example, and we will use it as a theoretical paradigm for many "large" physical systems. Because of the high degree of nonlinearity, analytic solutions to the field equations for such systems are known only in special cases. The standard way to study such models is either to perform experiments or simulate them on computers of the usual digital type.

Suppose a cellular space existed that evolved to a solution of a fluid system with given boundary conditions. Suppose also that we ask for the simplest possible such space that captured at least the qualitative and topological aspects of a solution. Later, one can worry about spaces that agree quantitatively with ordinary simulations. The problem is threefold: Find the least complex set of rules for updating the space, the simplest geometry for a neighborhood, and a method of analysis for the collective modes and time evolution of such a system.

At first sight, modeling the dynamics of large systems by cellular spaces seems far too difficult to attempt. The general problem of a so-called "inverse compiler"—given a partial differential system, find the rules and interconnection geometry that give a solution—would probably use up a non-polynomial function of computing resources and so be impractical if not impossible. Nevertheless, cellular spaces have been actively studied in recent years. Their modem name is cellular automata, and specific instances of them have simulated interesting nonlinear systems. But, until recently, no example of a cellular automaton simulated a large physical system, even in a rough, qualitative way.

Knowing that special cases of cellular automata are capable of arbitrarily complex behavior is encouraging, but not very useful to a physicist. The important phenomenon in large physical systems is not arbitrarily complex behavior, but the collective motion that develops as the system evolves, typically with a characteristic size of many elementary

length scales. The problem is to simulate such phenomena and, by using simulations, to try to understand the origins of collective behavior from as many points of view as possible. Fluid dynamics is filled with examples of collective behavior-shocks, instabilities, vortices, vortex streets, vortex sheets, turbulence, to list a few. Any deterministic cellular-automaton model that attempts to describe nonequilibrium fluid dynamics must contain in it an iterative mechanism for developing collective motion. Knowing this principle and using some very basic physics, we will construct a cellular automaton with the appropriate geometry and updating rules for fluid behavior…

Continuing from p. 203:

Using Lattice-Gas Methods to Approximate Hydrodynamics

In August 1985 Frisch, Hasslacher, and Pomeau demonstrated that one can approximate solutions to the Navier-Stokes equations by using lattice-gas methods, but their demonstration applied only to low-velocity incompressible flows near equilibrium. No one knew whether more interesting flows could be approximated. Consequently, computer codes were written to determine the region of validity of the lattice-gas method. An example of the results of some of the first simulations done at Los Alamos is shown in Fig. 2. (Most of the early calculations were done on a Celerity computer, and the displays were done on a Sun workstation.) All the results indicate qualitatively correct fluid behavior.

Figure 2 demonstrates that a stable trailing vortex pattern develops in a two-dimensional lattice-gas flowing past a plate… Stable vortices develop in a lattice gas at the interface between fluids moving in opposite directions. The Keivin-Helmholtz instability is known to initiate such vortices. The fact that lattice-gas methods could simulate vortex evolution was reassuring and caused several scientists to begin to study the new method… A complicated wake develops behind a V-shaped wedge in a uniform-velocity flow. With a Reynolds number of 76, the flow has a stable period of oscillation that slowly grows to its asymptotic limit. In a flow with a higher Reynolds number past an ellipse, the wake becomes chaotic and quite sensitive to details of the flow…

A three-dimensional flow around a square plate was one of the first results from Los Alamos in three-dimensional lattice-gas hydrodynamic simulations. The results compared well with theoretical predictions.

These calculations, and many others, have established some confidence that qualitative features of hydrodynamic flows are simulated by lattice-gas methods.

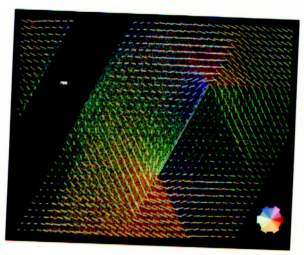
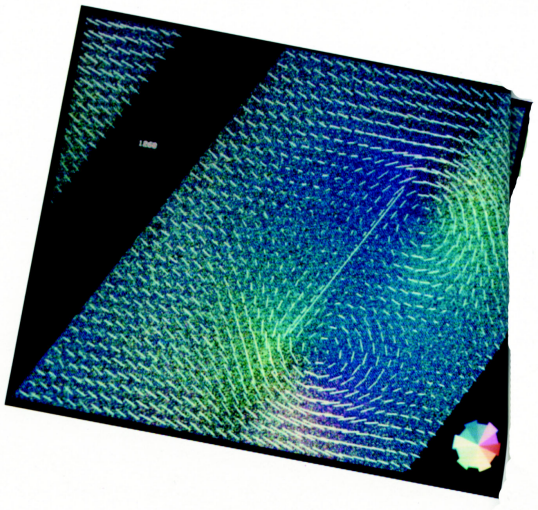
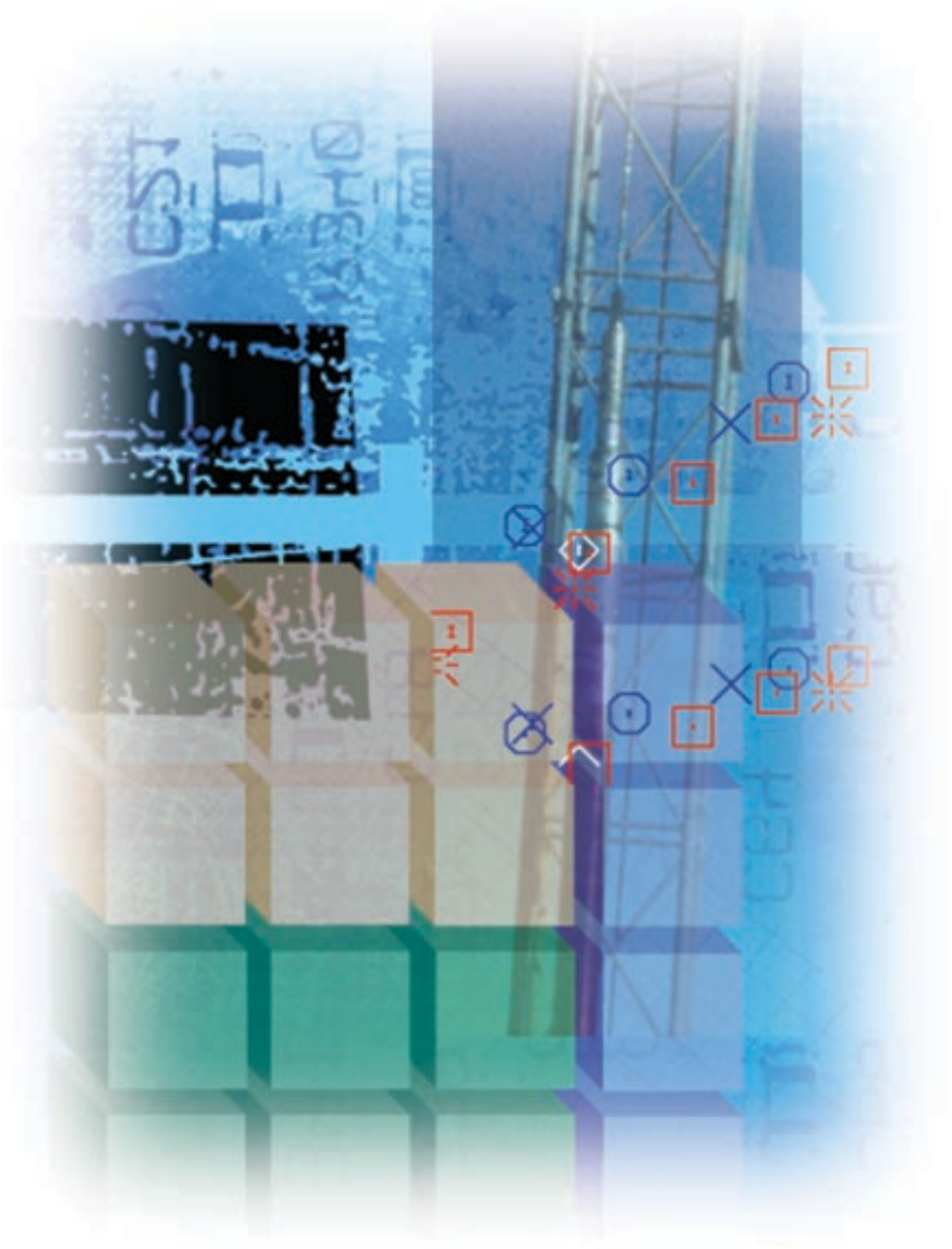


Fig. 2. Flow past a plate with periodic boundary conditions. This simulation, which was done in September 1985, shows vortices forming behind the plate. The average flow velocity has a magnitude of 0.2 lattice sites per time step and is perpendicular to the plate, pointing to the lower right. The direction of the flow velocity is color coded.





The Standard Model and Its Validation on the Computer

by Tanmoy Bhattacharya and Rajan Gupta

Baillie
Bhattacharya
Brickner
Canada
Daniel
DeLapp
Feynman
Gell-Mann
Goldman
Grandy
Gupta
Guralnik
Herczeg
Kilcup
Lee
Patel
Pope
Sharpe
Tamayo
Terning
Warnock
Zemach
Zweig

Lattice Quantum Chromodynamics (QCD) calculations at Los Alamos have had two goals—to push the envelope of high-performance computing and to validate QCD, a fundamental theory of nature. The beauty of lattice QCD calculations is that we create a laboratory in the computer—we can dial the six fundamental parameters of the theory, see how nature changes in response and determine whether there is a set of values for these parameters that reproduces all observables.

Advances in the simulations of lattice QCD are intertwined with the development of parallel- and high-performance computing. We developed a suite of highly optimized codes that stress every component of the machines and paved the way for efficient use of these machines by other users. Our code has also been the most reliable diagnostic for both the hardware and the software. From the CM2 to the CM5 to the Silicon Graphics Origin 2000 computer, this code was taken up by the manufacturer and used as a core component of their diagnostics. Even today, the QCD code continues to be used as a diagnostic and performance monitor on the state-of-the-art terascale computers at Los Alamos.

Advent of the Standard Model

The twentieth century has been called the century of physics. Over this period unprecedented advances were made in our understanding of the basic constituents of nature and the forces between them. At the end of the nineteenth century physicists viewed atoms as the indivisible particles arranged in the periodic table of elements. The only known forces were gravity and electromagnetism. By 1920, Bohr's picture of the atom as consisting of a nucleus surrounded by a cloud of electrons had taken hold.

In the mid-1920s quantum mechanics was developed to explain the stability and the excitation spectrum of atoms. This was quickly followed by Dirac's formulation of relativistic quantum mechanics which, remarkably, predicted the existence of antiparticles. The merger of quantum mechanics with the special theory of relativity ineluctably led physicists to quantum field theory (QFT); however, it took until the early 1970s for field theories to be accepted as the way to describe fundamental interactions in nature. The main reason for this long hiatus was the plethora of strongly interacting particles (hadrons like pions, kaons, hyperons, etc.) discovered in the 1950s and 1960s, that did not appear to be point-like elementary particles or fit in a coherent framework as demanded by field theory. Quantum electrodynamics (QED)

remained the only successful example until the early 1970s, and even QED was regarded by many as only a computational scheme due to the problems of infinite self-interactions.

Between 1945 and 1975 a number of developments took place: Feynman's formulation of quantum mechanics as an integral over paths and the development of the Feynman diagram techniques; the mathematical formulation of non-abelian gauge theories by Yang and Mills; the quark model by Gell-Mann and Zweig; the discovery of point-like entities identified with quarks inside hadrons in deep inelastic scattering experiments; and the need for three colors of quarks. By 1973 these ideas had been put together to formulate Quantum Chromodynamics (QCD) as the theory of strong interactions with quarks and gluons as the fundamental particles and force carriers. Over the same period the understanding of spontaneous symmetry breaking as a mechanism for generating mass (the so-called Higgs Mechanism), the discovery of neutral currents in weak interactions that foreshadowed the discovery of Z^0 bosons, and the renormalization of spontaneously broken gauge theories led to the unification of electrodynamics and weak interactions. Thus, by 1975 the weak, strong, and electromagnetic interactions had been understood within one comprehensive framework based on non-abelian field theories and local gauge invariance and encapsulated in a mathematical theory called the Standard Model (SM). [1,2]

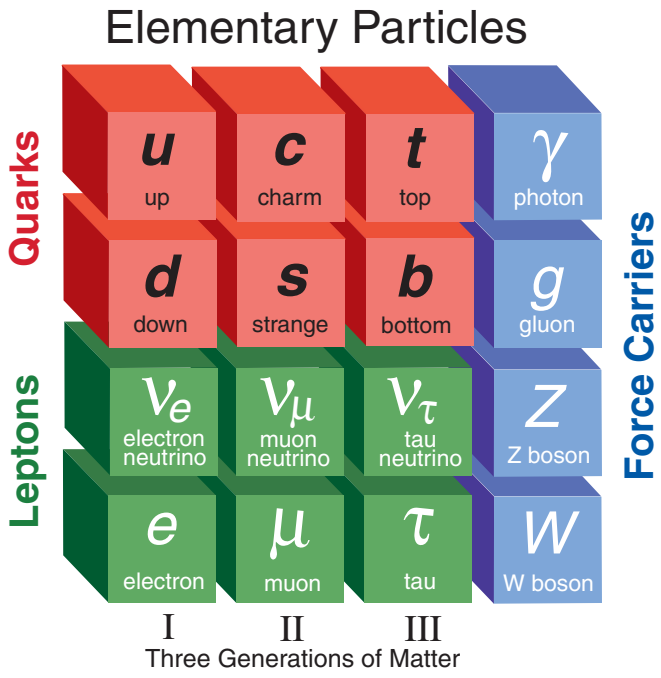


Fig. 1. The elementary particles in the Standard Model.

Our current understanding of the elementary particles of nature and their interactions is as follows. There are a total of 24 elementary particles that have spin 1/2 and obey Fermi statistics. These are the six flavors of quarks (up, down, strange, charm, bottom, and top), each of which comes in three colors—"red," "blue," and "green"—and the six leptons—electron, electron-neutrino, muon, muon-neutrino, tau, and tau-neutrino. Associated with each particle is its antiparticle. These particles interact through the exchange of spin 1 particles, called vector or gauge bosons. There are 12 such bosons that are related to the three different forces of nature—the eight gluons (g) mediate strong interactions between quarks, the photons (γ) mediate electromagnetic interactions, and the W^+, W^-, and Z^0 mediate weak interactions. Only the quarks and gluons carry the strong charge. A schematic representation of this set of elementary particles is shown in Fig. 1.

While this model, in its minimal form, is remarkably successful and consistent with all existing experimental data, it is not complete. Even ignoring the fact that it does not include gravity, it has a number of limitations. For example, the recently discovered phenomena of mixing between the different types of neutrinos and the associated tiny masses were not present in the original minimal SM.

More pertinent to our discussion is the origin of masses of the quarks, leptons, and the W^+, W^-, Z^0 particles. This is accomplished by adding spin 0 particles, called Higgs, to the quarks, leptons, and gauge bosons for the sole purpose of generating masses. The Higgs particles are coupled to the quarks and leptons in ratio of their masses and the common mass scale is the non-zero vacuum expectation value of the Higgs field. The final mass of a particle is the product of this scale and its coupling to the Higgs. One very important consequence of this construct is that it introduces at least one physical scalar particle into the theory. So, to confirm the Higgs mechanism one has to first observe this scalar particle and then show that its coupling to the quarks and leptons is in the ratio of their masses. Such a Higgs particle has not yet been observed (the current lower bound based on the SM on its mass, with 95% confidence level, is 114.8 GeV); consequently a very important aspect of the SM is unconfirmed.

The energy scale of the Higgs mechanism is fixed by the strength of weak interactions; however, within the minimal SM this scale is not stable. Many extensions like the Minimally Supersymmetric Standard Model, Technicolor, and extra spatial dimensions have been proposed to fix this problem, and each of these give rise to more fundamental particles and other observable consequences. So, a very large part of the current effort in high-energy physics is to decide which extension, if any, is correct (see article by Goldman, Herczeg, and Terning in this chapter). To answer this question requires the ability to make very precise predictions of both the SM and its extensions and to compare them to precision experiments.

Calculations in the electro-weak sector can be carried out with high accuracy because the weak and electromagnetic coupling constants are small; however this situation changes when one needs to consider the effects of quarks and gluons, even in intermediate states. At this order analytical methods, which are based on an asymptotic expansion in the coupling constant, become unreliable because the strong charge is of order unity. Similarly, all low-energy processes that involve quarks and gluons, either in initial or final state hadrons, cannot be calculated analytically, using perturbation theory, with sufficient precision. Strong interaction dynamics, as the name suggests, is intrinsically nonperturbative.

Quantum Chromodynamics

Very little of the low-energy interaction of quarks and gluons is understood at the quantitative level. For example, we cannot analytically calculate the spectrum of their bound states, in analogy to the energy levels of the hydrogen atom calculated using quantum mechanics given the masses of the proton and the electron and the force law between them. Furthermore, quarks and gluons are confined within hadrons, ostensibly because the strong force grows linearly with distance between quarks. One consequence of this "confinement" is that these elementary particles are not seen as asymptotic states in experiments; all one observes are bound states like protons and neutrons made up of quarks and gluons in color neutral combinations. Consequently, the masses of quarks cannot be directly measured but have to be inferred from the dependence of the masses of hadrons on them. Thus, the first computational challenge is to test QCD by computing its spectrum in terms of quark masses and the coupling constant and showing that this agrees with the observed masses of hadrons.

If all quark masses are zero, the fermionic sector of QCD has an additional symmetry, called chiral symmetry. A consequence of this symmetry is that opposite parity partners of every bound state of quarks and antiquarks must have the same mass, i.e., the states form parity doublets. This symmetry is observed to be broken in nature. This can happen in two ways—explicitly because the quarks have mass or through the phenomena of spontaneous symmetry breaking which, in a nutshell, means that while the interactions (the Hamiltonian) preserve the symmetry, the vacuum state chooses a particular representation that breaks the symmetry. In QCD both mechanisms are found to occur. The three light quarks, up, down and strange, have non-zero masses, but these are small in comparison to the scale of spontaneous chiral symmetry breaking (χSB); the other three quarks are heavier. The mechanism for χSB is not understood and its elucidation also requires new nonperturbative tools.

Lattice QCD

The starting point for nonperturbative analysis of QCD was Wilson's formulation in 1974 of quantum field theory on a Euclidian space time lattice using Feynman's path integral approach. [3] The lattice is simply an ultraviolet regulator, i.e., wavelengths shorter than the lattice spacing a are omitted, and the continuum theory is recovered by removing this scaffolding, i.e., by taking the limit $a \to 0$. Feynman's path integral approach in Euclidian time allows the generating functional for the QFT to be interpreted as the partition function of a statistical mechanics system. Approximating space-time by a finite set of discrete points makes the number of degrees of freedom finite and thus amenable to simulations on a computer.

The partition function for QCD in Euclidian space time is amazingly simple-looking, as follows:

$$Z = \int D\psi(x)\, D\overline{\psi}(x)\, DA_\mu(x)\, e^{-S_{gauge}-S_{fermion}} \quad ,$$

$$S_{gauge} = \frac{1}{4}\int F^{ab}_{\mu\nu}\, F^{\mu\nu}_{ab}\, d^4y \quad ,$$

$$S_{fermion} = -\int \overline{\psi}^a_i\, M^{ij}_{ab}\, \psi^b_j\, d^4y \quad ,$$

where S_{gauge} and $S_{fermion}$ are the actions for the gauge and fermion fields respectively; M is the Dirac operator; a,b are color indices; i,j are flavor indices; and μ,ν are Lorentz indices. The complexity arises because the functional integral in Z stands for a sum over all possible values of each field variable at each space time point. Since each bosonic variable $A\mu$ is continuous valued and space-time is continuous, this is an infinite-dimensional integral. In numerical simulations the problem is made tractable by choosing a finite hypercubic lattice of space-time points, and sampling a finite set of configurations—values of field variables at each space-time point. Also, the fermion degrees of freedom are integrated over at the expense of an additional, nonlocal interaction between the gauge fields, which is given by the determinant of the Dirac operator M, as

$$Z = \int DA_\mu(x)\, (Det\ M)\, e^{-S_{gauge}} \quad .$$

The theory in this form is simulated using a Markov chain Monte Carlo Method with the Metropolis algorithm. [4] The output of simulations consists of gauge configurations and quark propagators (calculated by inverting the Dirac operator M) evaluated on them. Physical observables are obtained by calculating correlation functions which, in the language of Statistical Mechanics, are expectation values:

$$\langle O \rangle = \frac{1}{Z}\int D\psi(x)\, D\overline{\psi}(x)\, DA_\mu(x)\, O\ e^{-S_{gauge}-S_{fermion}} \quad .$$

These are obtained by stitching together the gauge variable and quark propagators in appropriate combinations on each configuration and then averaged over the configurations. [4] A cartoon of the calculations is shown in Fig. 2.

Fig. 2. A cartoon of lattice QCD calculations. QCD, defined by six parameters, is simulated on the state-of-the-art computers ($$$) using the best algorithms (brains) and the outputs are gauge configurations and quark propagators from which expectation values are constructed. These contain systematic and statistical errors due to the discretization of the theory and Monte Carlo integration. These errors need to be resolved in order to make precise quantitative predictions.

Calculating these expectation values is an extremely demanding and challenging computational problem. In four dimensions even a reasonably sized ($32 \times 32 \times 32 \times 64$) lattice has two million points, and associated with each point are 36 complex valued variables for the colored gauge fields and twelve for each flavor of quarks (after integration over fermions, these are auxiliary variables used to calculate *Det M* and the quark propagators). Each step in the update of these lattice variables requires between $10^{10} - 10^{15}$ floating point operations depending on the approximations made and on the values of the light quark masses chosen. The estimates provided above are for masses equal to or larger than the strange quark and the complexity scales roughly as $m_q^{-4.5}$. The fact that the up and down quark masses are tiny ($m_u = 0.55\ m_d = 0.029\ m_s$) has necessitated that simulations involving light quarks be done at higher masses and the results extrapolated to their physical values. These extrapolations are expected to be reasonable if simulations are done with quark masses in the range $0.25\ m_s$–$2\ m_s$.

The most computationally intensive part is the calculation of the fermion determinant, *Det M*, in the update of the gauge fields. Even on tiny lattices, calculating this term is prohibitive. Thus, from very early days an approximation called the quenched approximation was designed. In this approximation, the determinant is set equal to a constant and therefore drops out of the functional integral. Physically this approximation corresponds to an alteration of the QCD vacuum—virtual quark-antiquark pair production and annihilation are eliminated. In the language of statistical mechanics this approximation corresponds to assigning the vacuum a static color dielectric constant. Even though this is a drastic modification of the theory, phenomenology indicated that the effect on a number of observables would be at the 10%–20% level provided the QCD coupling was appropriately tuned. It was therefore expedient to carry out initial simulations within this approximation to develop the methodology and to understand the statistical and systematic uncertainties, which are similar in the two theories.

The beauty of lattice QCD calculations is that we create a laboratory in the computer—we can dial the six fundamental parameters of the theory and see how nature changes in response and determine whether there is a set of values for these parameters that reproduces all effects of QCD. (These six parameters are the coupling α_s and the masses of up, down, strange, charm, and bottom quarks. The top quark is too short lived to form bound states and too heavy to affect other processes. Thus it is not included in the simulations.) Such calculations constitute a first principles test of the theory.

Lattice QCD Calculations at Los Alamos

Lattice QCD calculations at Los Alamos have had two goals—to push the envelope of high-performance computing and to validate QCD, a fundamental theory of nature.The first calculations undertaken at Los Alamos, by Guralnik, Warnock, and Zemach starting in 1983, were to estimate the masses of hadrons as a function of quark masses. These simulations were done on Cray supercomputers using a modular programming methodology. The most computationally intensive modules were written in assembly language to optimize performance.

From 1986 onwards the lattice QCD effort has been led by Rajan Gupta. He and his collaborators have performed a number of novel calculations of the hadron spectrum and weak matrix elements. They explored the glueball spectrum and wavefunctions, developed the methodology for elucidating the flavor symmetry breaking in staggered fermions, pioneered Monte Carlo Renormalization Group methods to develop improved versions of Lattice QCD, performed the first comprehensive set of simulations with two flavors of quarks, and carried out the first systematic analysis of the baryon spectrum.

Similar calculations have been done by other collaborations around the world. Over time, the numerical techniques and algorithms have been refined, calculations have been done with larger lattices and statistics, and choosing more realistic values of the parameters. Until a few years back, most of these calculations were done in the quenched approximation. With increases in the performance of, and access to, parallel supercomputers, current simulations have started to include the effect of dynamical quarks. Results from a recent state-of-the-art calculation by the MILC collaboration for the mass of the nucleon and the rho meson are shown in Fig. 3. The quantity r_1, which is defined in terms of the force between two static quarks, sets the mass scale. There is only one fundamental scale in QCD, and setting its value using a given physical quantity (which is chosen to be r_1 here) is equivalent to setting the strong coupling constant α_s to its physical value. The figure shows that in the quenched approximation, the nucleon is approximately 15% heavier than in the real world and that the rho meson is about 10% heavier. The improvement on adding three flavors of dynamical quarks is remarkable and suggests that the new generation of calculations will be able to test QCD.

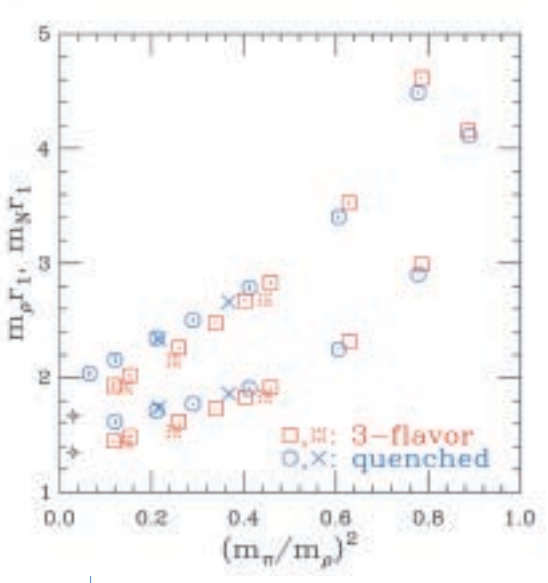


Fig. 3. Comparison of Lattice QCD estimates of the nucleon and rho meson masses in the quenched approximation and with three flavors of dynamical quarks. The left-most points (fancy diamonds) label the physical nucleon and rho masses in units of r_1.

Calculations of the hadron spectrum also provide a way of fixing the values of the fundamental parameters—the masses of quarks and the strong coupling constant. In this inverse problem, the quark masses are given by the set of input values such that all hadrons assume their experimental values. Of this set of unknowns, the greatest uncertainty was in the values of the three light quarks, up, down, and strange. Prior estimates had relied on a combination of phenomenology and analytical calculations and these had very large uncertainties. For example, the strange quark mass quoted by the particle data group in 1996 was 100–300 MeV, and the value $m_s(\overline{MS}, 2\ GeV) \approx 130\ (20)$ MeV was commonly used in phenomenology. Given m_s, the values of m_u and m_d are calculable with good precision using an analytical tool called chiral perturbation theory. In 1996, we analyzed the world lattice data and concluded that the light quark masses were significantly smaller, $m_s(\overline{MS}, 2\ GeV) < 100$ MeV. The reason previous lattice calculations had not been able to establish this result was because of systematic uncertainties—different lattice formulations gave widely different answers. We resolved these uncertainties and showed that consistent results are obtained once a proper extrapolation to the continuum limit is made. The current state-of-the-art result with two flavors of dynamical quarks by the Japanese collaboration CP-PACS is $m_s(\overline{MS}, 2\ GeV) = 89(7)$ MeV.

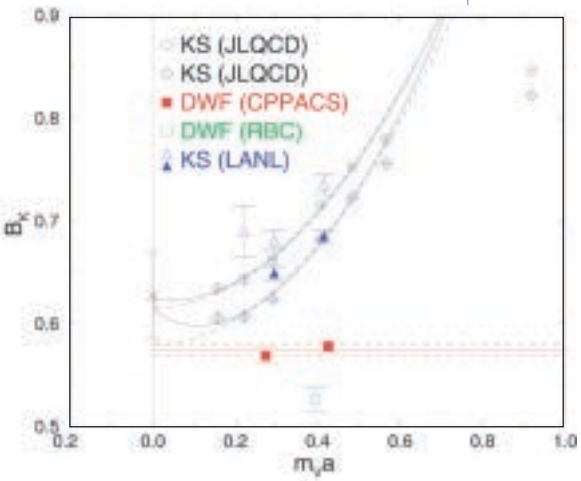


Fig. 4. Results of lattice QCD calculations of the kaon bag parameter B_K.

The final set of calculations that we would like to highlight are of the kaon bag parameter B_K, which measures the QCD corrections to the violation of charge conjugation and parity symmetry in the weak decays of kaons. Gupta and his collaborators have been the leaders in this area—from being one of the first to propose such calculations, to developing the technology using the staggered formulation of lattice fermions that has yielded the most reliable results so far. Their 1998 result $B_K(\overline{MS}, 2\ GeV) = 0.62(2)$ is consistent with the best estimate to date, $B_K(\overline{MS}, 2\ GeV) = 0.63(4)$. This result by the Japanese collaboration JLQCD is exported to the rest of the particle physics community and was obtained using the methods developed by Gupta, Kilcup, Patel, and Sharpe. By simulating the theory at many more values of the lattice spacing, the JLQCD collaboration was able to include a reliable estimate of systematic errors that was missing in the earlier calculation. A summary of the current results is shown in Fig. 4. The major uncertainty in this estimate is due to the quenched approximation, and our current work is aimed at eliminating it.

Over the years the lattice QCD effort in T-Division has attracted some outstanding computational physicists to Los Alamos. These include Clive Baillie, David Daniel, Jeffrey Grandy, Pablo Tamayo, and Weonjong Lee. The group has also maintained very strong ties with external collaborators, Gregory Kilcup (The Ohio State University), Apoorva Patel (Indian Institute of Science), and Stephen Sharpe (University of Washington). Together, they have worked very closely with computer scientists at Los Alamos, notably Jerry DeLapp, Ralph Brickner, Stephen Pope, and Curt Canada, to develop emerging parallel supercomputers (Floating Point Systems T-200 series, Connection Machine 2 and 5, and Silicon Graphics Origin 2000) into reliable computational tools.

Advances in the simulations of lattice QCD are intertwined with the development of parallel- and high-performance computing. One of the earliest parallel computing efforts, the cosmic cube project at Caltech, was driven by physicists, including Gupta, interested in solving lattice QCD. The lattice QCD group at Los Alamos, in collaboration with Brickner, were among the first users of the Connection Machines, CM2, and CM5. They developed a suite of highly optimized codes that stress every component of the machines and paved the way for efficient use of these machines by other users.

Because the collaboration developed the entire QCD code in-house, which is highly modular, they have complete control over all aspects of the computation, and the results at each step are predictable and quantifiable. As a result their code became the most reliable diagnostic for both the hardware and the software. In 1997 when the CM5 was replaced by Silicon Graphics Origin 2000 computer, this code was once again taken up by the manufacturer and used as a core component of their diagnostics. Even today, the QCD code continues to be used as a diagnostic and performance monitor on the state-of-the-art terascale computers at Los Alamos.

Investigation in Physics Beyond the Standard Model: A Review

by Terrance J. Goldman, Peter Herczeg, and John L. Terning

Antoniadis
Bhattacharya
Csáki
Dimopoulos
Engel
Erlich
Friar
Gelb
Gell-Mann
Goldhaber
Goldman
Graesser
Grojean
Gupta
Hall
Haxton
Hayes
Henley
Herczeg
Hoffman
Hollowood
Kaloper
Kolb
Kostelecky
Kribs
McKellar
Mohapatra
Mottola
Nieto
Nussinov
Oka
Raby
Ramond
Randall
Rosen
Ross
Senjanovic
Shirman
Slansky
Stephenson
Strottman
Terning
Towner
Wilczek
Yanagida

A major task of experimental physics has been to search for manifestations of new physics. The experimental efforts are aided by some very specific predictions of the Standard Model. These efforts concern the properties of neutrinos, aspects of CP-violation, proton stability, and some other phenomena. There are also implications for cosmology. The theoretical efforts have focused on model building, the study of the signatures of the associated new physics, and on providing precise predictions for observables. Physicists in the Theoretical Division have made important contributions to this field. The purpose of this article is to give a brief summary of some of this work.

Introduction

There are four fundamental forces known to date: electromagnetic, weak, strong, and gravity. The electromagnetic force is responsible, for example, for the existence of atoms, manifestations of the weak force include some types of radioactive decays, and the strong force is the one that binds neutrons and protons into nuclei. Finally, gravity, the weakest of all, produces an attraction between all forms of matter.

Experimental and theoretical efforts in particle physics culminated in the early 1970s in the formulation of the Standard Model (SM) of the fundamental forces. The SM contains the electroweak theory—a unified description of the electromagnetic and weak force, and quantum chromodynamics (QCD)—a theory of the strong force. For gravity there is as yet no established quantum theory.

The SM incorporates all the known particles that (to our present knowledge) are elementary (i. e., not composed of some more fundamental building blocks). One class of these is the leptons and the quarks. The leptons are spin-1/2 particles that do not interact by the strong force. They come in three "families," each family containing an electrically charged lepton (the electron, muon, and the tau, respectively) and a neutral lepton (the electron neutrino, muon neutrino, and the tau neutrino). The quarks, which also have spin-1/2, can interact through all the fundamental forces. Like the leptons, they also form three families, each containing two types of quarks. Unlike the leptons, each quark comes in triplicate, with the members distinguished by what is called "color." In addition to the leptons and the quarks, and their antiparticles, the SM contains spin-one particles, called gauge bosons, which mediate the fundamental forces. These are the photon (the carrier

of the electromagnetic force), the positively and the negatively charged W, and the electrically neutral Z (which mediate the weak force), and eight gluons (which are responsible for the strong force). The SM also contains a spin-zero particle (the Higgs boson) needed to generate the masses of the other particles. This particle has not been observed so far.

The SM has been extremely successful but has theoretical problems that lead to the conclusion that the model is incomplete. These problems include the mystery of why the mass scale characterizing the scale of gravity is almost a billion-billion times larger than the corresponding scale of the weak interactions. In the SM quantum corrections tend to drive this ratio toward one. In addition there are a large number of undetermined parameters in the model. Thus the existence of physics beyond the SM is widely expected. In fact, we have already, in the form of neutrino oscillations (see below), the first strong experimental evidence that some extension of the SM is required. To address the shortcomings of the SM, various extensions of the SM have been proposed. These invariably contain new particles and new forces. A major task of experimental physics has been to search for the manifestations of the new physics. The experimental efforts are aided by some very specific predictions of the SM. These efforts concern (see below) the properties of neutrinos, aspects of CP-violation, proton stability, and some other phenomena. There are also implications for cosmology. The theoretical efforts have focused on model building, study of the signatures of the associated new physics, and on providing precise predictions for observables. Physicists in T-Division have made important contributions to this field. The purpose of this article is to give a brief summary of some of this work.

Neutrino Physics

The neutrinos in the SM have zero rest mass. As already mentioned, three species of neutrinos are known: the electron neutrino, the muon neutrino, and the tau neutrino. The neutrinos are related to their charged lepton partners through the W-bosons: the W can decay to pairs consisting of a charged lepton and its associated neutrino. Masslessness of the known neutrinos would guarantee that the neutrinos cannot oscillate: a massless neutrino species has zero probability of transforming into a different neutrino.

Recently strong evidence was established from observations of solar and atmospheric neutrinos, that the neutrinos do oscillate and, consequently, that they have a non-zero rest mass. Evidence for neutrino oscillations comes also from the results of the Liquid Scintillator Neutrino Detector (LSND) experiment at the Los Alamos Neutron Science Center (LANSCE) accelerator. An independent experiment (MiniBooNE), which will probe the LSND results, is in progress at Fermilab.

If the neutrinos have mass, a puzzling question is why their masses are so much lighter than the masses of their charged lepton partners. For example, the experimental upper limit on the mass of the electron neutrino is about hundred thousand times smaller than the mass of the electron. In 1979 Murray Gell-Mann, Pierre Ramond, and Richard Slansky* [1] (and independently Tsutomu Yanagida, as well as Rabindra Mohapatra and Goran Senjanovic) [2] observed that in some theories where there is a mass scale much higher than the mass scale associated with the electroweak interaction (about 300 GeV), as occurs, for example, in Grand Unified Theories (theories that unify the electroweak and the strong interaction), it is possible to formulate models of the neutrino masses where the neutrino mass is proportional to the ratio of the square of the mass of the associated charged lepton to the new mass scale. Gerard Stephenson (then visiting from the University of Maryland and later a Laboratory staff member) also participated in the early stages of this work. [3] This idea (the so-called see-saw model) is widely regarded as the most plausible explanation of the smallness of neutrino masses.

Physicists in T-Division also made other important contributions to the physics of neutrinos: Peter Rosen* and Jim Gelb* to the physics of solar neutrino oscillations; [4] Stephenson, Terry Goldman,* and Bruce McKellar to the interpretation of experiments measuring the mass of the electron neutrino through studies of the electron end-point spectrum in nuclear beta decay; [5] Peter Herczeg to the question of the possibility of interpretations of the LSND results through new physics other than neutrino oscillations; [6] and Anna Hayes* and Ian Towner through improved calculations of neutrino-nucleus scattering cross-sections. [7] The latter are important for the interpretation of some of the neutrino oscillation experiments.

Baryon Number Violation

In the SM, baryon number (one third of the number of quarks minus antiquarks) is conserved in all processes at experimentally accessible temperatures. In Grand Unified Theories this is generally not the case. Grand Unified Theories have been popular since their inception, but variants with supersymmetry (a symmetry between fermions and bosons) have recently received a boost from the observation that the extrapolation of gauge boson interaction strengths from the weak scale up to the unification scale indicates that these forces do seem to unify in supersymmetric theories. Since supersymmetric theories can also stabilize the ratio of gravitational and weak scales, many theorists have taken this as indirect evidence that supersymmetric Grand Unified Theories are the most plausible extension of the SM. As mentioned above, an important implication of Grand Unified Theories is that the proton can decay.

* In the text an asterisk distinguishes the names of the authors who were members of T-Division at the time referenced.

In T-Division, Gell-Mann, Ramond, and Slansky* made a comprehensive study of the ways that the proton can be stabilized in Grand Unified Theories [8]. More recently, the effects of discretized extra dimensions on proton stability were studied by Oppenheimer Fellow Csaba Csáki,* Graham Kribs, and John Terning (now a staff member in T-Division). [9] Other work in this area included improved calculations of the proton lifetime by Goldman,* Doug Ross, Alfred Goldhaber, and Shmuel Nussinov, [10] and investigations of proton decay in supersymmetric models by Savas Dimopoulos, Stuart Raby,* and Frank Wilczek. [11]

Lepton Number Violation and Lepton Family Number Violation

In the SM, lepton number (the net number of leptons minus antileptons) is conserved in all processes at low temperatures, while Grand Unified Theories, for example, allow for lepton number conservation to be violated. An example of a process that violates lepton number conservation is neutrinoless double beta decay (the process neutron + neutron -> proton + proton + two electrons in nuclei). Searches for neutrinoless double beta decay provide information on neutrino masses that is complementary to that from neutrino oscillation and beta decay measurements. Moreover, neutrinoless double beta decay is the most sensitive process for determining whether the neutrinos are Dirac or Majorana particles. In the latter case the neutrinos would be identical to their antiparticles.

Wick Haxton,* Stephenson,* and Dan Strottman* carried out new calculations of neutrinoless double beta decay rates, [12] in which the uncertainties in the treatment of nuclear structure were reduced substantially. This is an essential requirement for extracting reliable information on the neutrino parameters. Further work in this area included an analysis by Goldman,* Rocky Kolb,* and Stephenson* of some implications of spontaneous lepton number violation. [13]

In addition to lepton number conservation, in the SM the individual lepton family numbers (LFN), i. e., the number of leptons from a particular lepton family minus the number of antileptons from that family, are also conserved. Many extensions of the SM (including supersymmetry) allow for LFN conservation to be violated. Examples of LFN nonconserving processes are the decay of the positive muon into a positron and a photon, and the decay of a neutral K-meson into an electron and a positive muon. The recently observed mixing of the neutrinos provides the first evidence that the conservation of the LFN is not exact.

Lawrence Hall, Alan Kostelecky, and Raby* contributed to the understanding of LFN violation in supersymmetric models, [14] and Goldman,* Stephenson,* Peter Herczeg,* and Takamitsu Oka* examined the phenomenology of LFN violating processes in a variety of extensions of the SM. [15, 16, 17, 18]

CP-Violation

Before 1964 it was believed that the laws of nature would be identical in a hypothetical world that is a mirror image of ours and in which the particles are replaced by their antiparticles. In a more technical language: physics was thought to be invariant under the product of C (charge conjugation, which takes a particle into its antiparticle) and P (parity, which changes the sign of the space coordinates). In 1964 an experiment investigating the decays of the neutral K-mesons (which contain a strange quark) demonstrated that physics is not CP invariant. More recently, CP-violation was also found in the neutral B-mesons (mesons containing a bottom quark), and an additional CP-violating effect was discovered in the decays of the neutral K-mesons. Although the SM incorporates CP-violation, and the observed effects are consistent with the predictions of the SM, the possibility that additional CP-violating interactions are involved is not ruled out. In fact, the most widely studied extension of the SM, supersymmetry, predicts a whole host of new CP-violating effects. Other extensions of the SM also predict additional CP-violation. Thus studies of CP-violation are important tools for probing physics beyond the SM. The most suitable observables for probing new CP-violating interactions are those, for which the contributions from the weak interaction of the SM are suppressed. Examples are CP-violating effects in neutron and nuclear beta decay, CP-violating nucleon-nucleon interactions, and the electric dipole moments of the electron and the neutron. For all these the weak interaction contributes only in second order.

In order to interpret many CP-violation experiments, precise calculations of the effects of the strong force are needed. T-Division has been a leader in this field (see the accompanying article in this chapter on lattice QCD by Rajan Gupta* and Tanmoy Bhattacharya*). Further work in T-Division included a study of CP-violation in beta decay in extensions of the SM by Herczeg,* [19] which provided the theoretical motivation for experiments searching for the so-called D-correlation (a correlation in the decay probability between neutron or nuclear spin, and the momenta of the electron and the neutrino), investigations of nuclear forces that violate simultaneously parity and CP conservation by Wick Haxton* and E. M. Henley [20] and Herczeg,* [21] and studies pertinent to atomic electric dipole moments by Jon Engel, Jim Friar,* and Hayes,* [22] and by Herczeg.* [23] A mystery of the SM is why CP-violation in the strong interaction is so weak. This suppression can be explained by the existence of a hypothetical particle, the axion. Goldman* and Cyrus Hoffman found a theoretical lower bound on the branching ratio for K-meson decay into a pion and axion, and deduced an experimental upper bound from existing data that was not much larger. [24] Dimopoulos and Raby* demonstrated how a supersymmetric Grand Unified Theory could predict the existence of an axion naturally. [25]

Theories with Extra Dimensions, Gravity and Cosmology

While supersymmetry has played a central role in many extensions of the SM, it is not the only possibility for stabilizing the weak scale relative to the gravitational scale. The most popular alternative is based on the recent revolution in our thinking about the possibility of extra spatial dimensions. Since Einstein forced physicists to think of time as a fourth dimension there have been speculations about a fifth and even higher dimensions. In order to avoid gross violations with experiment and even everyday experience, it was thought that any extra dimensions would have to be tiny and compact (like a circle or a sphere) and that their effects on present-day experiments would be unobservably small. There was the possibility that the ordinary particles that we know are somehow restricted to a membrane (with three spatial dimensions) that is embedded in a higher dimensional spacetime. This possibility was not taken very seriously, since it was not known how to realize this in a quantum field theory in a robust way. However, recent developments in string theory show that such membranes are required for consistency. This new understanding of the feasibility of localizing four-dimensional worlds (like ours) on a membrane in a higher dimensional spacetime has led to a variety of viable models, and even to the possibility of localizing gravity. Unlike older theories of extra dimensions, much of the focus now is on extra dimensions with sizes on the order of one thousandth of the size of the proton or even larger! Thus, there is a potential for discovery at current and soon-to-be-completed colliders and, in some cases, tabletop experiments. In addition, there are tremendous implications for cosmology.

Csáki,* along with Joshua Erlich,* Tim Hollowood,* and current Feynman Fellow Yuri Shirman, studied the universal effects of models where our universe exists on such a membrane. [26] Csáki,* Erlich,* and Terning* have also shown that models which affect the origin of the W and Z gauge boson masses (and hence also the origin of fermion masses) are tightly constrained by current precision electroweak data. [27] Csáki* and Erlich,* and Christophe Grojean found that in certain circumstances gravitational signals could propagate more quickly off the membrane than on it, leading to apparent violations of causality. [28] Csáki,* Michael Graesser, Lisa Randall, and Terning showed how the constraint of reproducing the observed cosmological expansion since the time of nucleosynthesis also greatly constrains these theories. [29]

In an alternate take on particle cosmology, Csáki,* Nemanja Kaloper, and Terning* showed that the observed dimming of distant supernovae may indicate the existence of an ultralight particle (an axion) rather than an accelerating expansion of the universe as is usually inferred. [30] Such an accelerating expansion requires the existence of an incredibly small "cosmological constant" or some heretofore-unknown "dark energy." A possible alternative explanation for a small "cosmological constant" was provided by Ignatios Antoniadis and Emil Mottola* in the form of the hypothesis that gravity becomes scale invariant at large distances. [31] Significant work was also done on how new interactions could modify gravity at long distances (see the accompanying article in this chapter by Michael Nieto*).

As can be seen from the brief summaries above, there has been an enormous range of new ideas that have been generated and probed by T-Division members along with their external collaborators. Also, much of the theoretical work mentioned above provided theoretical support for a number of experiments carried out at Los Alamos.

Beam Dynamics for Next-Generation Accelerators

by Salman Habib

Habib
Ji Qiang
Ryne

Accelerators are crucial to forefront discoveries in high-energy physics and nuclear physics. Indeed, experimental efforts in these fields have been largely defined by advances in accelerator science and technology. More recently, accelerators have become essential tools for structural studies in materials science and biology. Aside from fundamental research, accelerators are essential components in diverse applications, such as medical isotope production, irradiation therapy, ion implantation, beam lithography, transmutation of waste, and hydrodynamic imaging. T-Division remains connected to the national effort by helping to provide the basic scientific underpinnings for next-generation accelerators.

The national investment in accelerator technology is enormous: major facilities include the Tevatron at Fermilab; SLC, PEP-II, and Stanford Synchrotron Radiation Laboratory at the Stanford Linear Accelerator Center (SLAC); Spallation Neutron Source at Oak Ridge National Laboratory; Continuous Electron Beam Accelerator Facility at the Thomas Jefferson National Accelerator Facility; Relativistic Heavy Ion Collider and National Synchrotron Light Source at Brookhaven National Laboratory; Intense Pulsed Neutron Source, Argonne Tandem-Linear Accelerator System, and Argonne Photon Source at Argonne National Laboratory; and Los Alamos Neutron Center/Lujan Center at Los Alamos National Laboratory.

The development of next-generation accelerators requires very significant advances in simulation capability to address issues of performance optimization and cost reduction. Given the expense of a major facility, investment in a predictive design capability is not only prudent, it is fast becoming essential. Moreover, the development of new methods to radically improve the acceleration gradient (to perhaps GeV/m to tens of GeV/m) also requires high-resolution numerical modeling. The importance of developing such a capability for high-current linear accelerators was recognized at the Laboratory by Robert Ryne and me 10 years ago. Initial work carried out at the Laboratory's Advanced Computing Laboratory (ACL) led to a Department of Energy Grand Challenge award in 1997. This award was instrumental in putting together a national program in advanced accelerator modeling encompassing electromagnetics (spearheaded by SLAC) as well as beam dynamics.

The Grand Challenge was followed by one of the first-ever awards in DOE's Scientific Discovery through Advanced Computing Program (SciDAC).

Our work on beam dynamics began with the development of a parallel, particle-in-cell (PIC) code for modeling beam transport in linear accelerators including the effects of space charge (i.e., not only taking into account the external guiding, focusing, and accelerating forces on the particles but also their mutual Coulomb interaction). Moving the state of the art in the field rapidly from handling hundreds of thousands of particles now to billions of particles, this initial effort culminated in the IMPACT (Integrated Map/Particle Accelerator Tracking) code that combines high-order magnetic optics with a PIC treatment of space-charge forces. The use of a split-operator methodology allows the code to retain symplectic time evolution.

IMPACT is a large parallel program that runs on state-of-the-art computing platforms such as the IBM SP-3 at the National Energy Research Scientific Computing Center (NERSC). In fact, it is one of the largest such codes in the world. (Unfortunately, while IMPACT was originally developed at Los Alamos on the Thinking Machines CM-5 and the SGI Origin 2000 cluster at the ACL, there appears to be no future for the code at LANL due to lack of sufficient open computational resources.) Development of IMPACT continues with the addition of higher-order elements, integration with the Lie algebraic code MaryLie, implementation of ever-more sophisticated Poisson solvers (multigrid with adaptive mesh refinement), and exploration of new applications. The present research effort on beam dynamics is focusing on beam-beam interactions, intra-beam collisions, and extension of our previous work on linear accelerators to circular machines. An exciting prospect is the implementation of IMPACT kernels in a new code to study the formation and evolution of structure in the universe. This code will provide the computational backbone for a new collaboration between Los Alamos and the Sloan Digital Sky Survey, the largest astronomical survey implemented so far.

One of our recent innovations came out of the recognition that it is possible to carry out fully self-consistent calculations for the damping and diffusion coefficients that characterize "soft" collisions in a mean-field dominated charged plasma. These damping and diffusion coefficients appear in a Fokker-Planck equation and characterize the effects of dynamical friction and multiple small-angle scattering respectively, the strengths of the two being connected by a generalized fluctuation-dissipation relation that enforces exact energy conservation. The algorithmic structure of our code LANGEVIN3D is based on the PIC method familiar from the solution of Vlasov-Poisson equations. Extension of this method to the collisional case, especially for inhomogeneous

situations, is very demanding numerically. The full simulation is broken up into a set of "super cells" and Monte Carlo integrations using tracer particles are carried out in each super cell to obtain the local transport coefficients. The particles are then stepped forward using both systematic force terms (these include external, space-charge, and damping forces) and stochastic force terms that are the trajectory (Langevin) analog of the diffusive term in the Fokker-Planck equation. Since the transport coefficients vary with position, so must the strength of the stochastic forces in the Langevin equations: the noise appearing in these equations is multiplicative (in contrast to the more usual case of additive noise). An interesting feature of the multiplicative noise Langevin equations for the particle dynamics is that the energy is explicitly conserved. This reflects the essential self-consistent nature of the process, in contrast to the energy fluctuations characteristic of canonical ensembles or those seen in simulations of Brownian motion.

We are now extending our method to the case of multicharged species (important for beam cooling and study of collisional instabilities) as well as making careful checks of the numerics for spatially inhomogeneous distributions. In some situations of physical interest in the dynamics of intense beams, collisional effects can compete with collisionless fluctuations. For most linear accelerators, the collisional terms are small and relaxation effects due to them are subdominant over the time scales of interest. In circular machines, however, the dwell times are very much longer and collisional effects can be significant.

The advanced accelerator simulation effort was an excellent example of a cross-divisional collaboration at Los Alamos. Although the two main researchers at LANSCE (Ji Qiang and Robert Ryne) have now moved to LBNL, T-Division remains connected to the national effort by helping to provide the basic scientific underpinnings of the project.

Ander
Anderson
Chamblin
Csaki
Erlich
Giovanielli
Goldman
Holzscheiter
Hughes
Laing
Lau
Liu
Luther
Nieto
Raby
Shirman
Terning
Turyshev

Tests of Gravity

by Michael Martin Nieto

Although the concept of testing gravity on antimatter was unconventional in the 1980s, the idea amounted to another way to look at modern theories modifying gravity. These theoretical ideas also helped stimulate experimental tests of gravity, many of them done by Los Alamos, on lab scales (mm to cm), geophysical scales (km), and even deep space scales (many times the distance from the Earth to the Sun). The work on antimatter has progressed to the point where the ATHENA antihydrogen production experiment has recently made over 50,000 cold antihydrogen atoms—one step closer to the ultimate goal.

It all started in 1981 when, walking down the hallway, I heard Stu Raby ask out of his office in Theoretical Division, "How well do we know that Newtonian gravity is correct in the solar system?" Stu was interested because, at the time, there were ideas floating about that a relatively heavy axion particle might exist. It could mimic a change in gravity on large-length scales. I answered, what was then true, "Actually, not very well at all." As I walked away, for some reason I thought, "And when it comes to antimatter we know virtually nothing!"

There were then standard arguments against "antigravity," which are not as tight as often thought. Further, modern particle theories were producing predictions that, at some length scale, as a matter of principle gravity must be different for antiparticles, even if only infinitesimally. [1] The more I thought about it, the more interested I became. Terry Goldman joined me in the work. Years later, Terry told me he had originally joined me because he thought this might be totally crazy and he wanted to keep his friend from going off of the deep end. At the time the idea that one should test gravity on antimatter was certainly quite far out.

We proposed that an experiment to test gravity on antiprotons be done at the Low-Energy Antiproton Ring (LEAR) at CERN. Richard Hughes, a postdoctoral student, later joined us and funding was obtained to start the experiment, eventually under the leadership of Michael Holzscheiter of Physics (P) Division.

Simultaneously, Principle of Equivalence tests had become of interest because of the proposed "5th (hypercharge) force" of Ephraim Fischbach. The connection was that, "hypercharge force" or not, this idea amounted to another way to look at modern theories trying to modify gravity. In particular, geophysical experiments were of use to find if there were forces with length scales on the order of kilometers.

At Los Alamos, Mark Ander of the former Earth and Space Sciences Division (now called Energy and Sustainable Systems Division) wanted to perform a test through 1.6 km of ice down a borehole at the DYE-3 site. DYE-3 was an Early Warning DEW-line station, east of Søndrestrømfjord on top of the Greenland icecap. (See Fig. 1.) The experiment came off and participating in it was a great adventure. The measurements were a gravity field survey and topographic study on the surface, a study of the underground topography, and of gravity down a borehole built for glacier studies. (See Fig. 2.)

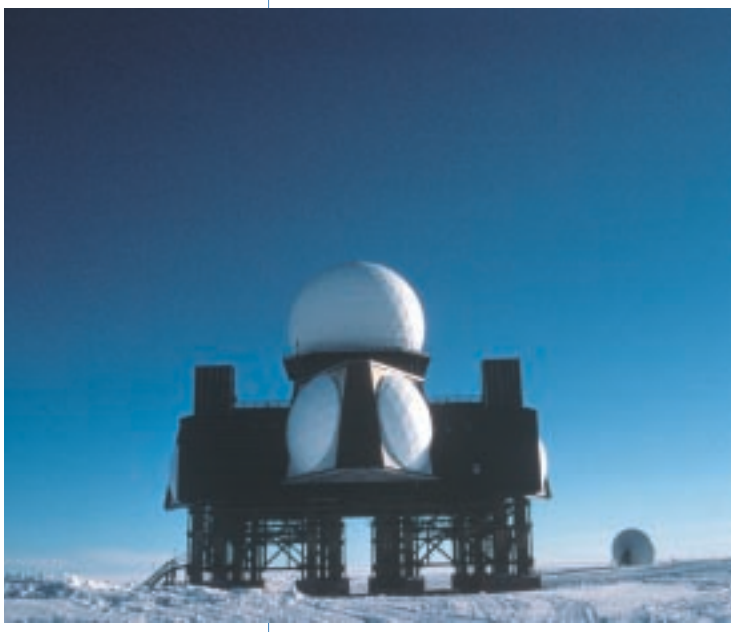


Fig. 1. The DYE-3 station on the Greenland icecap. It is situated over more than 2 km (1-1/4 miles) of ice. To appreciate the scale, note the staircase in the lower left side of the station.

This first experiment was inconclusive because of too-large uncertainties in the type of rock under the ice. But it made us realize something very important. We were testing Gauss' Law with gravity on a plane and a line down the center. What is preferable is to measure gravity on many planes and/or many vertical lines so one has data from a 3-dimensional region. That was next. Joining the Scripps Institute of Oceanography on their ship, New Horizon, we went off the California coast where the water has a depth of over 5 km. A surface gravity survey was made, the topography of the ocean floor determined, a gravity meter was lowered to numerous places on the ocean floor (See Fig. 3.), underwater gravity planes were measured on a navy submarine, and vertical profiles were made in a submersible. The end result in 1991 ruled out a deviation greater than 2 parts in a thousand over scales of a few kilometers.

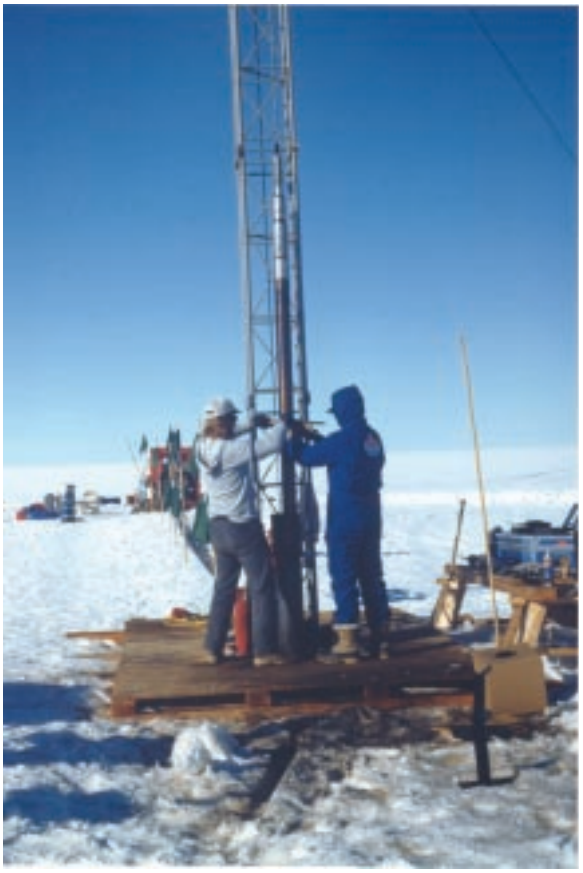


Fig. 2. Casey Rohn of the University of Nebraska and James Wirtz of Amoco carefully lift the borehole gravity meter from the ice. Gravity measurements were made down to 1683 meters below the surface. In the background is the cabin housing the winch.

These types of geophysical experiments, laboratory experiments such as those performed by Gabe Luther of P-Division, and

ranging space experiments, have now ruled out deviations from Newtonian gravity greater than a fraction of a percent on scales from about a millimeter to many astronomical units (AU).

Returning to antimatter, especially with the aid of Damon Giovanielli of P-Division, equipment reached CERN, and the largest number of antiprotons ever made and controlled in a trap was accomplished (greater than a million). Unfortunately, funding and time constraints at CERN led to the decommissioning of the LEAR accelerator.

However, a new storage ring (the AD) was commissioned and the original Los Alamos equipment was reconfigured in a new experiment (ATHENA), with a first goal of making cold antihydrogen. The ATHENA trap, an upgrade of our original PS-200 antiproton "catching trap," has detected about 130 reconstructed, low-energy, antihydrogen events. [2] (See Fig. 4 for an example antiproton catching trap.) Given the low detection efficiency, this signal implies that on the order of 50,000 cold antihydrogen atoms were created. The next step will be to confine these antiatoms in a (perhaps AC) magnetic trap, and then to eventually measure gravity on this easy-to-calibrate neutral system.

Fig. 3. The submersible gravity meter is just being taken aboard the Scripps Institution of Oceanography research vessel New Horizon. In a given run, the meter is lowered to the sea bottom, 5 km down, at a known flat region. Measurements are made, the meter is lifted a few hundred meters, the ship then tows its "5-km anchor" for a few km, and finally the meter is lowered again.

In 1994, during the period of collecting antiprotons at CERN I was asked to give a talk at the 1994 Low-Energy Antiproton workshop. Observing that over most of the universe (everything outside the solar system) we do not know gravity precisely, I decided to find out how close in we do know gravity well. (Note that the standard "solution" to the problem of gravitational effects in the cosmos not agreeing with the observed matter is to assume that the majority of the matter of the universe is "dark matter" or in the form of "dark energy" and, hence, that is the reason it has not been observed.)

Coming to the conclusion that our best bet to find deviance was near the outer edge of the solar system, I eventually was led to John Anderson of the Jet Propulsion Laboratory (JPL). He informed me that, because they are spin-stabilized, the Pioneer spacecraft are the most precisely tracked probes. (See Fig. 5.) Then he added something that had been nagging them for a number of years, "By the way, the biggest systematic in our acceleration residuals is a bias of 8×10^{-13} km/s^2 directed toward the Sun." I almost fell off of my chair because, although this is an anomalous acceleration of only 8 Å/s^2, over years this can translate into a huge distance.

Fig. 4. A photograph of the Los Alamos antiproton "catching trap" on the floor at CERN. Nick King of the Laboratory's P-Division is shown by the trap. This trap was upgraded to become the ATHENA antihydrogen production trap.

Joined by other colleagues at JPL (Eunice Lau and Slava Turyshev) and at The Aerospace Corporation (Philip Laing and Tony Liu), we began a detailed study of the signal and all possible systematics, external to and internal to the spacecraft, that could produce the effect. We found none and in 2002 published a 50-page manuscript in *Physical Review D* saying that between about 20 AU and 70 AU there is an anomalous Doppler signal that can be interpreted as an anomalous acceleration towards the Sun of $(8.73 +/- 1.34) \times 10^{-8}$ cm/s^2.

Even though one has to concede that ultimately the signal is probably some as-yet-not understood systematic, the result has elicited enough interest, both experimental and theoretical, that we hope it can be tested by another mission in the not-too-distant future.

Coming full circle, I want to conclude with the observation that recent tests of gravity on small, submillimeter scales have been inspired by theoretical studies of gravity in extra dimensions. Recent theoretical work has pointed out that these dimensions possibly could appear in our world on these scales. Among those who have been importantly active and productive in this area, are several people in T-Division: Andrew Chamblin, Csaba Csaki, Josh Erlich, Yuri Shirman, and John Terning. [3]

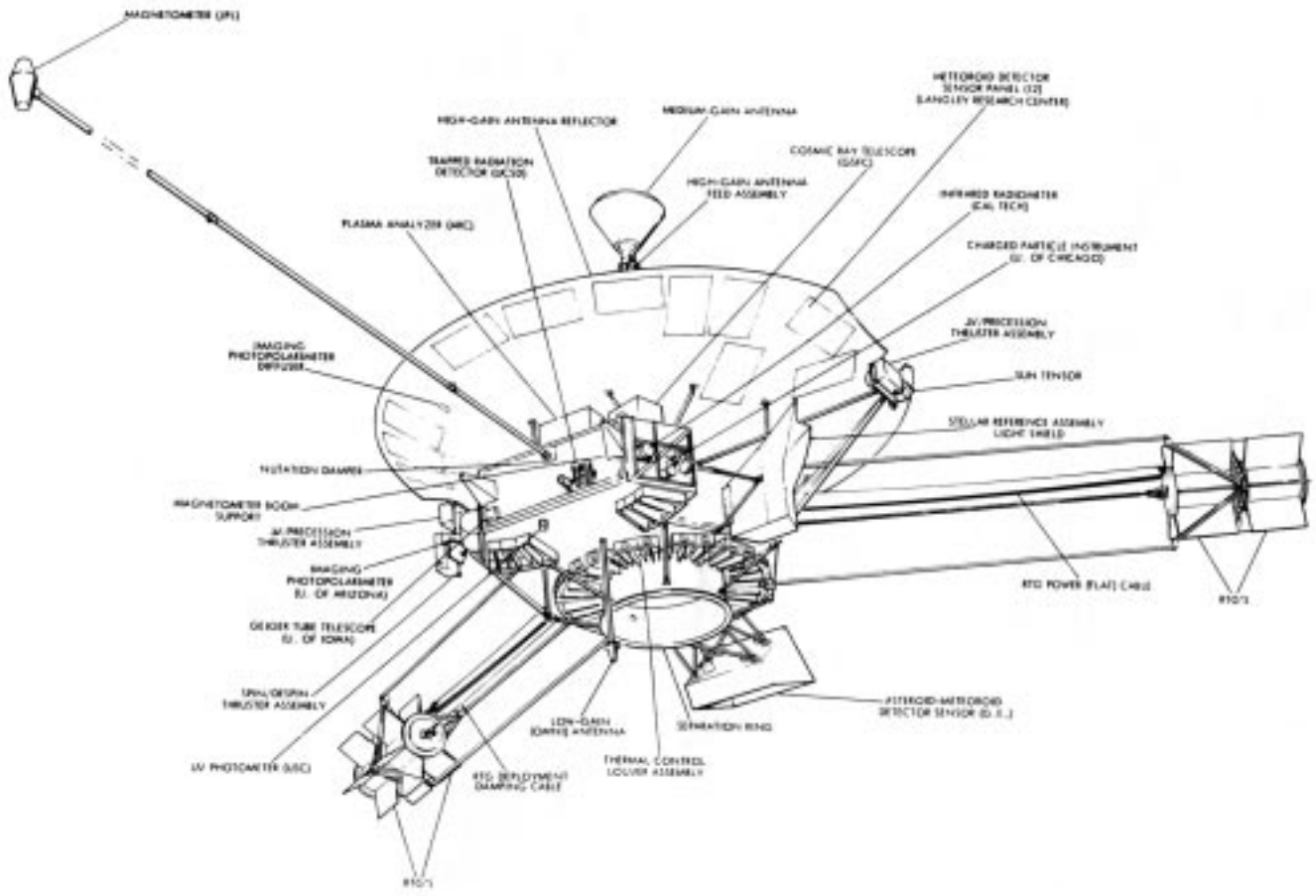


Fig. 5. A drawing of the Pioneer spacecraft.

Catching the Free Neutrino

by Harris L. Mayer

The team of Frederick Reines and Clyde Cowan at Los Alamos used a strange-looking detector to detect the elusive free neutrino for the first time in 1956, thus giving reality to the Pauli suggestion that such a particle must exist to conserve both the energy and momentum in beta decay. The seemingly impossible experiment using that detector earned Frederick Reines the Nobel prize in physics in 1995. Clyde Cowan had died and was therefore ineligible to share in the prize. The reality of the existence of the free neutrino and of its detectability legitimized the whole field of neutrino experiments. The neutrino is now included as the doublet partner of the electron, its rightful role in the classification of elementary particles.

How the Quest Started

In the lobby of the Administration Building at the Los Alamos National Laboratory is a strange looking piece of equipment with the words Herr Auge scrawled with magic marker on it. (See Fig. 1.) It is the detector the team of Reines and Cowan at Los Alamos used to detect the elusive free neutrino for the first time, thus giving reality to the Pauli's 1930's suggestion that such a particle must exist to conserve both the energy and momentum in beta decay. The seemingly impossible experiment using that detector earned Frederick Reines the Nobel prize in physics. Although the work was completed in 1956, the prize came belatedly in 1995. Clyde Cowan had died and was therefore ineligible to share in the prize. Fred Reines was frail. The prize was a retrospective on their accomplishments, rather than a prospective on their promise. The detector that had sparkled at the beginning, is now a relic. How did this somewhat ugly but sentimentally beautiful piece of equipment get to its resting place?

Fig. 1. Fred Reines tinkering with the "Herr Auge." He won the 1995 Nobel prize for his leadership of the Los Alamos team that first detected the free neutrino.

The direct answer is that at the conclusion of the experiment it was deposited as surplus in the Laboratory's salvage yard from where it was picked up by Ed Grothus of the "Black Hole" and forgotten. After many years Ed, in a historically sensitive moment, donated the detector to the Laboratory.

But the real answer goes back to the day in 1948 when Fred Reines came into my office at the Laboratory and suggested that we should make up a handbook for the upcoming Sandstone nuclear test operation in the Pacific. Thereafter Fred became actively interested in the entire testing regime. In his usual insightful way, Reines made significant contributions to the scientific understanding of the principles of weapon design and of the weapon effects revealed by the test operations. Thus he was initiated into the realm of "Big Physics," excellent preparation for his future career in neutrino experimentation.

Fred Reines received his Ph.D. in nuclear physics under Richard Present at New York University in 1944. Soon after he joined the Los Alamos Laboratory in the Theoretical Division. In contrast to his tutelage under a quiet, competent but unassuming thesis advisor, Fred now found himself in a position to learn from his group leader Dick Feynman, young, outwardly antic, but keenly capable of illuminating the technical path to follow to solve even the most difficult problem. The division was headed by the hearty Hans Bethe of almost universal knowledge and capability. What a stimulating change! Fred's mind expanded in the new environment. During the war he worked on many theoretical aspects of nuclear weapon design. It is unclear what specific influence Feynman had on Fred's development as a mature physicist. However, nobody working closely with Feynman, as his group members certainly did during the war, could avoid being influenced by that magical free-ranging spirit.

After the war, Fred, freed from the focused pressures of the specific design of Fat Man and Little Boy, let his imagination range over the whole field of nuclear energy applications. Almost alone among the Lab scientists, Reines opened up the field of nuclear weapon effects—what happens to the tremendous energy released after it leaves the outer boundary of the exploding bomb. He pioneered the application of scaling methods to blast and thermal radiation, as well as to the γ-ray and neutron emissions. That endeared him to the military. He also had the vision, typical of much of his later work, to realize that the bomb would generate a tremendous lightning-like electromagnetic pulse, even if only a small fraction of its energy, say 10^{-4} or 10^{-5}, was emitted in that form. These ideas led him naturally into participation in the nuclear weapons testing program. On leave from the T-Division he became the leader of the Pogo staff in the Field Testing (J) Division, which recommended and then coordinated the execution of the diagnostic tests and the weapons effects tests conducted at the Pacific test range. Fred had become a "hands off" experimentalist, deciding what could and what should be done, and also what would not be done in the test program.

Several years later, Reines, feeling that he was losing his edge as a scientist, resigned from J-Division and returned to the T-Division. He felt uneasy with his career—the excitement of the participation in the test operations were like an addiction. He went to Carson Mark, the fatherly leader of the division, and asked if he could be relieved of programmatic responsibilities "to sit and think," to put his scientific life back on track. With wisdom and grace characteristic of his handling of the division, Carson said that Fred had already overpaid his dues to the Laboratory and was entitled to the chance to make some contribution to pure science.

Fred literally sat and thought for a year in his office at T-Division. Much later he confessed to me that this was the hardest thing that he ever did in his life. Finally he decided that he would detect the neutrino, primarily because everyone thought that it was impossible. Fred felt that there must be an obviousness to the realities of the fundamental particles of nature. They had to be more than an abstraction made up to balance the equations to conserve the sacred quantities of physics. To be accepted as real, not a convenience device, the free neutrino must be detected. And Fred had an advantage that the physicists of academe did not—he knew about fission bombs. He had access to their tests; he felt that the bomb gave him real possibilities for neutrino detection, because it was a fantastic pulsed source of these particles.

Fred went with his idea for detection of the neutrinos from a fission bomb explosion to the "wise old men" of the Los Alamos Lab—Fermi, Bethe, Teller. They told him that the neutrino would not be detected, that its cross section was too small to be measured. The neutrino mean free path was greater than the diameter of the Earth—the neutrinos at that time were even thought to proceed unimpeded from the reactive core of the sun to its surface. The sages were both right and wrong. They were right—the neutrinos from the bomb could not be detected but primarily because of the fantastic background of neutrons and γ-rays accompanying the explosion. But they were wrong because the free neutrino could be detected. Fred would not give up.

Fred had the soaring imagination of the theoretical physicist untrammeled by deep knowledge of the intricacies of that craft. In his experimental approach to the free neutrino search he also was imaginative, and he paid careful attention to experimental details. However he was still a novitiate, undeterred by prior hard experience with the obstinacies of inanimate equipment. He was thus the ideal person to attempt the impossible experiment. He didn't know any better. He did know that he was embarked on big science. He would need a group that had to be managed; there would be scientific sponsoring agencies to assuage. But mere bigness of a project would not deter him. He had experience with the large-scale efforts in the nuclear test program that more than compensated for his naiveté in scientific experimentation.

Fred Reines had felt the seductive attraction of the thermonuclear bomb. Under its spell, he had seen the armada of Joint Task Force Eight sail into a Pacific atoll to test a thermonuclear device. He was not only to overcome that seduction of weapons program; he was to make its lessons the means of jump-starting his scientific career. Reines used that knowledge wisely in a lifetime of research in understanding the neutrino.

The Initial Experiment: Hanford

In the initial Reines/Cowan neutrino detection experiment, the source was one of the powerful nuclear reactors at Hanford Site. The products of the nuclear chain reaction in the reactors emitted neutrinos in their beta decay. The initial nuclei, such as uranium-235 at the upper end of the periodic table, had an excess of neutrons over protons compared to the stable fission daughters in the middle of the periodic table. After fission, the basic reaction is the transformation of a neutron into a proton by the process of beta decay, that is

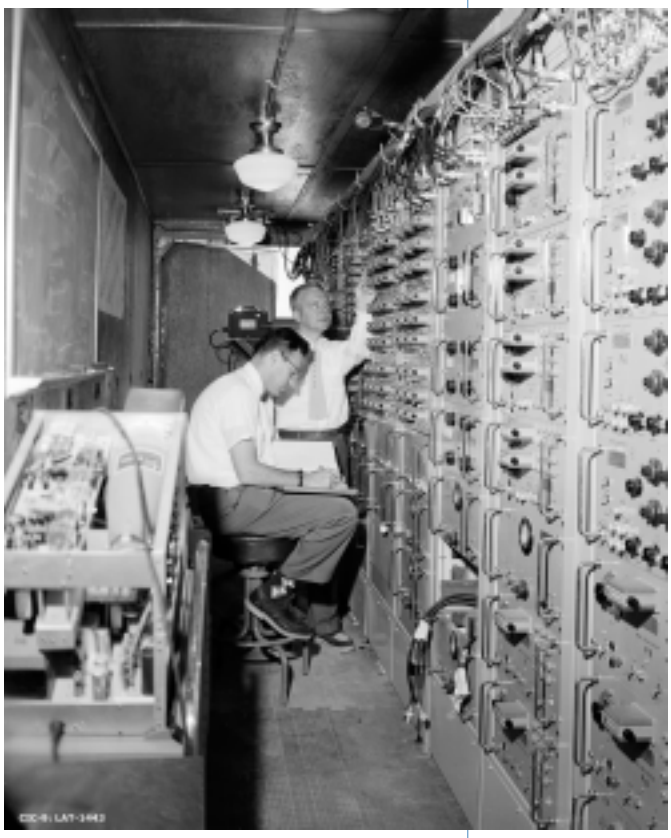


Fig. 2. Cowan and Reines calibrating the electronics.

$$n \longrightarrow p + e + \bar{\nu}_e \quad .$$

Here the negatively charged electron "e," which compensates for the charge of the proton, is paired with an electron antineutrino $\bar{\nu}_e$. Of course it was not the decay of the free neutrons that produced the major signal, but the change of neutrons into protons within the daughter nuclei. The energy available in this transformation varied with the particular daughter, but was of the order of 1 to a few MeV. Divided between the electron and the neutrino in the ratio 1:2, the neutrinos had a continuous spectrum with the average energy around 3 MeV.

The detection process made use of the inverse beta decay reaction

$$\bar{\nu}_e + p \longrightarrow e^+ + n \quad .$$

Here it is the positive-charge positron e^+ that carries away the charge of the proton. The defining concept which made the experiment feasible was the use of both products, the positron and the neutron, in delayed coincidence. This produced an unmistakable signal of the neutrino capture that stood out in the vast background produced by neutrons and gamma rays from the active reactor.

Fortunately for Reines and Cowan, the essential components for this type of detector were available from technology in use in Los Alamos in connection with nuclear weapons tests. It was known that ionizing radiation produced visible light in some transparent liquids having the

benzene ring in their structure. The response was very fast due to the radiative transitions of the specific molecules excited by the electrons produced by the ionizing incident particles. At this time large-area fast-response photomultiplier tubes to detect this light were just being produced. The neutrino group, Reines, Cowan, and associates, built a detector from a large tank of scintillator liquid with an array of photomultiplier tubes sitting like patterned warts on the tank skin. Considerable trial and error, empowered by some dark art, was invoked in preparing the actual liquids used in this and future detectors. By 1953 the initial Hanford experiment was ready.

The characteristic signal of the neutrino was obtained as follows. The positron from neutrino capture almost immediately found an electron in the scintillator liquid and was annihilated. The two resulting 0.511 MeV gamma rays produced a visible light pulse in the detector. Meanwhile, the neutron wandered around, slowing down in the hydrogen-rich liquid. After a few microseconds it thermalized and was captured by cadmium atoms in cadmium chloride in the liquid. The gamma rays emitted after this neutron capture excited a second light pulse in the scintillator liquid. These two pulses recorded by photomultipliers were fed into delayed coincidence discriminating electronics. An acceptable signal had amplitude and time delay easily distinguishable from the very many randomly occurring pulses due to background radiation.

The detector was placed as close as possible to the Hanford reactor but heavily shielded to reduce the gamma rays from the source. The shielding of course had essentially no effect on the neutrinos. For various operational reasons the reactor was shut down at intervals, a circumstance not under control of the experimentors. This was fortunate, Reines and Cowan could compare the number of delayed coincidence counts (putative neutrino detections) when the reactor was running and when it was off. They found an encouraging difference but not enough to establish neutrino captures clearly.

The Second Experiment: Savannah River

The Neutrino is Real

The difficulty with the Hanford experiment was a undesirable capture signal due to cosmic rays which produced a background 10 times the genuine neutrino signal. Of course by taking a long series of measurements with the reactor on and with it off, the real signal could be pulled out of the counterfeit background. Better, however, to find an additional characteristic of the cosmic ray to act as a veto signal. The Reines/Cowan group retreated from Hanford to design a new detector that had this veto power.

The basic idea was simple. If two neutrino detectors were placed one atop the other, the cosmic ray would pass through both of them, giving a counterfeit signal in each. A single neutrino could not do this. There were plenty of neutrinos simultaneously incident on the detectors, but

the probability that one would register in the top detector and another at the same time register in the bottom detector was nil. It is hard enough to catch even one neutrino! Therefore by vetoing a dual signal the cosmic rays background was eliminated. Actually the new detector was more complicated than this simple picture implies.

With the new detector, the group went to a new and more powerful reactor at Savannah River Site and repeated the neutrino capture experiment. When the reactor was in operation, they found many more true coincidence counts than when it was shut down. By laboratory calibration of the detector with positron and gamma ray sources, Reines and Cowan could calculate the neutrino capture cross section of protons in the inverse beta reaction.

The Ingredients of Success

At that time nowhere else in the world could this neutrino detection experiment have been performed. There were three essential ingredients. First there was Fred Reines. Fred wanted to do something outstanding in physics. The very difficulty of neutrino detection, the initial negative advice of the sages Bethe and Fermi, made it all the more compelling that Fred should do it. He was not overburdened with the detailed knowledge of the day which held that the Fermi beta decay theory or its possible modification was so satisfactory that the neutrino hypothesis needed no further confirmation. Instead he had the instinct that something so apparently insignificant as a particle with no mass and no charge and no appreciable interactions, must somehow be of very great importance in the scheme of physics. That instinct was well validated by later developments in elementary particle theory and experiment. To enable future understandings the free neutrino had to be directly detected. Furthermore, Fred had the self-confidence to follow his instinct, and the perseverance to see it to the end. He had honed his skills in the big science of nuclear tests. With the reputation that brought him, he could have had a nuclear explosion test as his neutrino source. Although later reflection showed that a nuclear reactor was the more appropriate source, initially he was inspired to consider the experiment because the bomb would have been his to exploit.

Second, there was Carson Mark, the mild, paternalistic, and withal wise leader of the T-Division at Los Alamos. Carson encouraged Reines to try his wings in pure physics, to find his dream project, and although it was in experimental rather than theoretical physics, to follow that dream.

Third, there was the Los Alamos Scientific Laboratory at its finest after its great achievement in the Mike shot, a 10-megaton thermonuclear device which, though a test, presaged the deployment of a whole stockpile of nuclear weapons. The Laboratory provided the infrastructure enabling the successful completion of the neutrino experiment. Although Reines kept his small group rather isolated from the rest of the Lab, and single-mindedly dedicated to their task, always in backup there were the resources of the whole organization.

The Reines/Cowan experiments certainly confirmed qualitatively the existence of the free neutrino. Moreover, after revised analysis of the experimental results, the Savannah River measurement was quantitatively in agreement with the predictions of the 1934 Fermi theory of beta decay. Accounting for parity violation, that theory predicted a cross section of 1.2×10^{-43} cm^2 with an uncertainty of 25% for proton capture of an antineutrino. Subsequent experiments agreed with this value to within 5%. But the significance of the experiment is more profound. It agrees with the current theory of elementary particles and their interactions, the so-called Standard Model (SM), and the Reines/Cowan success instructed and inspired other experiments on neutrinos which aided in the development of that theory.

The Los Alamos group had successfully done the experiment that opened up the experimental field of neutrino physics. It merited the Nobel Prize, which was finally awarded to Frederick M. Reines in 1995. Unfortunately, Clyde Cowan, his close collaborator, had died by then.

A Career in Neutrino Physics

Reines' Viewpoint of Physics

The detection of the free neutrino was not the culmination but the start of Reines' career in neutrino physics. This complex scientist was guided by several different characteristics. To him physics had to have a concrete realism—the neutrino was not a placeholder in a set of complicated equations as in the theory of β decay, but a full partner with the familiar electron in the catalog of elementary particles.

Fred Reines time and again showed an instinct for the significant. After the canonization of the SM, he and collaborators undertook the difficult experiment of the elastic scattering of antineutrinos on electrons to support it. They then did the experiment to measure the lifetime of the proton, which if found would indicate physics beyond the SM. They succeeded, however, in only establishing a lower limit. Earlier Fred had looked at his old data to find evidence for neutrino oscillations, and for a few days thought he had it. Later, critical review of the data did not confirm the oscillations.

Career Change: Reines in Academia

The success of the neutrino detection experiments immediately established Fred Reines as a major experimental physicist. And rightly so. But it was not merely the difficulty of the experiment and the ingenuity of the detection concept, or even the significance of the contribution to basic physics that earned him that acceptance. He had brought careful workmanship to the task. He knew every aspect of the equipment. He checked and rechecked his result, testing and yet again testing to see if it was only an unsuspected background effect which he had interpreted as a true neutrino. Add more cadmium—decrease the time delay for the

neutron capture signal. Check. Replace half the protons with deuterons—decrease the neutrino signal rate. Check. Add more shielding—decrease the background but not the neutrino signal. Check. Always careful, the signal must be justifiable—the wish must not be father to the thought. He did not rush to archival publication, but redid the experiment, moving from Hanford to Savannah River to change a merely probable qualitative detection to a quantitative measurement of the capture cross section with valid statistical support.

Fred realized the importance of his work and wished to continue it. He realized too that this was academic science and, although big science and bound to become bigger, it was not in line with the main mission of the Los Alamos Scientific Laboratory. His experimental group was disbanding; Fred himself had always regarded the peak of the physics profession as a full professorship at a major university. Therefore when Case Institute of Technology in Cleveland offered him the chair of its physics department, and the Atomic Energy Commission promised continued support for neutrino experimentation, he accepted the position. In 1959 he left, but he never deserted Los Alamos—too many good friends, and a vibrant laboratory with capabilities in many areas beyond compare. He remained affiliated with the Lab, and in later years served as chairman of the Academic Advisory Committee for the University of California, following Herb York in that position.

After years at Case, he transferred to the new University of California campus at Irvine (UCI) in 1966, where he became chairman of the physics department and dean of physical sciences. But in keeping with his value system, once again as at Los Alamos, after overpaying his dues to his organization for many years, he set aside his duties and returned to active status as a full professor of physics.

He continued work in neutrino-related experiments in the years after he left Los Alamos. Over the years he was a major researcher in three remarkable experiments fully justifying his place as a significant contributor to the progress of physics. His detection of the free neutrino was not the climax of his career, merely a sensational beginning.

Three Other Great Experiments

Elastic Scattering of Neutrinos on Electrons

Of all the interactions of the elementary particles, from the viewpoint of experimental determination, the hardest is the elastic scattering of a neutrino by an electron. There are two possible ways in which this can happen. The more novel and not included in Fermi's theory is the so-called neutral current interaction. Here the neutrino emits a virtual uncharged Z^o boson which is absorbed by the electron. In the second type of interaction the incident neutrino emits a virtual W^+ and is changed into an electron. The target electron absorbs the W^+ and is changed into an electron neutrino. This is called a charged current

interaction as the charged W^+ is the mediator. Fred had started contemplating this experiment in the mid-1950s at which point neutral currents were not yet anticipated. The very difficult experiment to observe and measure the neutral current scattering of the neutrino was done by Reines and collaborators after 20 years of hard thinking and experience with neutrinos.

The Lifetime of the Proton and Super Nova 1987A

In 1973 Fred and his wife Sylvia were on vacation combined with speaking engagements in Hawaii, completing the last leg of their three-month ever-eastward trip around the world. They met me and my wife Rosalie, by appointment, at the Rockefeller hotel Mona Kao on the big island of Hawaii. The families had been close friends ever since their days at Los Alamos. Now I was with The Aerospace Corporation in Los Angeles, and Fred was close by as dean of physical sciences at UCI.

"There is delight in the illusion
That given wild enough confusion
There follows such precise conclusion."

Fred Reines 11/17/77

Sitting on the beach in front of the Mona Kao, Fred and I were trying to understand whether it was possible to measure the lifetime of the proton. The sea was very rough that day after a severe Pacific storm which had killed several foolhardy tourists at Lahaina. These tourists thought they were very safe on an overlook 25 feet above the water when a great wave plucked them from their perch and sucked them into the ocean. Sunning on the beach, Fred and Harris were watching the waves, mindful of the power of nature and the foolhardiness of men. There was to be no swimming that day, and even talking was difficult above the roar of the breakers.

Talking was difficult too because neither of them knew what they were talking about. Mayer claimed that if the proton were eternally stable, then there must be a symmetry principle that caused it. Symmetries are exact, therefore the proton could not decay. This circular reasoning convinced neither man, but made them wonder if symmetry could be a teeny bit broken. Fred said he wouldn't be convinced that protons are stable unless it was experimentally demonstrated. Both agreed that it was hard to experimentally confirm a nonoccurrence. Fred, as was his nature, said that he would like to try. That was one conversation of many indicating his resolve, in the face of theoretical uncertainties, to build a detector which could at least put a reasonable lower limit on the lifetime of the proton.

The counter was built and about 1980 wax put into operation. Prophetically, Fred put a sign on it with the words "Super Nova Detector." After operation for about a year, the counter, having recorded no approved proton decay events, gave the result that the proton lifetime was greater than 10^{32} years. This lower limit to the lifetime increased with further counting.

Fortunately the detector was ready and working well in time to record the neutrinos from Super Nova 1987A in the greater Magellanic Cloud, the small satellite galaxy 170,000 light years from Earth. Altogether eight

precious neutrino events were unmistakably confirmed. Another 12 neutrino events were found by the Kamiokande detector in Japan on the other side of the Earth. The time coincidence confirmed that the counts came from the same celestial event. Understanding the significance of this result, Fred immediately predicted that hundreds of papers would be written based on this slim score of counts. He was right as usual, partly because of later gamma ray and visible light observations of the remnant supernova. Thus was born a new field, observational neutrino astrophysics, with the almost undetectable particle, because of its penetrating power through matter, revealing information from the inner depths of the exploding star.

Lasting Significance of Reines' Contributions to Physics

What was the significance of the neutrino experiments done at the Los Alamos National Laboratory? They were initiated by one man, Fred Reines, sitting alone in a room in the old T-Division building near Ashley Pond. With the concurrence of Carson Mark, T-Division leader, he was encouraged to "sit and think" and find a fitting pure-science project. Fred was allowed to dream the impossible dream—the Reines/Cowan neutrino capture group turned that dream into reality.

The reality of the existence of the free neutrino and of its detectability legitimized the whole field of neutrino experiments. The neutrino was now to be included as the doublet partner of the electron, its rightful role in the classification of elementary particles. Later, the other two flavors of leptons were arranged as doublets, the muon paired with the mu neutrino, and the tau lepton with its tau neutrino. The combination of Fred's insight and the Los Alamos detection technology was essential to pursue such research. Success depended upon a highly discriminating signature for the rare neutrino interaction event to eliminate uncorrelated backgrounds from reactor neutrons and gamma rays, as well as statistical noise from the detector photomultiplier tubes. This was complemented by a strong veto of a counterfeit signal from cosmic rays. This combination went far beyond the usual discrimination of signal to background and signal to noise using techniques prevalent at that time.

The Reines/Cowan experiment, [1] which demonstrated that in the first family of electron-neutrino pair, both partners were detectable, stimulated search for the free muon neutrino and the tau neutrino of the second and third lepton families. For the detection of the tau, Martin Perl shared the 1995 Nobel prize with Fred Reines. The tau neutrino was detected only recently. These discoveries of particles that were predicted by the SM further increased our confidence in the theory. (See article by Bhattacharya and Gupta in this chapter.)

The Savannah River experiment gave a value for the cross section for the antineutrino capture by the proton, thereby confirming the strength of the weak interaction originally deduced from the lifetimes of nuclei undergoing β decay. Such consistency increased our confidence in the quantitative predictions of the SM.

The small Herr Auge detector with 300 liters of scintillator fluid was superceded, in time, by improved designs culminating in the 8000-ton Cherenkov water detector in the Norton Salt mine in Cleveland, Ohio. The latter experiment was designed primarily to search for proton decay, not surprisingly an experiment in which Reines played a significant leadership role. This gigantic 8000-ton detector of the Irvine, Michigan, Brookhaven (IMB) collaboration, was serendipitously ready in time to catch the neutrinos from the 1987 supernova event. The detection of that explosion initiated the field of neutrino astronomy. Observation of a supernova is opportunistic, several new detectors are now operating, and awaiting an event in our own galaxy. The neutrino flux from an event within 10,000 light years would be 400 times or more larger than that from SN1987A!

But near at hand is our sun. Another pioneer in neutrino observation, Ray Davis, has paved the path to its study. Finally the solar neutrino problem—the large deficit in the observed electron neutrinos compared to those predicted by solar models—seems to be understood. In the deep interior of the sun, some electron neutrinos are converted to mu or tau neutrinos. The populations of the different neutrino flavors oscillate as the neutrinos move outward and to some extent continue on their free space passage to the Earth. Terrestial apparatus however observe only the electron neutrinos, missing those converted to mu or tau neutrinos. The oscillation mechanism, a theoretical addendum to the simple SM, is now being investigated by accelerator-generated neutrinos and observed by detectors descended from the original Reines/Cowan Herr Auge! Thus we find Fred Reines' influence in the entire field of neutrino experiments.

Reines and his contemporary Ray Davis in his measurement of the solar neutrinos, have both shown the virtue of patience in accumulating data. Patience in the extreme is the characteristic of the seekers for the lifetime of the proton. Reines had waited for years without finding a single verified event signifying proton decay. Davis in his experiment captured solar neutrinos in clorine-37 atoms to produce radioactive argon atoms at the rate of only 150 atoms per year. Both men had proved by experience that very rare events could be profitably studied by the exercise of very great patience. Rare events in elementary particle physics may indicate the small breakdown of a symmetry, suggesting a significant modification of theory.

"The Sun and Earth conspire
To dress the moon in
gay attire."

Fred Reines 9/27/89

In the field of astrophysics and cosmology, neutrinos play a very important, but poorly understood, role. Shortly after the Big Bang, much of the energy was in the form of neutrinos which, because of the weak interaction, uncoupled from baryonic matter. That happened much before the decoupling of electromagnetic radiation that formed the currently observable microwave background. The universe is now filled with the remnant neutrino background, in number and in total energy far surpassing the microwave quanta and comparable to all luminous matter. What have these neutrinos experienced in their lifetimes, and what memories may they reveal to us of their birth and infancy if we examine them today? Fred Reines was pondering such thoughts in the 1970s when he was looking over the precious few thousand neutrino events captured in all of his neutrino experiments. Even without the background of the professional cosmologists, Fred's inquiring mind and his universal interests spurred him on to fanciful speculations. Then and to the end he was still dreaming the impossible dream.

Too Late, the Great Prize

In 1995 Fred Reines, professor emeritus at UCI, received the Nobel Prize in physics for his detection of the neutrino. The prize was shared with Martin Perl of Stanford, now also at UCI, for detecting the tau lepton. Fred was pleased but the prize was too late in coming for Clyde Cowan. Alfred Nobel, the philanthropist-entrepreneur discoverer of dynamite, had willed the Nobel prize to be awarded to young achievers who could use the money and prestige of the prize to propel their careers to new heights. Reines' career was at its end, and he had been hospitalized for a long time. He was recovering, but it was not clear until the last moment whether he would be well enough to travel to Stockholm for the award ceremony. He went, but he had one of his collaborators read his acceptance address.

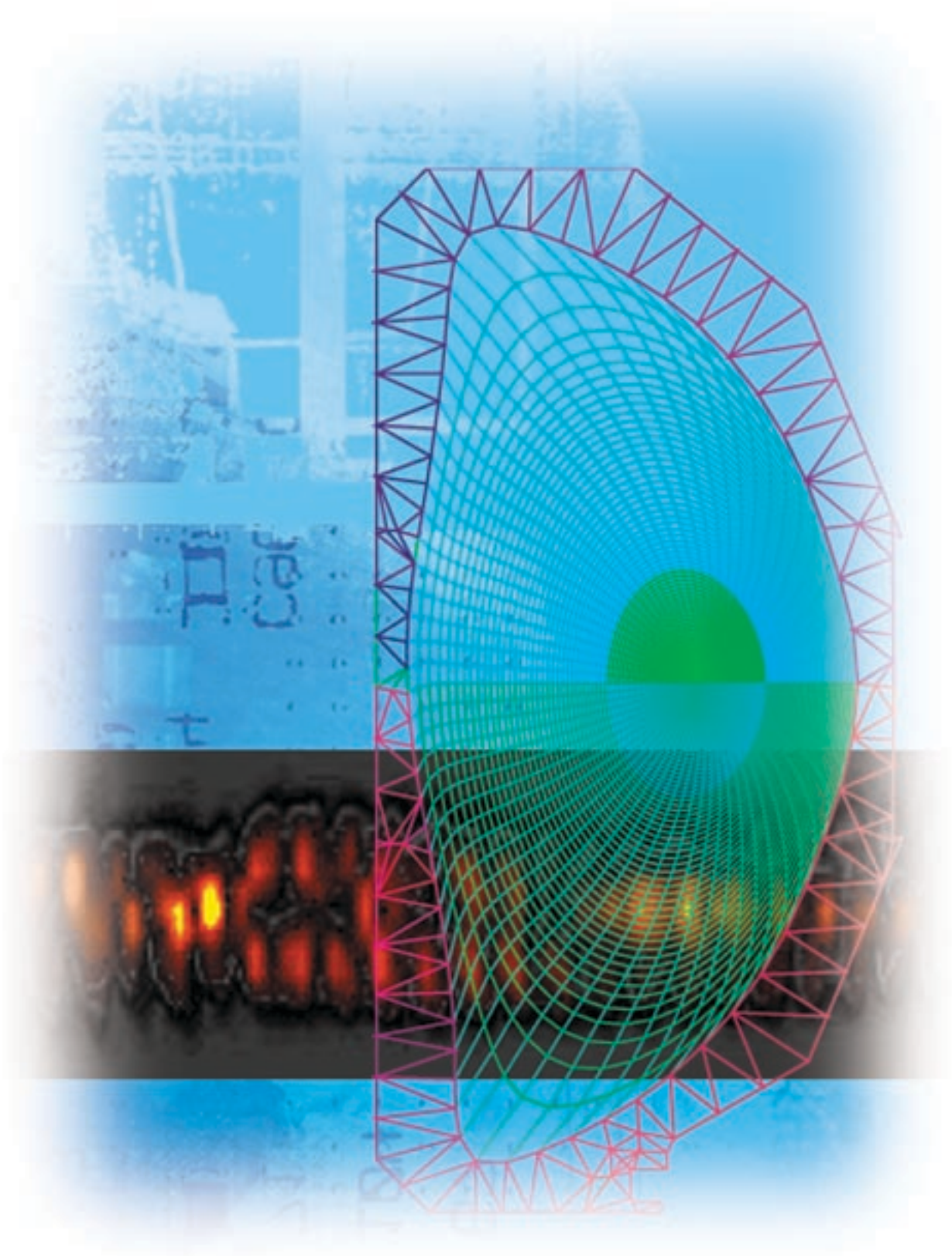
Fred was very proud of the award. Although he professed a modesty over many years, and explicitly never stated that his work deserved the Prize, inwardly he desired it. Many of his friends could not understand why he had been passed over for so many years. Hans Bethe wrote Fred a moving handwritten congratulatory letter. With a bit of cutting humor the note said that Reines had surpassed Bethe in his Nobel. Bethe received his about 20 years after his discovery of the carbon nuclear cycle that generates the energy of one class of stars. Reines received his Nobel 40 years after the detection of the neutrino.

The University of California at Irvine renamed the physical science building after Reines and put a bust of Fred outside the entrance. They also renamed an adjacent building after their other Nobel laureate Martin Perl. Fred was again very proud. But he no longer would do any real scientific work in the office within the building bearing his name.

The Los Alamos National Laboratory, which had so long been involved with the neutrino field which Reines had opened up, had a ceremony in celebration of the award. Fred Reines and his wife Sylvia and his children came. Clyde Cowan was missed. Cowan's wife and daughter were gracious but could not entirely hide their disappointment. They felt that Clyde should have been included in the award. Fred Reines did not speak at the ceremony. Before it was over Sylvia led him out of the auditorium because he was too tired to remain.

Shortly after the ceremony at Los Alamos, Fred returned to the hospital. He went through a long series of illnesses there and in nursing homes. The pioneer of neutrino experiments, Frederick M. Reines, died on August 26, 1998.

Fig. 3. The Poltergeist Project team sharing a light moment in its quest for the elusive neutrino.



A Brief History of Plasma Modeling in Theoretical Division

by Jeremiah U. Brackbill, Dana A. Knoll, and Giovanni M. Lapenta

The potential usefulness of plasma physics in controlled fusion, e.g., controlled fusion in magnetically confined plasmas, has caused strong support for directed research at Los Alamos. Theoretical Division members have contributed to large experiments in magnetically confined plasmas and laser-plasma interactions. T-Division is a world leader in spacecraft instrumentation and observations, especially for the Earth's magnetosphere. Providing basic data equations of state and opacities has made other important contributions. T-Division has also pioneered in modeling plasmas as a fluid. A convergence of many trends is removing many of the constraints imposed on plasma modeling by computer capacity and the limited ability to solve nonlinear systems of equations. There is now both a greater understanding of the problems and an improvement in technique.

The potential usefulness of plasma physics in controlled fusion e.g., controlled fusion in magnetically confined plasmas, has caused there to be strong support for directed research at Los Alamos. Consequently, there have been large plasma theory groups, such as the "Controlled Thermonuclear Theory Group" and the "Laser-Fusion Theory Group," which provided direct support for large experiments. Nevertheless, T-Division has been home for some very distinguished plasma theorists. Among them are Marshall Rosenbluth, who received the Fermi Award from the President of the U.S. in 1985 for "his leadership in the development of modern plasma theory," and Conrad Longmire and Bergen Suydam, who wrote seminal papers on the stability of magnetically confined plasmas while members of T-Division. (Rosenbluth made many important contributions while at Los Alamos in the 1950s, not least of which is his collaboration with Metropolis on the famous Monte-Carlo method. Longmire published a very early text on plasmas, "Elementary Plasma Physics").

Los Alamos has had large experiments in magnetically confined plasmas and laser-plasma interactions and is a world leader in spacecraft instrumentation and observations, especially for the Earth's magnetosphere. T-Division members have contributed to these projects in many ways, some of which are described by Finn, Glasser, Rose, and DuBois in this volume. Other important contributions have been made by providing basic data equations of state and opacities.

Here, we describe T-Division's contributions to our understanding of nonlinear plasma behavior, such as shocks in collisionless plasmas and magnetic reconnection, using plasma simulation.

As an ordinary fluid is heated at constant pressure, molecules dissociate, atoms ionize, and close-range encounters or collisions become less and less frequent. Although one can define a mean velocity and a temperature, these constructs have less and less meaning as individual charged particles travel longer and longer distances without colliding with each other. Instead of colliding, particles interact through the electromagnetic fields their charge and motion produce, and since electromagnetic fields have a very long range, collisions are replaced by "collective interactions" and "wave-particle interactions." Essentially, plasma theory has been about all the ways that a plasma finds to restore thermodynamic equilibrium without collisions.

The most important technique for modeling nonlinear plasma kinetic behavior is based on the particle-in-cell (PIC) method. PIC was invented in the mid-1950s by Harlow, and was a very successful method for hydrodynamics in its time. In Harlow's PIC, each computational particle represents a fluid blob, a piece of a continuous fluid whose motion is determined by interactions with its immediate neighbors. Kinetic plasma simulation uses computational particles that represent blobs of charge. These pass freely through each other but interact over long distances through electric and magnetic fields. PIC is an extremely effective model for a collisionless plasma, which gives good results with an unreasonably small number of computational particles. This happy accident has been exploited in literally thousands of publications and is the main theoretical tool for understanding problems as diverse as energy confinement in tokamaks and magnetic reconnection in the Earth's magnetosphere.

Collisionless shocks occur in very hot, tenuous gases of ions and electrons, in which close encounters between individual ions and electrons are rare. A gas shock is associated and characterized by an abrupt change in the state of the gas with an accompanying increase in entropy. In the absence of collisions, dissipation is provided by long-range, collective interactions among ions and electrons. In a very early paper, PIC simulations produced strikingly beautiful figures representing particle motion in position-velocity or phase space, [1] which explained the basic mechanisms by which shocks convert directed upstream of the shock energy to heat downstream. Major features of this early study have been confirmed many times by experiments, satellite observations, and subsequent simulation studies. From our current world of massively-parallel computers, it is remarkable that these seminal calculations were performed with just 25,000–80,000 particles on a 1,024 grid in one space dimension. However, it would be wrong to assume that the early simulations were easy. They were performed on a CDC 6600, the most advanced mainframe computer available, and they required many hours of computing time. Later, much more ambitious simulations were performed with the WAVE code, which modeled collisionless plasmas in electromagnetic fields in two dimensions.

Modern simulation studies use implicit plasma simulation and focus on understanding the basic mechanisms that cause magnetic reconnection in the Earth's magnetosphere and solar corona. An implicit simulation code, CELESTE, which developed in T-Division as part of an initiative by the U.S. Department of Energy/Office of Fusion Energy to develop a "numerical tokamak" and a figure depicting trapped particle orbits in a toroidal geometry was displayed in the Laboratory's Administration Building main entrance for several years. Implicit methods are designed to model plasmas with realistic electron/ion mass ratios. Explicit methods of the type described above resolve all time scales, for which the computer time scales as

$$(m_{ion}/m_{electron})2d + 1\backslash 2 \quad ,$$

where d is the number of space dimensions. Simulations with

$$m_{ion}/m_{electron} = 1836 \text{ and } d = 3 \quad ,$$

cost the Earth to run, even with the largest, massively parallel machines available now or in the foreseeable future. By contrast, the cost of implicit simulations scale with the mass ratio as

$$(m_{ion}/m_{electron})^{1/2} \quad .$$

(With implicit methods, simulations with $m_{ion}/m_{electron} = 1836$ and $d = 3$ could be run on a single processor PC, if one were willing to wait months for the results.) A general discussion of multiple time-scale problems is given in, [2] and a detailed description of the CELESTE method is given in [3].

A recent result of reconnection studies with CELESTE that not only is nonlinear, but couples kinetic and fluid instabilities in an interesting way, is described in [4]. In the simulations, a well-known kinetic instability, the lower-hybrid drift instability, modifies an initial plasma equilibrium causing shear in the mean flow velocity, which, in turn, drives a well-known fluid instability, the Kelvin-Helmholtz instability. (Secondary instabilities are not unknown, but are typically parametric and involve interactions of waves on similar length and time scales.) This is an important result in two respects. First, it may explain how reconnection can occur in the Earth's magnetotail by the breaking of an invariant. Second, the growth of a secondary instability is confirmed by an independent explicit calculation.

T-Division has also pioneered in modeling plasmas as a fluid. A collisional plasma, in which plasmas are described by the magnetohydrodynamic model, is less expensive than a kinetic simulation and can model larger-scale phenomena. Sometimes, it even gives the right answers.

An early project studied the role of plasmas in the phenomenology of high-altitude nuclear weapons detonations. As one can imagine, the release of so much energy in a relatively small volume has many interesting consequences. One of these, the generation of electromagnetic pulses, has been a lifelong interest of Longmire. Another, the interaction of the plasma formed by the detonation with the Earth's magnetic field, required the development of new techniques for the numerical solution of the MHD equations.

In support of Field Testing Division's program on high-altitude weapons phenomenology, T-Division developed an implicit, arbitrary-Eulerian-Lagrangian MHD code, MOQUI, to model the expansion of a bubble of hot, ionized gas and its rise through the atmosphere. [5]

MOQUI caught the interest of the Controlled Thermonuclear Research Division that needed a three-dimensional numerical model to describe the SCYLLAC experiment, a toroidal theta pinch. SCYLLAC failed to reach its performance goals, and the complexity of the confinement geometry was so great that analysis seemed impossible. Despite the limitations of computer hardware (a CDC 7600 with 132,000 words of fast memory, and 400,000 words of slow memory), a computational accuracy of 0.1 in the computation of the balance of forces was achieved with MALICE, and the underperformance was understood as a natural consequence of the diffuse currents in the actual experiment. Experiments confirmed the accuracy of the computations. [6] MALICE had many innovations, including an arbitrary Eulerian-Lagrangian formulation, an implicit solver coupling hyperbolic and elliptic equations across a free surface, and an elliptic mesh generator with body-fitted coordinates. [7]

Thirty years later MOQUI lives on as MACH2, a workhorse code for the Air Force Research Laboratory in the design of plasma opening switches, and the just-developed MACH3, in which the same basic algorithm is parallelized and supported by a sophisticated grid generator for complex geometries and a graphical user interface.

Some of our current MHD studies also use a PIC method, which has many features in common with kinetic plasma simulations. One recent result with this method is a study of the effect of unstable shear flows on magnetic reconnection in three dimensions. Fluid instabilities are surprisingly effective in causing rapid, localized reconnection. [8] The resulting field line structures reproduce the observed features of flux transfer events, which are thought to be the principal mechanism for reconnection on the side of the magnetopause facing the sun. Very highly resolved calculations of shear flow in two dimensions show that reconnection driven by fluid instabilities occur at a rate that is only weakly dependent on the resisitivity. [9] Studies underway suggest that shearing footprint motion can cause accelerated reconnection in helmet streamers in the solar corona.

A convergence of many trends is removing many of the constraints imposed on plasma modeling by computer capacity and our limited ability to solve nonlinear systems of equations. There is both a greater understanding of the problems and an improvement in technique. Sometimes technique wins, and sometimes understanding wins. Two examples illustrate this point. A Newton-Krylov solver with a multigrid preconditioner yields linear cost scaling with the number of nodes for an elliptic boundary problem with strongly varying coefficients. [10] This new technique, while complex, removes a major stumbling block for implicit codes. More recently, a reformulation of the boundary conditions eliminates the need to solve this equation entirely. [11] Finally, there is just so much more computer capacity now that computer simulations can be incredibly more ambitious than a few years ago. One can do simulations with a million particles on a laptop, and simulations with billions of particles have been performed on massively parallel platforms.

Plasma Theory

Brackbill
DuBois
Finn
Glasser
Longmire
Rose
Rosenbluth
Suydam

by John M. Finn

Theoretical research in plasma physics in Theoretical Division originated in the 1950s when a small group of distinguished theorists began work in controlled fusion, which concentrated on the development of basic plasma theory and magnetohydrodynamics (MHD), with a focus on the study of the stability of pinch devices that were under experimental investigation in the Los Alamos Physics (P) Division at the time. A major focus of the work in plasma theory in the 1990s was the study of magnetic fusion, inertial confinement fusion (ICF), and particle-in-cell (PIC) methods for particle simulations. In addition, theorists in T-Division have worked in several other areas, namely inertial-electrostatic confinement (IEC), astrophysics, the development of MHD codes, strongly coupled plasmas, and several areas of basic plasma physics.

Theoretical research in plasma physics in Theoretical Division began in the 1950s when a small group of distinguished theorists, including Marshall Rosenbluth, Conrad Longmire, and Jerry Suydam, began work in controlled fusion. This program, named "Project Sherwood," was an offshoot of the successful program on fusion weapons. The work in T-Division concentrated on the development of basic plasma theory and magnetohydrodynamics (MHD), with a focus on the study of stability of pinch devices that were under experimental investigation in the Los Alamos Physics (P) Division at the time. The controlled fusion program was created as a classified program, and the work in T-Division continued after declassification in 1958.

With the formation of the Controlled Thermonuclear Research (CTR) Division in the early 1970s, theoretical and experimental magnetic confinement fusion research was focused there, until CTR Division was terminated in the early 1990s. At that time, a number of the theorists from CTR Division came to T-Division, and the Plasma Theory Group was formed.

A major focus of the work in plasma theory in the 1990s was the study of magnetic fusion, and some of this work is described in the article by Alan Glasser in this volume. Harvey Rose and Don Dubois also describe a continuing T-Division effort in inertial confinement fusion (ICF) in this volume. T-Division work over the years in developing particle-in-cell (PIC) methods for particle simulations is described in the article by J. U. Brackbill, also in this volume. In addition, theorists in T-Division have worked in several other areas, namely inertial-electrostatic confinement (IEC), astrophysics, the development of MHD codes, strongly coupled plasmas, and several areas of basic plasma physics.

IEC work has concentrated on forming nonneutral spherical plasmas with ions strongly focused at the center. At present, T-Division researchers are operating an IEC device, in collaboration with experimentalists in P-Division. They have looked at both ion-focused systems and virtual cathode (electron-based) systems. The ion-based systems can operate as neutron sources and have many potential near-term applications. These applications are related to assaying of materials. For example, nuclear assay can be used to find concealed high explosives, highly enriched uranium, chemical weapons, and other items of interest to homeland defense.

The astrophysics work in T-Division has focused on plasma effects, including the possible dynamo (creating of magnetic fields) in accretion disks, and the formation of powerful jets associated with disks, which appear to be related to the dynamics of magnetic fields and are well described by the MHD equations. In space and astrophysics plasmas, the scale of interest is very large; but plasmas, by their nature, involve a wide variety of smaller scales that determine the large-scale behavior and are in turn affected by it. The complete description of macroscopic processes cannot neglect the smallest microscopic level. The implicit kinetic approach to plasma simulation (implicit PIC) described in the article by Brackbill can effectively treat this problem. The method has been applied successfully to a number of problems in solar, magnetospheric, and astrophysics problems. Particularly, in the area of magnetic reconnection described in the following discussion, new discoveries have been made regarding the problem of stability of specific magnetic configurations, such as the Earth's magnetotail.

In addition to the developments in particle simulation methods, T-Division researchers have made a major effort to develop numerical methods for MHD and codes to deal with MHD problems. The article in this volume by Alan Glasser describes the development of the NIMROD code. In addition, T-Division scientists have put much effort into developing completely implicit methods for MHD. These methods allow time steps much larger than those for nonimplicit methods, up to the dynamical time scale for the phenomenon being studied. The breakthrough that has made this possible is the development of Newton-Krylov methods with physics-based preconditioners. These methods have been successfully developed here in T-Division, and codes using the methods have been used in studies of fusion and space plasmas.

Strongly coupled plasmas are plasmas that have correlations strong enough that the usual mean-field descriptions are not applicable. Such conditions can occur in a wide variety of plasmas, including dusty plasmas, ICF plasmas, cooled trapped ions, astrophysical plasmas, and ultracold plasmas. Generally, the conditions for strong coupling are a combination of low temperature, very high-charge state, or high density. Plasma Theory Group personnel have put forth a very active effort in strongly coupled plasma physics from the pure theoretical approach (e.g., integral equations and quantum kinetic theory) and molecular dynamics simulations (e.g., classical pair interaction and semiclassical multispecies). These approaches are applied to calculations generalizing known plasma physics results. These calculations include stopping

power, temperature relaxation, and electrical conduction. We have worked in describing other novel aspects of strongly coupled plasmas, such as freezing, disorder-induced heating/cooling, and lattice-like shear modes.

Basic plasma studies have included work on Penning trap plasmas, magnetic reconnection, and turbulence in plasmas and fluids.

A Penning trap is a trap for a pure electron plasma, based on a uniform magnetic field and electrostatic fields created by electrodes at the ends. The magnetic field confines the electrons in the direction perpendicular to itself, and the applied electric field confines the electrons parallel to the magnetic field. The nonneutral plasma contributions related to Penning traps, in addition to the work described above under IEC devices, include the discovery that the curvature in the electrostatic sheaths at the ends of the device leads to effects that are analogous to those present in geophysical fluid dynamics, namely the existence of Rossby waves and an instability previously observed in experiments but predicted not to exist in earlier theories. Since this discovery, the analogous instability has been observed in geophysical fluid dynamics studies.

Magnetic reconnection studies in T-Division have application to astrophysical plasmas, as discussed previously, and to magnetic confinement fusion. Much of the work in basic reconnection studies has used the Newton-Krylov codes, the development of which was discussed previously. These basic studies involve studies of the nonlinear behavior of tearing instabilities, modes which spontaneously change the topology of the magnetic field lines and the response of the plasma to applied magnetic perturbations (the so-called forced reconnection problem). These studies have included many different physical effects, including the Hall effect, finite electrical conductivity, hyperresistivity (anomalous electron viscosity), and electron inertia. Investigation of these physical effects has proceeded by means of the implicit MHD codes and the implicit PIC codes described previously.

Basic plasma and fluid turbulence studies in T-Division have concentrated on understanding the turbulent MHD dynamo. Such a mechanism is believed to be responsible for the sustainment of magnetic fields in astrophysical bodies, such as the Earth, and in laboratory fusion experiments, such as the reversed-field pinch (RFP). The earliest analysis of this phenomenon used a statistical approach for an inviscid, perfectly conducting magnetofluid that employs the concept of an absolute equilibrium ensemble based on the conserved quantities of energy and magnetic helicity. Although this approach led qualitatively and naturally to the “reversed” magnetic field observed in RFPs, it also had the characteristic pathology of continuum statistical mechanics, the Rayleigh-Jeans ultraviolet catastrophe.

More recent T-Division developments include a highly compact formalism, which has been applied to inhomogeneous fluid turbulence. This formalism allows one to remove this catastrophe by incorporating the dissipative effects of resistivity and viscosity. Our formalism has two novel aspects: (1) an adaptation of helicity basis functions to represent an arbitrary solenoidal field and (2) the appropriate use of an approximation of random phase when helicity is present. At the same time, MHD simulations of the RFP dynamo have led to a fuller understanding of the role of mode coupling and velocity shear on the turbulent RFP dynamo.

The current efforts in T-Division described previously have a highly developed interdisciplinary aspect. For example, the IEC work, with the objective of developing neutron sources, is complemented by the basic Penning trap studies. In addition, there is considerable overlap in the MHD investigations in magnetic fusion, in plasma astrophysics, and in basic magnetic reconnection.

Theoretical Division and the NIMROD Plasma Simulation Code

Glasser
Knoll

by Alan H. Glasser

Theory of plasmas, or ionized gases, the fourth state of matter, after solids, liquids, and gases, is a central goal for Theoretical Division. Although the division has recently expanded and diversified into many applications of plasma theory, including strongly coupled plasmas, plasma astrophysics, and nonenergy applications of fusion reactions, the theory and computer simulation of plasmas for magnetic fusion energy (MFE) remains at the core of group interests and is strongly supported by the Department of Energy Office of Fusion Energy Sciences (DOE/OFES). The goal of MFE research is the controlled release of the energy of thermonuclear fusion reactions that power the sun, the stars, and the hydrogen bomb—for peaceful uses.

The central focus of Theoretical Division plasma research is the theory of plasmas, or ionized gases, the fourth state of matter, after solids, liquids, and gases. The plasma researchers were originally the theory group of the Los Alamos Controlled Thermonuclear Research (CTR) Division until the termination of that division in 1991, when the group moved into T-Division. While it has recently expanded and diversified into many applications of plasma theory, including strongly coupled plasmas, plasma astrophysics, and nonenergy applications of fusion reactions, the theory and computer simulation of plasmas for magnetic fusion energy (MFE) remains at the core of group interests and is strongly supported by the DOE/Office of Fusion Energy Sciences (DOE/OFES).

The goal of MFE research is the controlled release of the energy of thermonuclear fusion reactions that power the sun, the stars, and the hydrogen bomb, for peaceful uses. The key difficulty is isolating the 100-million-degree plasma fuel from its surroundings. The concept of magnetic plasma confinement exploits the fact that a magnetic field can be very effective at inhibiting the motion of particles and fluids across field lines but not along them.

The best-known plasma confinement system is the tokamak, a Russian acronym for "toroidal chamber, magnetic coil," in which magnetic field lines wind helically through a toroidal chamber, generated in part by external field coils and in part by currents induced to flow through the plasma itself, pinching the plasma inward. Because the magnetic field lines, in principle, remain within the torus, the plasma cannot escape along the field lines and remains well confined across them. There are many other variations on this theme of toroidal confinement, with names such as stellarator, spheromak, and reversed-field pinch.

The theory of MFE plasmas is concerned with the equilibrium, stability, and transport of plasmas in such magnetic confinement systems, and the nonlinear evolution of instabilities and their effects on confinement.

A direct consequence of this concept is a fundamental difference in the requirements for accurate numerical study of magnetically confined plasmas, compared with other areas of computer simulation—extreme anisotropy, the difference in the behavior of the plasma in different directions, along and across magnetic field lines. The ratio of parallel to perpendicular thermal conductivity can be as high as 10^{10}. If numerical approximations cause a small amount of the fast parallel conductivity to "leak" into the transverse direction, the results can be nonsense.

In addition, a large range of time scales of phenomena is in the plasma. Waves propagate much faster across the magnetic field than along it. Time scales of interest range from periods of gyration of electrons and ions about the magnetic field to the periods of slow and fast waves to the growth times of slowly growing instabilities to the collisional and turbulent transport of plasma across the magnetic fields. A successful numerical simulation must study the longest-time behavior while dealing accurately and efficiently with the short-time behavior.

The Plasma Theory Group has been at the center of the NIMROD plasma simulation code development project since its initiation by DOE/OFES in February 1996. The goal of this work is a code capable of long-time simulation of MFE plasmas. To deal with the demands of extreme anisotropy, the code uses two-dimensional (2-D) finite elements on a nonorthogonal coordinate system based on the equilibrium magnetic field, allowing for a clean separation of the physics along and across this field. Variations in the third direction, around the torus the long way, are treated by expansion in a Fourier series. Multiple time scales are treated by means of a semi-implicit time step, a process that requires solution of large systems of simultaneous equations.

The code was written from the beginning for efficient use of massively parallel computers, using message-passing to coordinate the distributed effort. A geographically distributed team from many institutions developed the code, but the core of the code was mainly developed at Los Alamos. Other features of the NIMROD code include advanced visualization, a user-friendly graphical interface, and object-based coding that uses Fortran 90. The code is freely available for download from the Internet. NIMROD has been, and continues to be, used by a large number of researchers around the country and has had a major impact on the development of MFE plasma theory.

Figure 1 shows an example of the computational grid used in NIMROD. The inner region, where the magnetic field lines are close and plasma is confined, uses a logically rectangular grid aligned with the equilibrium

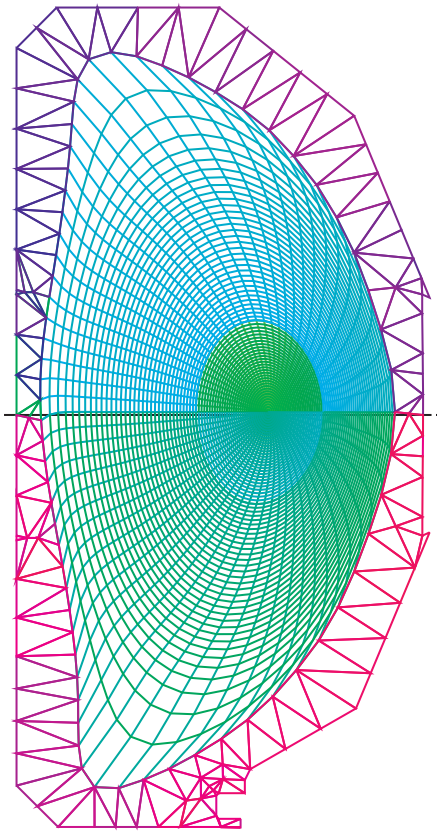


Fig. 1. NIMROD computational grid.

magnetic field. The outer vacuum region, where the field lines are open, uses an unstructured grid of triangles. Figure 2 shows a Poincare surface of section, illustrating how the magnetic field is broken up and made stochastic by nonlinearly evolving plasma instabilities, destroying its ability to confine the plasma.

Although the existing NIMROD code has thus proved very useful, it has some essential limitations. In the presence of nonlinearly evolving helical distortions, the static, axisymmetic coordinate system does not remain aligned with the dominant magnetic field. In addition, unlike tokamaks, stellarators are initially and inherently nonaxisymmetric and cannot be treated by NIMROD. Furthermore, the semi-implicit time step is not capable of efficiently treating some important instabilities. Two new efforts are underway in the Plasma Theory Group to improve these capabilities. One is to develop a 3-D moving adaptive grid that automatically aligns itself with the evolving magnetic field and packs itself into regions of sharp gradients across this field. The other is to develop an efficient fully implicit time step, using physics-based preconditioning of Newton-Krylov methods. The latter project is part of a larger Laboratory-Directed Research and Development Directed Research project in T-Division organized by Dana Knoll in T-Division.

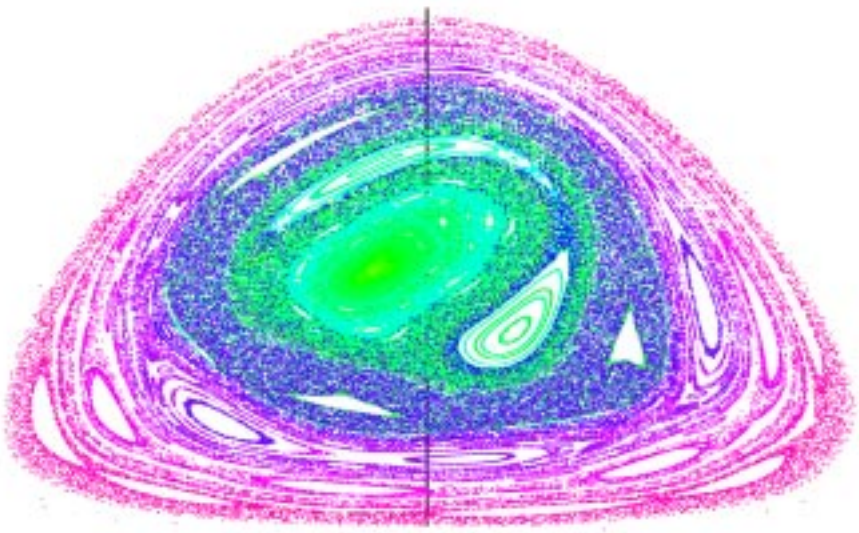


Fig. 2. Poincare surface of section.

Nonlinear Coupling of Coherent Electromagnetic Waves with Plasma

by Donald F. DuBois and Harvey A. Rose

DuBois
Rose
Zakharov

Because of its relevance to laser fusion the interaction of intense laser beams with plasma has sustained the attention of scientists. Theoretical Division has provided the intellectual environment in which researchers coming from diverse subfields of physics, including quantum mechanics and fluid turbulence, could work together in order to understand the rich set of phenomena which result when these beams pour energy into a plasma faster than it can be dissipated. The resultant nonequilibrium plasma is home to a zoo of alternate pathways of energy absorption, including the generation of plasma waves and scattered light waves. Deducing an engineering model of these microscopic processes to arrive at a practical predictive tool is a major challenge.

The interaction of intense laser beams with plasma has sustained the attention of scientists because of its relevance to laser fusion. Theoretical Division has provided the intellectual environment in which researchers coming from diverse subfields of physics, including quantum mechanics and fluid turbulence, could work together in order to understand the rich set of phenomena which result when these beams pour energy into a plasma faster than it can be dissipated. The resultant nonequilibrium plasma is home to a zoo of alternate pathways of energy absorption, including the generation of plasma waves and scattered light waves. The size (duration) of fusion relevant plasma is so large compared to the wavelengths (period) of these secondary waves that direct numerical simulation will not be a viable tool for the foreseeable future. A major challenge is to deduce an engineering model of these microscopic processes so as to arrive at a practical predictive tool.

Some of the basic processes associated with laser-produced plasma are also found in the ionosphere when it is irradiated by intense radio waves. A generalization of a plasma wave model introduced by Zakharov in 1972, to include coupling to laser/radio beams, elucidates a particular pathway for energy absorption in a regime appropriate both to the ionospheric and laser examples. Plasma waves condense into small regions of space, leading to regions of diminished plasma density. These "cavitons" are roughly akin to the cavitation bubbles that form in the wake of a high-speed propeller. During their life cycle, they violently shrink in size, thereby changing the plasma wave frequency, just as the frequency of a plucked string changes with its length. Our work made several novel advances and predictions concerning the caviton cycle, the frequency spectrum of plasma waves, and the density dependence of such spectra. Some of these spectral predictions were unlike any based on previous theories. This research had a seminal effect on

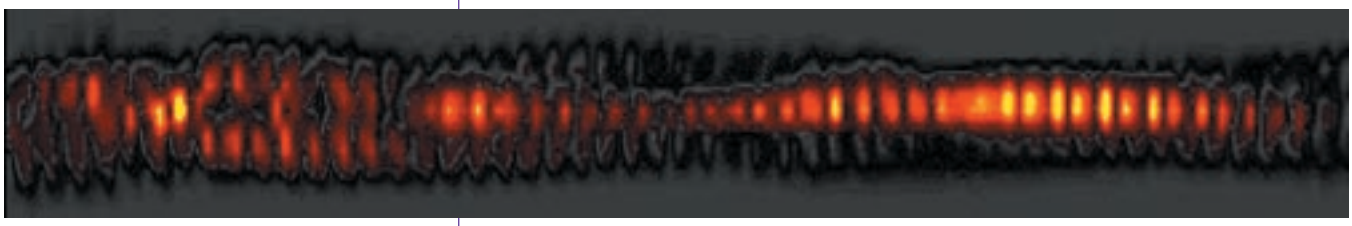


Fig. 1. When laser light is focused onto a fusion target, elongated regions of elevated laser intensity are formed, generating intense plasma wave activity. A portion of the simulation [3] region is shown in which the wave-energy density is modulated, forming a necklace of cavitons.

research on plasma waves excited by powerful radio waves in the F layer of the ionosphere. A decade of experiments at the Arecibo radio telescope and the EISCAT Observatory in Norway have, [1, 2] as of 2001, substantially verified all of our theoretical predictions. (These predictions were not initially accepted by the community due to the novel elements of the theory.) Further research showed that similar phenomena may occur in the laser case, limiting the efficiency of some laser fusion schemes. An example is shown in Fig. 1.

A limitation of this modeling is that only plasma waves are explicitly considered, while the independent electron and ion kinetics are omitted. A new, efficient, simulation technique was developed in Applied Physics Division's Plasma Physics Group, in collaboration with T-Division, which for the first time also allowed for full particle kinetics. In the low-temperature regime of a fusion plasma, excellent agreement with the generalized Zakharov model was obtained. [4] But in the high temperature regime the kinetic simulations showed that electron trapping [5] in the plasma wave is a dominant mechanism. This has led to a re-evaluation of what kinds of models are required to simulate the entire duration of a laser fusion plasma.

Another effect of electron trapping led to the prediction, by T-Division theorists, that laser absorption/scatter should also occur by coupling to the electron acoustic wave—a plasma wave which defies the conventional wisdom about what kind of plasma oscillation is possible. This "stimulated electron acoustic scatter," SEAS, was then observed at experiments performed at the Laboratory's TRIDENT laser facility. Not only was this the first SEAS observation, but the first definitive experimental proof that electron trapping effects can play a dominant role in stimulated scatter.

The attainment of fusion through inertial confinement, with energy provided by intense laser beams, requires precision control of their energy deposition in the target. Experimental observation of beam deflection at NOVA led to the prediction by T-Division theorists that this was caused by the coupling between the beam and plasma flow. Measurements of deflection [6] at TRIDENT are in quantitative agreement with the predicted dependence on laser and flow parameters, thus allowing for *a priori* control of this process by the engineers.

Theoretical Division, The Future

by Francis H. Harlow

The future of the Theoretical Division is filled with exciting challenges for both the near-term research directions and the long-term problems that should keep the division active and productive for at least the next 60 years. An excellent summary of the issues is contained in the "Science Thrusts" section of the 2002 T-Division Self-Assessment, Los Alamos National Laboratory report. Excerpts are presented here with minor revisions.

Innovative Science in Support of Stockpile Stewardship

A primary mission of the Laboratory is to certify the stockpile in the absence of nuclear testing. T-Division undertakes research in nuclear and atomic physics, materials science, hydrodynamics, and uncertainty quantification—capabilities that are all used in the Laboratory's stockpile stewardship and design codes. The thrust brings together, and provides coherence and enhanced visibility for, numerous new and innovative scientific concepts being pursued at the Laboratory and other U.S. Department of Energy (DOE) laboratories. The research is conducted by collaborative teams that bring T-Division scientists into working groups with designers, code integrators, and experimentalists in other Laboratory divisions.

The thrust focuses on the following research areas:

(1) nuclear and atomic physics and interactions of matter with radiation (nuclei far from stability; quantum molecular dynamics simulations of density effects; novel diagnostics for mix; material processes for "Boost"; and opacities);

(2) materials science (constitutive formulations developed from multiscale modeling for metals, high explosives, polymers, and other weapons materials; reaction laws; aging analysis; and equations of state);

(3) numerical methods and software for hydrodynamics (e.g., tetrahedral mesh hydro, interface treatments, constitutive laws for particular mesh resolutions); and

(4) uncertainty quantification (metrics for stockpile certification and weapons physics in general; analysis of experimental data [both historical and modern]).

These efforts are coordinated with one another and with parallel basic research initiatives for nonweapons materials and issues. All of these

issues are increasing in significance because of subtle variations in manufacturing and aging characteristics that need to be accounted for with enhanced fidelity in the maintenance of the stockpile. The demonstration of applications to problems outside the weapons arena will broaden the approach beyond weapons materials and issues and will leverage investments in those areas.

Simulation-Based Predictions for Multiscale Real-World Problems

Simulation-based predictions play an expanding and increasingly essential role in the solution of critical national problems. To be effective in this role, such predictions must reflect the highest-quality science that is available for the problem. The simulation-based predictions must also be accompanied by rigorous estimates of their level of accuracy and range of applicability. Generally, predictive simulations of complex systems must incorporate multidisciplinary expertise in modeling of multiscale and nonequilibrium processes, uncertainty quantification, numerical algorithms, and modern software techniques. T-Division is uniquely positioned to lead this thrust owing to its leading role in initiating efforts in multiscale science and uncertainty quantification, the existence of in-place multidisciplinary teams, and the broad scientific resource base represented in the division.

This thrust requires integration of the following capabilities:

- uncertainty quantification for simulation-based predictions;
- advanced software design;
- multiuse, reliable, and efficient software;
- robust and accurate algorithms;
- powerful techniques for modeling multiscale and nonequilibrium processes;
- effective methods for making maximum use of data to constrain uncertainties; and
- new approaches for modeling complex stochastic processes.

The capabilities represented by this thrust map directly onto needs arising in the national security and civilian missions within the DOE, as well as those of the U.S. Department of Defense (DOD) and of certain parts of the private sector. Application areas that will benefit mankind include stockpile stewardship, energy and environment, critical infrastructure, bioscience, global climate change, manufacturing, and the nuclear fuel cycle.

Extreme States of Matter

A new generation of large-scale laboratory facilities will enable us to probe matter at very high energy densities, often in the presence of intense (multimegagauss) magnetic fields. Rapid time variations in temperature, density, and other parameters usually preclude thermal equilibrium from being achieved in the experiments performed at these large facilities.

A quantitative understanding of the experimental data, especially strong field and/or nonequilibrium effects, is needed in order to effectively use the data to improve equations-of-state (EOS), opacity, and particle-transport models in weapons simulation codes. New astronomical facilities are probing exotic cosmic "laboratories," such as supernovae and the environs of compact objects (white dwarfs, black holes, and neutron stars), in which very high-energy particles and radiation are produced. Detailed modeling of these astrophysical phenomena relies on accurate descriptions of matter under extreme conditions. All these astronomical facilities are coming on line in the same time frame that the nuclear weapons program is providing the DOE weapons laboratories with tremendous enhancements in computational capabilities. Consequently, we have a great opportunity to capitalize on these contemporaneous developments and dramatically improve our understanding of extreme states of matter important to both the weapons program and the international astronomical community.

Staff expertise and experimental capabilities at Los Alamos, combined with recent technical developments, suggest three specific topics for investigation: elementary processes in strongly coupled plasmas, properties of warm dense matter, and properties of intensely magnetized matter.

These efforts will lead to new models of opacity, EOS, reaction rates, etc., that are valid under a much broader range of physical conditions than those of currently available models. Besides helping to implement these new models in codes now being used to simulate weapons and to understand and plan high-energy density aboveground experiments, we intend to develop a similar suite of new capabilities to analyze astrophysical observations.

Fundamental Processes

Experimental and observational advances over the last decade have led to unprecedented progress in our knowledge and understanding of the universe, both in the realms of small and of large scales. The most intriguing aspect of this enhanced understanding is the strong coupling between information obtained on both large and small scales. Present cosmological observations (microwave background, large-scale structure) not only provide solid evidence for a new form of "dark" matter but also strongly constrain the relevant particle and nuclear physics (e.g., particle masses). Conversely, present-day thinking in particle physics often drives model-building in cosmology (e.g., inflation). So tight is this coupling that one now speaks of connecting "quarks to the cosmos." This new cross-disciplinary field encompasses key aspects of astrophysics, atomic physics, and nuclear and particle physics: these aspects we refer to as "fundamental processes." Fundamental processes include the basic building blocks of nature and their interactions; the properties of nuclei and nuclear reactions; the evolution of the universe and the formation of large-scale structure and matter in novel states, such as Bose-Einstein, fermion, and molecular condensates and the quark gluon plasma.

These new problems include understanding the consequences of extensions of the Standard Model for cosmology, mapping the dark matter in the universe using gravitational lensing, understanding neutrino flavors and masses in the light of recent and planned experiments, improving the nuclear physics underlying big bang nucleosynthesis and supernovae, and generating, understanding the dynamics of, and applying new types of quantum condensates. Solving these difficult and complex problems will require active collaboration of physicists who possess expertise in multiple areas of research.

Institute for Quantum Computation at Los Alamos

Several recent developments in quantum computing indicate that now is an opportune time to create at Los Alamos an Institute for Quantum Computation. Reflecting this Laboratory's acknowledged position as one of the world leaders in experimental and theoretical efforts on quantum computation, the National Security Agency is currently funding four independent quantum-computing research efforts here. A new version of a quantum computer (called a "Type II" quantum computer) is being funded by the US Air Force. The Type II quantum computer is designed to solve general finite-difference equations, including many algorithms of interest to the main missions of Los Alamos and not requiring long decoherence times. Advances in self-assembly of quantum bits (qubits) appear capable of creating arrays containing a very large number of such bits orders of magnitude more than previously assembled. There is a great opportunity to capitalize on these recent developments and dramatically improve our understanding of the design and optimization of quantum computers.

Five specific topics for investigation are

- solid-state quantum computation and spin measurements;
- quantum cryptography and ion-trap quantum computation;
- quantum measurements and optical quantum computers;
- nuclear magnetic resonance quantum computation and quantum error correction; and
- modeling dynamics of large quantum computers, qubit measurements, and decoherence.

Expected outcomes include operational quantum computers before the year 2020 and the creation of many new applications for quantum computers and quantum measurement capabilities. Benefits include a many orders-of-magnitude increase in computational searches, numerical modeling, and for a specific example, the factoring of large integers. Computational abilities of the DOE weapons laboratories will be tremendously enhanced.

Understanding and Designing Complex Materials: Bridging Multiple Length and Time Scales

Many high-performance materials needed for industry, energy, and the Laboratory's internal needs are so complicated that new, predictive, integrated, multiscale theoretical techniques must be developed to understand and design them for varied applications. Experimental data alone will be insufficient. The ability to bridge multiple length and time scales is critical to success and predictability because these materials often have properties that are controlled by phenomena at many connected functional length scales to form "systems." (Examples include nanosystems, dopants, dislocations, surface additives, contaminants, interfaces, grains, and composites.)

An integration of synthesis, structure, and property relations is, of course, also essential. Specific applications potentially span the following areas: organic, inorganic, biological, and hybrid materials (including nanomaterials), biological and biomimetic materials, sensors, catalysts, polymers and soft materials, organic/inorganics, metals, and weapons materials. The range of techniques that must be integrated includes electronic-structure, atomistic, coarse-grained, chemical, engineering, and systems. New statistical, stochastic reduced models must be developed that capture the critical scientific parameters and variables at intermediate length and time scales. Experimental guidance and validation of novel theoretical and modeling techniques will be critical.

Modeling Biological Systems: Whole-Cell Modeling

Studies of biological systems have reached a level of detail and sophistication requiring the development of mathematical, computational, and physical models that integrate multiple length and time scales. New developments in experimental techniques can now probe the interrelationship among large numbers of proteins during performance of all of a cell's functions. Areas where physical and phenomenological models will be necessary are in the modeling of immune-regulatory and transcription-control networks. Another research area related to the understanding of cell function is the study of the structure and dynamics of protein machines. The techniques that must be integrated in order to excel in this new area of research are those of bioinformatics, sequence databases, disease modeling, cell signaling, molecular dynamics, protein-structure dynamics and prediction, and dynamical systems. New phenomenological models must be developed in collaboration with experimental groups working on cell-networks characterization.

This thrust brings the uniquely powerful combined expertise of T-Division into collaboration with that of other divisions to study problems at the cutting edge of present-day research, including the following:

- viral disease dynamics;
- gene regulatory networks;
- immune regulatory networks;
- protein-folding and structure prediction;
- biomolecular dynamics, molecular machines, and enzyme catalysis; and
- bioinformatics.

Additional Thrust Directions

Going beyond the science thrusts contained in the *2002 Self-Assessment*, we can anticipate a succession of new problems of crucial concern to national welfare but not yet precisely identified. Issues of homeland defense, for example, are leading to the realization of numerous problems to which T-Division experience and capabilities can be applied. With the astonishing evolution of high-speed supercomputers, we shall be more and more challenged to use our basic analysis techniques to solve problems that go far beyond the physical and biological sciences. Our stochastic mathematical techniques will no doubt be adapted more and more to problems involving interacting populations of people, including world economics and the psychological roots of terrorist activities, for which precise quantification of whole new sets of relevant variables will be required.

Perhaps the most pressing national (and world) problems to which T-Division can make significant contributions center on environment and energy. Examples are

- safe and efficient nuclear power sources without the problems of noxious waste;
- desalinization and pumping of seawater to remote areas;
- new ways to replace petroleum dependence by other sources of energy for vehicle transportation; for example, liquid fuels from garbage and/or seaweed; and
- global atmospheric and oceanic issues; such as storms, climate, global-warming, pollution removal, and mitigation of effects from extraterrestrial bombardment.

The ultimate goal of science is to understand nature and to use that understanding for the betterment of all facets of being on this remarkable planet. T-Division can and will make lasting and significant contributions toward the accomplishment of this quest.


n
n
n
n
p
n

Appendix A: References

Astrophysics

John R. Baumgardner, "Unlocking the Secrets of the Rocky Planets"

[1] H.-P. Bunge, M. A. Richards, and J. R. Baumgardner, "The Effect of Depth-Dependent Viscosity on the Planform of Mantle Convection," *Nature* **379**, 436 (1996).
[2] H.-P. Bunge, et al., "Time Scales and Heterogeneity Structure in Geodynamic Earth Models," *Science* **280**, 91 (1998).
[3] W.-S. Yang and J. R. Baumgardner, "Matrix-Dependent Transfer Multigrid Method for Strongly Variable Viscosity Infinite Prandtl Number Thermal Convection," *Geophys. and Astrophys. Fluid Dyn.* **92**, 151 (2000).
[4] M. A. Richards, W.-S. Yang, and J. R. Baumgardner, "The Role of a Low-Viscosity Zone in Stabilizing Plate Tectonics: Implications for Comparative Terrestrial Planetology," *Geochem., Geophys., Geosys.* **2** (2001).
[5] C. C. Reese, V. S. Solomatov, and J. R. Baumgardner, "Survival of Impact-Induced Thermal Anomalies in the Martian Mantle, *J. Geophys. Res.* **107**, 143 (2003).
[6] D. R. Stegman, et al., "An Early Lunar Core Dynamo Driven by Thermochemical Mantle Convection," *Nature* **421**, 143 (2003).

Stanislaw M. Ulam, "On the Possibility of Extracting Energy from Gravitational Systems by Navigating Space Vehicles"

[1] J. Oxtoby and S. Ulam, "Measure-Preserving Homeomorphisms and Metrical Transitivity," *Ann. Math.* **42**, 874 (1941).
[2] K. A. Ehricke and G. Gamow, "A Rocket around the Moon," *Scientific American* **196**, 47 (1957).

Atomic Physics

John J. Keady, "Opacity Efforts"

[1] H. Mayer, "Methods of Opacity Calculations," Los Alamos Scientific Laboratory report LA-647 (1948).
[2] A. L. Merts, R. D. Cowan, and N. H. Magee, Jr., "The Calculated Power Output from a Thin Iron Seeded Plasma," Los Alamos Scientific Laboratory report LA-6220-MS (1976).
[3] W. F. Huebner, et al., "Astrophysical Opacity Library," Los Alamos Scientific Laboratory report LA-6760-MS (1977).
[4] A. N. Cox, "Stellar Absorption Coefficients and Opacities," in *Stellar Evolution*, L. H. Aller and D. B. McLaughlin, Eds. (University of Chicago Press, Chicago, 1995), p. 195.
[5] P. T. Springer, et al., *J. Quant. Spectrosc. Ra.* **58**, 927 (1997).
[6] R. D. Cowan, *The Theory of Atomic Structure and Spectra* (University of California Press, Berkeley, 1981).

James S. Cohen, "Muon-Catalyzed Fusion"

[1] F. C. Frank, "Hypothetical Alternative Energy Sources for the 'Second Meson' Events," *Nature* **160**, 525 (1947).
[2] A. D. Sakharov, *Rep. Phys. Inst. Acad. Sci. USSR* (1948); republished (in English) "Passive Mesons," *Muon Catal. Fusion* **4**, 235 (1989).
[3] Y. B. Zeldovich, "Reactions Initiated by μ-Mesons in Hydrogen," *Dokl. Akad. Nauk USSR* **95**, 493 (1954).
[4] L. W. Alvarez, et al., "Catalysis of Nuclear Reactions by μ-Mesons," *Phys. Rev.* **105**, 1127 (1957).
[5] S. E. Jones, et al., "Experimental Investigation of Muon-Catalyzed *d-t* Fusion," *Phys. Rev. Lett.* **51**, 1757 (1983).
[6] S. S. Gershtein and L. I. Ponomarev, "μ-Meson Catalysis of Nuclear Fusion in a Mixture of Deuterium and Tritium," *Phys. Lett.* **72B**, 80 (1977).

[7] J. S. Cohen, R. L. Martin, and W. R. Wadt, "Diabatic-State Treatment of Negative-Meson Moderation and Capture. I. The Hydrogen Atom," *Phys. Rev. A* **24**, 33 (1981).

[8] J. S. Cohen, R. L. Martin, and W. R. Wadt, "Diabatic-State Treatment of Negative-Meson Moderation and Capture. II. Mixtures of Hydrogen and Helium," *Phys. Rev. A* **27**, 1821 (1983).

[9] J. S. Cohen, "Slowing Down and Capture of Negative Muons by Hydrogen: Classical-Trajectory Monte Carlo Calculation," *Phys. Rev. A* **27**, 167 (1983).

[10] J. D. Garcia, N. H. Kwong, and J. S. Cohen, "Time-Dependent Quantal Treatment of Muon-Hydrogen Collisions," *Phys. Rev. A* **35**, 4068 (1987).

[11] N. H. Kwong, J. D. Garcia, and J. S. Cohen, "Classical-Quantal Coupling in the Capture of Muons by Hydrogen Atoms," *J. Phys. B* **22**, L633 (1989).

[12] J. S. Cohen, "Molecular Effects on Antiproton Capture by H_2 and the States of $\bar{p}p$ Formed," *Phys. Rev. A* **56**, 3583 (1997).

[13] J. S. Cohen, "Isotope Effects on Antiproton and Muon Capture by Hydrogen and Deuterium Atoms and Molecules," *Phys. Rev. A* **59**, 1160 (1999).

[14] M. C. Struensee, J. S. Cohen, and R. T Pack, "Improved Adiabatic Calculations of the Vibrational-Rotational States of HD^+," *Phys. Rev. A* **34**, 3605 (1986).

[15] J. S. Cohen and M. C. Struensee, "Improved Adiabatic Calculation of Muonic-Hydrogen-Atom Cross Sections. I. Isotopic Exchange and Elastic Scattering in Asymmetric Collisions," *Phys. Rev. A* **43**, 3460 (1991).

[16] J. S. Cohen, "Improved Adiabatic Calculation of Muonic-Hydrogen-Atom Cross Sections. II. Hyperfine Transitions and Elastic Scattering in Symmetric Collisions," *Phys. Rev. A* **43**, 4668 (1991).

[17] J. S. Cohen, "Improved Adiabatic Calculation of Muonic-Hydrogen-Atom Cross Sections. III. Hyperfine Transitions in Asymmetric Collisions," *Phys. Rev. A* **44**, 2836 (1991).

[18] J. S. Cohen and R. L. Martin, "Interaction Matrix Elements for Resonant Muonic-Molecule Formation," *Phys. Rev. Lett.* **53**, 738 (1984).

[19] V. Zeman, E. A. G. Armour, and R. T Pack, "Treatment of the $t\mu + D_2$ Reaction by the Methods of Quantum Reactive Scattering," *Phys. Rev. A* **61**, 052713 (2000).

[20] J. S. Cohen and M. Leon, "New Mechanism for Resonant $dt\mu$ Formation and Epithermal Effects in Muon-Catalyzed Fusion," *Phys. Rev. Lett.* **55**, 52 (1985).

[21] M. C. Fujiwara, et al., "Resonant Formation of $d\mu t$ Molecules in Deuterium: An Atomic Beam Measurement of Muon-Catalyzed *d-t* Fusion," *Phys. Rev. Lett.* **85**, 1642 (2000).

[22] M. C. Struensee, et al., "Nuclear Effects on the Eigenvalues of the $dt\mu$ Molecule," *Phys. Rev. A* **37**, 340 (1988).

[23] C.-Y. Hu, G. M. Hale, and J. S. Cohen, "Variational Calculations for the $dt\mu$ Molecule including Nuclear Effects on Sticking by Means of the Bloch Operator," *Phys. Rev. A* **49**, 4481 (1994).

[24] J. S. Cohen, "Kinetics of Muonic Helium in Muon-Catalyzed *d-d* and *d-t* Fusion," *Phys. Rev. Lett.* **58**, 1407 (1987).

[25] G. M. Hale, "Nuclear Physics of the Muon-Catalyzed *d+d* Reactions," *Muon Catal. Fusion* **5**, 227 (1990/91).

Peter W. Milonni, "Chaos in Laser-Matter Interactions"

[1] P. W. Milonni, J. R. Ackerhalt, and H. W. Galbraith, "Chaos in the Semiclassical N-Atom Jaynes-Cummings Model: Failure of the Rotating-Wave Approximation," *Phys. Rev. Lett.* **50**, 966 (1983).

[2] J. R. Ackerhalt, H. W. Galbraith, and P. W. Milonni, "Chaos in Multiple-Photon Excitation of Molecules," *Phys. Rev. Lett.* **51**, 1259 (1983).

[3] A. L. Gaeta, et al., "Instabilities and Chaos in the Polarization of Counterpropagating Light Fields," *Phys. Rev. Lett.* **58**, 2432 (1987).

Joseph Abdallah, "The Los Alamos Suite of Atomic Physics and Kinetics Codes"

[1] J. Abdallah, Jr., and R. E. H. Clark, "X-ray Transmission Calculations for an Aluminum Plasma," *J. Appl. Phys.* **69**, 23 (1991).

[2] J. Abdallah, Jr., R. E. H. Clark, and J. M. Peek, "Excited-State 1s-2p Absorption by Aluminum Ions with Partially Filled L Shells," *Phys. Rev. A* **44**, 4072 (1991).

References

[3] T. S. Perry, et al., "Opacity Measurements in Hot, Dense Medium," *Phys. Rev. Lett.* **67**, 3784 (1991).
[4] J. Abdallah, Jr., et al., "Time-Dependent Kinetics Model for a Helium Discharge Plasma," *J. Phys. B* **32**, 1001 (1999).
[5] T. Auguste, et al., "About the Role of Prepulse in the Heating of Clusters by High-Intensive Femtosecond Laser Pulse," *JETP Lett.* **72**, 55 (2000).
[6] J. Abdallah, Jr., et al., "Hot Electron Influence on the X-ray Emission Spectra of Ar Clusters Heated by a High-Intensity 60-fs Laser Pulse," *Phys. Rev. A* **63**, 032706 (2001).
[7] G. C. Junkel-Vives, et al., "High-Resolution X-ray Spectroscopy Investigations of fs Laser-Irradiated Ar Clusters by Varying Cluster Size and Laser Flux Density," *J. Quant. Spectros. Radiat. Transfer* **71**, 417 (2001).
[8] G. C. Junkel-Vives, et al., "Observation of H-Like Ions within Ar Clusters Irradiated by 35-fs Laser via High-Resolution X-ray Spectroscopy," *Phys. Rev. A* **64**, 021,201 (2001).
[9] G. C. Junkel-Vives, et al., "Spatially Resolved X-ray Spectroscopy Investigations of Femtsecond Laser-Irradiated Ar Clusters, " *Phys. Rev. E* (March 2002).
[10] G. C. Junkel-Vives, et al., "Supercritical Density Measurements of Laser-Irradiated Ar Clusters," Los Alamos National Laboratory document (submitted to *Phys. Rev. A*).

George Csanak, "Electron-Ion Collision Studies in Theoretical Division"

[1] J. J. Keady, "Opacity Efforts," *Theory in Action: Highlights in the Theoretical Division at Los Alamos, 1943–2003*, F. H. Harlow and H. J. Shepard, Eds., Los Alamos National Laboratory document LA-14000-H (April 2003), p. 59.
[2] B. E. Clements, J. D. Johnson, and D. C. Wallace, "Equation of State Research and the SESAME Library," *Theory in Action: Highlights in the Theoretical Division at Los Alamos, 1943–2003*, F. H. Harlow and H. J. Shepard, Eds., Los Alamos National Laboratory document LA-14000-H (April 2003), p. 319.
[3] W. F. Huebner, J. J. Keady, and S. P. Lyon, "Solar Photo Rates for Planetary Atmospheres and Atmospheric Pollutants," *Astrophys. Space Science* **195**, 1 (1992).
[4] R. D. Cowan, "The Theory of Atomic Structure and Spectra," *Theory in Action: Highlights in the Theoretical Division at Los Alamos, 1943–2003*, F. H. Harlow and H. J. Shepard, Eds., Los Alamos National Laboratory document LA-14000-H (April 2003), p. 81.
[5] A. L. Merts, R. D. Cowan, and N. H. Magee, Jr., "The Calculated Power Output from a Thin Iron-Seeded Plasma," Los Alamos Scientific Laboratory report LA-6220-MS (February 1976).
[6] R. D. Cowan, "Spectra of Highly Ionized Atoms of Tokamak Interest," Los Alamos Scientific Laboratory report LA-6679-MS (January 1977).

Robert D. Cowan, "The Theory of Atomic Structure and Spectra"

[1] R. D. Cowan and N. J. Peacock, "Identification of Some Intense Iron Lines in the Solar Spectrum near 170 Angstroms," *Astrophys. J.* **142**, 390 (1965); "Erratum," *Astrophys. J.* **143**, 283 (1966).
[2] R. D. Cowan, "Theoretical Calculation of Atomic Spectra Using Digital Computers," *J. Opt. Soc. Am.* **58**, 808 (1968).
[3] E. Källne, J. Källne, and R. D. Cowan, "High-Resolution X-ray Spectra from Molybdenum Ions in the Alcator C Tokamak," *Phys. Rev. A* **27**, 2682 (1983).
[4] C. D. Caldwell, et al., "Innershell Photoexcitation in an Open-Shell Atom: The Cl 2p–>ns,md Spectrum as a Case Study," *Phys. Rev. A* **59**, R926 (1999).
[5] A. L. Merts, R. D. Cowan, and N. H. Magee, Jr., "The Calculated Power Output from a Thin Iron-Seeded Plasma," Los Alamos Scientific Laboratory report LA-6220-MS (March 1976).
[6] T. D. Cowan and J. B. Mann, "The Atomic Structure of Super-Heavy Elements," *Atomic Physics 2*, P. G. H. Sandars, Ed. (Plenum Press, London, 1971), p. 215.

Biology

Angel E. Garcia, "Protein-Folding Dynamics"

[1] C. B. Anfinsen, et al., "The Kinetics of Formation of Native Ribonuclease during Oxidation of the Reduced Polypeptide Chain," *Proc. Natl. Acad. Sci. U.S.A.* **47**, 1309 (1961).
[2] C. Levinthal, "Are There Pathways for Protein Folding?" *J. Chem. Phys.* **65**, 44 (1968).
[3] P. S. Kim and R. L. Baldwin, "Specific Intermediates in the Folding Reactions of Small Proteins and the Mechanism of Protein Folding," *Annu. Rev. Biochem*. **59**, 631 (1990).
[4] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, "The Energy Landscapes and Motions of Proteins," *Science* **254**, 1598 (1991).
[5] P. E. Leopold, M. Montal, and J. N. Onuchic, "Protein-Folding Funnels: Kinetic Pathways through Compact Conformational Space," *Proc. Natl Acad. Sci. U.S.A.* **89**, 8721 (1992).
[6] J. N. Onuchic, et al., "The Energy Landscape Theory of Protein Folding: Insights into Folding Mechanisms and Scenarios," *Adv. Protein Chem.* **87** (2000).
[7] J. N. Onuchic, H. Nymeyer, and A. E. Garcia, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5921 (May 1998).
[8] A. F. Voter, "Parallel Replica Method for Dynamics of Infrequent Events," *Phys. Rev. B*, **57**, 13,985 (1998).
[9] E. Marinari and G. Parisi, "Simulated Tempering: A New Monte Carlo Scheme," *Europhys. Lett.*, **19**, 451(1992).
[10] A. E. Garcia and K. Y. Sanbonmatsu, "Alpha Helical Stabilization by Side Chain Shielding of Backbone Hydrogen Bonds," *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2782 (2002).

Alan Lapedes, "The Antigenic Evolution of Influenza"

[1] A. Lapedes and R. Farber, "The Geometry of Shape Space: Application to Influenza," *J. Theoretical Biology* **212**, 57 (2001).
[2] D. J. Smith, et al., "Modeling the Effects of Updating the Influenza Vaccine on the Efficacy of Repeated Vaccination," in *Options for the Control of Influenza IV, ADME*, E. Osterhaus, N. Cox, and A. Hampson, Eds., *Excerpta Medica* (International Congress Series 1219, Amsterdam, 2001), p. 655.
[3] D. J. Smith, et al., "Thirty-two Years of Antigenic and Genetic Evolution of Influenza A (H3N2) Virus," (submitted, *Science*, 2002).

Christian Burks and Linda K. Wood, "GenBank—From Idea to Worldwide Indispensability"

[1] A. J. Burke, "GenBank Turns 20," *Genome Technology*, p. 44 (October 2002).
[2] C. Burks, et al., "The GenBank Nucleic Acid Sequence Database," *Comput. Appl. Biosci.* **1**, 225 (December 1985).
[3] P. Sherrid, "Christian Burks: Slicing and Dicing the Facts of Life," *U.S. News and World Report*, **129**, 58 (December 24, 2000).
[4] M. J. Cinkosky, et al., "Electronic Data Publishing and GenBank," *Science* **252**, 1273 (1991).
[5] National Center for Biotechnology Information, "GenBank Overview," http://www.ncbi.nlm.nih.gov/Genbank/GenbankOverview.html (September 24, 2002).
[6] The University of Texas, "GenBank," *BioTech Life Sciences Resources and Reference Tools*, http.//biotech.icmb.utexas.edu/pages/science/genbank/genbank.html (July 23, 1998).
[7] Los Alamos National Laboratory, "Walter Goad, GenBank Founder, Dies," *Daily NEWSBulletin* [Online], http://www.lanl.gov/orgs/pa/News/112100.html (November 21, 2000).
[8] J. McLaughlin, "Smithsonian Winners Announced: Computerworld Smithsonian Awards Given to Three Sun Users," *The Florida SunFlash* [Online], http://sunsite.lanet.lv/ftp/sun-info/sunflash/1992/Jun/42.18.smith (June 15, 1992).
[9] G. H. Anthes, "Smithsonian Awards: 10 Years of Heroes," *Computerworld* [Online], http://www/computerworld.com/news/1998/story/0,11280,31280,00.html (June 8, 1998).

References

Bette Korber and Tanmoy Bhattacharya, "HIV Evolution and Its Implications for Understanding Viral Origins"

[1] T. Zhu, et al., "An African HIV-1 Sequence from 1959 and Implications for the Origin of the Epidemic," *Nature* **391**, 594 (1998).
[2] B. T. Korber, J. Theiler, and S. Wolinsky, "Limitations of a Molecular Clock Applied to Considerations of the Origins of HIV-1," *Science* **280**, 1868 (1998).
[3] B. T. Korber, et al., "Timing the Ancestor of the HIV-1 Pandemic Strains," *Science* **288**, 1789 (2000).
[4] B. T. Korber, et al., "Search for the Origin of HIV and AIDS," *Science* **289**, 1140 (2000).
[5] K. Yusim, et al., "Using HIV-1 Sequences to Infer Historical Features of the AIDS Epidemic and HIV Evolution," *The Philosophical Transactions of the Royal Society, series B*, **356**, 855 (2001).
[6] B. Gaschen, et al., "Diversity Considerations in HIV-1 Vaccine Selection," *Science* **296**, 2354 (2002).
[7] F. Gao, et al., "Central Derived Sequences in HIV Vaccine Design," (submitted, *Science*, August 2002).

Alan S. Perelson, "Modeling HIV Infection and the Introduction of Combination Therapy"

[1] D. D. Ho, et al., "Rapid Turnover of Plasma Virions and CD4 Lymphocytes in HIV-1 Infection," *Nature* **373**, 123 (1995).
[2] A. S. Perelson, P. Essunger, and D. D. Ho, "Dynamics of HIV-1 and CD4+ Lymphocytes In Vivo," *Aids* **11** (Suppl. A), S17 (1997).
[3] A. S. Perelson, et al., "HIV-1 Dynamics In Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time," *Science* **271**, 1582 (1996).
[4] A. S. Perelson, et al., "Decay Characteristics of HIV-1-Infected Compartments during Combination Therapy," *Nature* **387**, 188 (1997).
[5] A. U. Neumann, et al., "Hepatitis C Viral Dynamics In Vivo and the Antiviral Efficacy of Interferon-Alpha Therapy," *Science* **282**, 103 (1998).
[6] S. R. Lewin, et al., "Analysis of Hepatitis B Viral Load Decline under Potent Therapy: Complex Decay Profiles Observed," *Hepatology* **34**, 1012 (2001).

Geoffrey B. West, "The Origin of Universal Scaling Laws in Biology from Molecules and Cells to Whales and Ecosystems"

[1] See, for example, K. Schmidt-Nielsen, *Scaling: Why is Animal Size So Important* (Cambridge University Press, 1984); W. A. Calder, III, *Size, Function and Life History* (Harvard University Press, Cambridge, Massachusetts, 1984); R. H. Peters, *The Ecological Implications of Body Size* (Cambridge University Press, Cambridge, United Kingdom, 1983); K. J. Niklas, *Plant Allometry: The Scaling of Form and Process* (University of Chicago Press, Chicago, 1994).
[2] G. B. West, W. H. Woodruff, and J. H. Brown, "Allometric Scaling of the Metabolism from Molecules and Mitochondria to Cells and Mammals," *Proc. Nat. Acad. Sc.* **99**, 2473 (2002).
[3] G. B. West, J. H. Brown, and B. J. Enquist, "A General Model for the Origin of Allometric Scaling Laws in Biology," *Science* **276**, 122 (1997).
[4] G. B. West, J. H. Brown, and B. J. Enquist, "A General Model for the Structure and Allometry of Plant Vascular Systems," *Nature* **398**, 573 (1999).
[5] G. B. West, J. H. Brown, and B. J. Enquist, "A General Model for Ontogenetic Growth," accepted for publication in *Nature*, October 2001.
[6] B. J. Enquist, et al., "Allometric Scaling of Production and Life-History Variation in Vascular Plants," *Nature* **401**, 907 (1999).
[7] G. B. West, J. H. Brown, and B. J. Enquist, "The Fourth Dimension of Life: Fractal Geometry and Allometric Scaling of Organisms," *Science* **284**, 1677 (1999).
[8] J. F. Gillooly, et al., "Effects of Size and Temperature on Metabolic Rate," *Science* 293, 2248 (2001).
[9] J. F. Gillooly, et al., "Biological Time: Effects of Size and Temperature on Development," *Nature* **417**, 70 (2002).

Chemistry

P. Jeffrey Hay and Willard R. Wadt, "Extending Quantum Chemistry through the Periodic Table"

[1] L. R. Kahn, P. J. Hay, and R. D. Cowan, "Relativistic Effects in *Ab Initio* Effective Core Potentials for Molecular Calculations. Application to the Uranium Atom," *J. Chem. Phys.* **68**, 2368 (1978).

[2] P. J. Hay, et al., "*Ab Initio* Studies of the Electronic Structure of UF_6, UF_6^+ and UF_6^- Using Relativistic Effective Core Potentials," *J. Chem. Phys.* **70**, 1767 (1979).

[3] W. R. Wadt and P. J. Hay, "*Ab Initio* Studies of the Electronic Structure and Geometry of UF_5 Using Relativistic Effective Core Potentials," *J. Amer. Chem. Soc.* **101**, 5198 (1979).

[4] P. J. Hay, "*Ab Initio* Studies of Excited States of Polyatomic Molecules including Spin-Orbit and Multiplet Effects: The Electronic States of UF_6," *J. Chem. Phys.* **79**, 5469 (1983).

[5] W. R. Wadt, "An *Ab Initio* Study of Low-Lying 5f-5f Excitations in PuF_6," *J. Chem. Phys.* **86**, 339 (1987).

[6] P. J. Hay, "Electronic States of the Quadruply Bonded $[Re_2Cl_8]^{2-}$. An *Ab Initio* Theoretical Study," *J. Am. Chem. Soc.* **104**, 7007 (1982).

[7] R. L. Martin and P. J. Hay, "Relativistic Contributions to the Low-Lying Excitation Energies and Ionization Potentials of the Transition Metals," *J. Chem. Phys.* **75**, 4539 (1981).

[8] P. J. Hay and W. R. Wadt, "*Ab Initio* Effective Core Potentials for Molecular Calculations. Potentials for the Transition Metal Atoms Sc to Hg," *J. Chem. Phys.* **82**, 270 (1985).

[9] W. R. Wadt and P. J. Hay, "*Ab Initio* Effective Core Potentials for Molecular Calculations. Potentials for the Main Group Elements Na to Bi," *J. Chem. Phys.* **82**, 284 (1985).

[10] P. J. Hay and W. R. Wadt, "*Ab Initio* Effective Core Potentials for Molecular Calculations: Potentials for K to Au including the Outermost Core Orbitals," *J. Chem. Phys.* **82**, 299 (1985).

[11] G. Schreckenbach, P. J. Hay, and R. L. Martin, "Theoretical Study of Stable Trans and Cis Isomers in $[UO_2(OH)_4]^{2-}$ Using Relativistic Density Functional Theory," *Inorganic Chemistry* **37**, 4442 (1998).

[12] Donald E. Stokes, *Pasteur's Quadrant: Basic Science and Technological Innovation* (The Brookings Institute Press, Washington, D.C., 1997).

Russell T Pack, "Body-Frame Axes and Angular Momentum Decoupling in Molecular Collisions"

[1] M. Born and J. R. Oppenheimer, *Ann. Phys.* (*Leipzig*) **84**, 457 (1927).

[2] T. P. Tsien and R. T Pack, *Chem. Phys. Lett.* **6**, 54 and 400 (1970) and **8**, 579 (1971).

[3] R. T Pack, *Chem. Phys. Lett.* **14**, 393 (1972).

[4] K. Takayanagi, *Prog. Theor. Phys.* (Kyoto) Suppl. **25**, 40 (1963).

[5] W. A. Lester, Jr., and R. B. Bernstein, *Chem. Phys. Lett.* **1**, 297 and 347 (1967).

[6] T. P. Tsien, G. A. Parker, and R. T Pack, *J. Chem. Phys.* **59**, 5373 (1973).

[7] A. M. Arthurs and A. Dalgarno, *Proc. R. Soc. London, Ser. A* **256**, 540 (1960).

[8] C. F. Curtiss and F. T. Adler, *J. Chem. Phys.* **20**, 249 (1952).

[9] C. F. Curtiss, ibid. **21**, 2045 (1953).

[10] G. Gioumousis and C. F. Curtiss, *J. Math. Phys.* **2**, 96 (1961).

[11] C. F. Curtiss, *J. Chem. Phys.* **49**, 1952 (1968).

[12] L. W. Hunter and C. F. Curtiss, ibid. **58**, 3884 (1973).

[13] R. T Pack, *J. Chem. Phys.* **60**, 633 (1974).

[14] P. McGuire and D. J. Kouri, *J. Chem. Phys.* **60**, 2488 (1974).

[15] For a review, see D. J. Kouri, in *Atom-Molecule Collision Theory: A Guide for the Experimentalist,* R. B. Bernstein, Ed. (Kluwer Academic/Plenum Publishers, New York, 1979), p. 301.

[16] D. Secrest, *J. Chem. Phys.* **62**, 710 (1975).

[17] L. W. Hunter, *J. Chem. Phys.* **62**, 2855 (1975).

[18] G. A. Parker and R. T Pack, *J. Chem. Phys.* **68**, 1585 (1978).

References

[19] A. Kuppermann and G. C. Schatz, *J. Chem. Phys.* **62**, 2502 (1975).
[20] A. B. Elkowitz and R. E. Wyatt, *J. Chem. Phys.* **62**, 2504 (1975).
[21] F. T. Smith, *J. Math. Phys.* **3**, 735 (1962).
[22] R. C. Whitten and F. T. Smith, ibid. **9**, 1103 (1968).
[23] R. T Pack, *Chem. Phys. Lett.* **108**, 333 (1984).
[24] G. A. Parker, et al., *Chem. Phys. Lett.* **137**, 564 (1987).
[25] R. T Pack and G. A. Parker, *J. Chem. Phys.* **87**, 3888 (1987).
[26] R. T Pack and G. A. Parker, ibid. **90**, 3511 (1989).
[27] R. T Pack, in *Advances in Molecular Vibrations and Collision Dynamics,* J. M. Bowman, Ed. (JAI Press, Greenwich, Connecticut, 1994), Vol. IIA, p. 111.
[28] B. K. Kendrick and R. T Pack, *J. Chem. Phys.* **104**, 7475 and 7502 (1996).
[29] B. K. Kendrick, et al., *J. Chem. Phys.* **110**, 6673 (1999).
[30] B. K. Kendrick, *J. Chem. Phys.* **112**, 5679 (2000).
[31] B. K. Kendrick, ibid. **114**, 4335 (2001).
[32] B. K. Kendrick, ibid. **114**, 8796 (2001).
[33] F. D. Colavecchia, et al., *J. Chem. Phys.*, submitted for publication (2003).
[34] G. A. Parker, et al., *J. Chem. Phys.* **117**, 6083 (2002).

Dimitri Babikov and Joel D. Kress, "Semiclassical Dynamics—Insight into Physics of Complex Systems"

[1] E. J. Heller, "Quantum Dynamics of Localized Wavepackets on Smooth Potential Energy Surfaces," *J. Chem. Phys.* **62**, 1544 (1974).

Computers and Computing

Francis H. Harlow, "Computers"

[1] R. B. Lazarus, et al., "Computing at LASL in the 1940s and 1950s," Los Alamos Scientific Laboratory report LA-6945-H (May 1978).
[2] D. W. Forslund, C. A. Slocomb, and I. A. Agins, "Windows on Computing, *Los Alamos Science*, Number 22, p. 4 (1994).

Paul Ginsparg, "e-Print arXiv Project"

[1] S. Loken, et al., *Bull. Am. Phys. Soc.* **36**, 1119 (1991). http://publish.aps.org/eprint/reports/lokenrep.html
[2] J. Langer, *Physics Today* **53**, 35 (2000). http://www.aip.org/pt/vol-53/iss-8/p35.html
[3] Adrian Cho, *Science* **287**, 1899 (March 2000). http://www.sciencemag.org/cgi/content/full/287/5460/1899b
[4] http://www.openarchives.org/

Concentrated Energy

J. N. Johnson, "The JTF Model for Shock Initiation of High Explosives"

[1] J. N. Johnson, P. K. Tang, and C. A. Forest, "Shock-Wave Initiation of Heterogeneous Reactive Solids," *J. Appl. Phys.* **57**, 4323 (1985).
[2] C. L. Mader and C. A. Forest, "Two-Dimensional Homogeneous and Heterogeneous Detonation Wave Propagation," Los Alamos Scientific Laboratory report LA-6259 (1976).
[3] T. C. Germann, et al., "Molecular Dynamics Studies of Detonation in Defective Explosive Crystals," in *Proceedings of the 12th International Symposium on Detonation,* San Diego, California, August 11–16, 2002 (to be published).
[4] J. N. Johnson, "Micromechanical Considerations in Shock Compression of Solids," in *High-Pressure Shock Compression of Solids*, J. R. Asay and M. Shahinpoor, Eds. (Springer-Verlag New York, New York, 1992), p. 217.

David E. Hanson, "Inertial Confinement Fusion Program"

[1] D. Eimerl, J. M. Auerbach, and P. W. Milonni, "Paraxial Wave Theory of Second and Third Harmonic Generation in Uniaxial Crystals. I. Narrow-Band Fields," *J. Mod Optic* **42**, 1037 (1995).

[2] P. W. Milonni, J. M. Auerbach, and D. Eimerl, "Frequency Conversion Modeling with Spatially and Temporally Varying Beams," in *Solid-State Lasers for Application to Inertial Confinement Fusion*, The International Society for Optical Engineering document SPIE 2633, M. André and H. T. Powell, Eds. (SPIE Press, Bellingham, Washington, 1995), p. 230.

[3] J. M. Auerbach, et al., "Frequency—Conversion Modeling," *Inertial Confinement Fusion Quarterly Report 6* (1996), p. 199.

[4] J. M. Auerbach, et al., "Perturbation Theory for Electric Field Amplitude and Phase Ripple Transfer in Frequency Doubling and Tripling," *Appl. Optics* **36**, 606 (1997).

[5] J. M. Auerbach, et al., "Perturbation Theory for Frequency Doubling and Tripling of Electric Field Amplitude and Phase Ripples," *Solid-State Lasers for Application to Inertial Confinement Fusion*, The International Society for Optical Engineering document SPIE 2633, M. André and H. T. Powell, Eds. (SPIE Press, Bellingham, Washington, 1995), p. 655.

[6] J. M. Auerbach, et al., "Alternate Frequency Tripling Schemes," The International Society for Optical Engineering document SPIE 3047 (SPIE Press, Bellingham, Washington, 1997), p. 853.

[7] D. Eimerl, et al., "Multicrystal Designs for Efficient Third Harmonic Generation," *Opt. Lett.* **22**, 1208 (1997).

Condensed Matter

Brad Lee Holian, Timothy C. Germann, and Peter S. Lomdahl, "Molecular Dynamics Simulations"

[1] A. F. Voter, F. Montalenti, and T. C. Germann, *Annu. Rev. Mater. Res.* **32**, 321 (2002).

[2] B. J. Alder and T. E. Wainwright, *J. Chem. Phys.* **31**, 459 (1959).

[3] B. J. Alder and T. E. Wainwright, *J. Chem. Phys.* **27**, 1208 (1957).

[4] W. W. Wood and J. D. Jacobson, *J. Chem. Phys.* **27**, 1207 (1957).

[5] W. G. Hoover and W. T. Ashurst, "Nonequilibrium Molecular Dynamics," in *Theoretical Chemistry: Advances and Perspectives*, H. Eyring and D. Henderson, Eds. (Kluwer Academic/Plenum Publishers, New York, 1975), p. 1.

[6] B. L. Holian and D. J. Evans, *J. Chem. Phys.* **78**, 5147 (1983).

[7] B. L. Holian, *J. Chem. Phys.* **117**, 1173 (2002), and references therein.

[8] W. G. Hoover and W. T. Ashurst, "Nonequilibrium Molecular Dynamics," in *Theoretical Chemistry: Advances and Perspectives*, H. Eyring and D. Henderson, Eds. (Kluwer Academic/Plenum Publishers, New York, 1975), p. 25.

[9] J. J. Erpenbeck and W. W. Wood, *Phys. Rev. A* **26**, 1648 (1982).

[10] B. L. Holian and G. K. Straub, *Phys. Rev. B* **18**, 1593 (1978).

[11] B. L. Holian and G. K. Straub, *Phys. Rev. Lett.* **43**, 1598 (1979).

[12] B. L. Holian, et al., *Phys. Rev. A* **22**, 2498 (1980).

[13] W. C. Swope and H. C. Andersen, *Phys. Rev. B* **41**, 7042 (1990).

[14] W. G. Hoover, et al., *Phys. Rev. A* **42**, 5844 (1990).

[15] N. J. Wagner, B. L. Holian, and A. F. Voter, *Phys. Rev. A* **45**, 8457 (1992).

[16] D. M. Beazley and P. S. Lomdahl, *Parallel Comput.* **20**, 173 (1994).

[17] P. S. Lomdahl, et al., *Proc. of Supercomputing 93* (Institute of Electrical and Electronics Engineers, Inc., Computer Society Press, Los Alamitos, California, 1993), p. 520.

[18] D. M. Beazley and P. S. Lomdahl, *Comput. Phys.* **11** (3), 230 (1997).

[19] S. J. Zhou, et al., *Science* **279**, 1525 (1998).

[20] S. J. Zhou, et al., *Engin. Fract. Mech.* **61**, 173 (1998).

[21] B. L. Holian, J. E. Hammerberg, and P. S. Lomdahl, *J. Comput.-Aided Mater.* **5**, 207 (1998).

[22] B. L. Holian and P. S. Lomdahl, *Science* **280**, 2085 (1998).

[23] K. Kadau, et al., *Science* **296**, 1681 (2002).

[24] T. C. Germann, et al., "Molecular Dynamics Simulations of Detonation in Defective Explosive Crystals," in *Proceedings of the 12th International Symposium on Detonation,* San Diego, California, August 11–16, 2002 (to be published).

Lee A. Collins, Joel D. Kress, and Stephane F. Mazevet, "Simulations of Warm, Dense Matter"

[1] L. A. Collins and A. L. Merts, "Electronic Structure of Clusters of Atoms in a Dense Plasma," in *Radiative Properties of Hot Dense Matter*, Eds. J. Davis, C. Hooper, R. Lee, A. Merts, and B. Rozsnyai (World Scientific: Singapore, 1985), p. 385.
[2] L. Collins, et al., "Quantum Molecular Dynamics Simulations of Hydrogenic Systems," *Phys. Rev. E* **52**, 6202 (1995).
[3] M. Desjarlais, J. Kress, and L. Collins, "Electrical Conductivities for Warm, Dense Aluminum Plasmas and Fluids," *Phys, Rev. E* **66**, 025401 (2002).
[4] I. Kwon, et al., "First Principles Study of Solid Ar and Kr at High Compression," *Phys. Rev. B* **52**, 15165 (1995).
[5] S. Bickham, et al., "Molecular Dynamics Simulations of Expanded, Liquid Sodium," *Phys. Rev. B* **58**, R11813 (1998).
[6] J. Kress, I. Kwon, and L. Collins, "Simulation of Impurity Line Shapes in a Hot, Dense Plasma," *J. Quant. Spectrosc. Rad. Transp*. **54**, 237 (1995)**.**
[7] T. Lenosky, et al., "Density-Functional Calculation of the Hugoniot of Shocked Liquid Hydrogen," *Phys. Rev. B* **61**, 1 (2000); B. Militzer, et al., "Calcualtion of Deuterium Double Shock Hugoniot from *Ab Initio* Simulations," *Phys. Rev. Lett.* **87**, 275502 (2001).
[8] S. Mazevet, et al., "Density Functional Calculation of Multiple-Shock Hugoniots of Liquid Nitrogen," *Phys. Rev. B* **65**, 014204 (2002); S. Mazevet, et al., "Quantum Molecular Dynamics Study of the Electrical and Optical Properties of Shocked Liquid Nitrogen," *Phys. Rev. B* **67**, 054201 (2003).
[9] J. Kress, et al., "Tight-binding Molecular Dynamics of Shock Waves in Methane," *Phys. Rev. Lett.* **83**, 3896 (1999).
[10] S. Bickham, et al., "*Ab Initio* Molecular Dynamics Studies of Off-Center Displacements in CuCl," *Phys. Rev. Lett.* **83**, 568 (1999).
[11] S. Mazevet, L. Collins, and J. Kress, "Evolution of Ultracold, Neutral Plasmas," *Phys. Rev. Lett.* **88**, 055001 (2002).

Lawrence R. Pratt, "Modeling and Molecular Theory of Liquids"

[1] W. W. Wood, "Early History of Computer Simulations in Statistical Mechanics," in *International School of Physics "Enrico Fermi"* (Societa Italiana di Fisica, Bologna, Italy, 1986), p. 3.
[2] W. W. Wood, "Some Additional Recollections, and the Absence Thereof, about the Early History of Computer Simulations in Statistical Mechanics," in *Euroconference on Computer Simulation in Condensed Matter Physics and Chemistry*, Vol. 49 (Societa Italiana di Fisica, Bologna, Italy, 1996), p. 908.
[3] M. S. Wertheim, "Exact Solution of Percus-Yevick Integral Equation for Hard Spheres," *Phys. Rev. Lett.* **10**, 321 (1963).
[4] E. Thiele, "Equation of State for Hard Spheres," *J. Chem. Phys.* **39**, 474 (1963).
[5] L. R. Pratt and S. B. Rempe, *Simulation and Theory of Electrostatic Interactions in Solution, AIP Conference Proceedings,* **492**, 172 (1999).
[6] M. E. Paulaitis and L. R. Pratt, "Hydration Theory for Molecular Biophysics," *Adv. Protein Chem.* **62**, 283 (2002).

Arthur F. Voter, "Accelerated Molecular Dynamics Methods"

[1] A. F. Voter, F. Montalenti, and T. C. Germann, "Extending the Time Scale in Atomistic Simulation of Materials," *Annu. Rev. Mater. Res.*, **32**, 321 (2002).
[2] A. F. Voter, "A Method for Accelerating the Molecular Dynamics Simulation of Infrequent Events," *J. Chem. Phys.* **106**, 4665 (1997).
[3] A. F. Voter, "Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events," *Phys. Rev. Lett.*, **78**, 3908 (1997).
[4] A. F. Voter, "Parallel Replica Method for Dynamics of Infrequent Events," *Phys. Rev. B* **57**, 13,985 (1998).

[5] M. R. Sorensen and A. F. Voter, "Temperature-Accelerated Dynamics for Simulation of Infrequent Events," *J. Chem. Phys.* **112**, 9599 (2000).
[6] W. F. Egelhoff, Jr., and I. Jacob, "Reflection High-Energy Electron-Diffraction (RHEED) Oscillations at 77 K," *Phys. Rev. Lett.* **62**, 921 (1989).

Cristian D. Batista and Gerardo Ortiz, "Unifying the Quantum Description of Matter"

[1] C. D. Batista and G. Ortiz, *Phys. Rev. Lett.* **86**, 1082 (2001).
[2] C. D. Batista and G. Ortiz, *Phys. Rev. B* **65**, 180, 402 (2002).
[3] G. Ortiz and C. D. Batista, cond-mat/0207077 at http://arXiv.org.
[4] G. Ortiz and C. D. Batista in *Recent Progress in Many-Body Theories*, Raymond F. Bishop, Tobias Brandes, Klaus A. Gernoth, Niels R. Walet, and Yang Xian, Eds. (World Scientific, Singapore, 2002).
[5] C. D. Batista and G. Ortiz, cond-mat/0207106 at http://arXiv.org.
[6] G. Ortiz, et al., *Phys. Rev. A* **64**, 22319 (2001); R. Somma, et al., *Phys. Rev. A* **65**, 42323 (2002).
[7] C. D. Batista, *Phys. Rev. Lett.* **89**, 166403 (2002).

Environmental Studies

Francis H. Harlow and William E. Pracht, "A Theoretical Study of Geothermal Energy Extraction"

[1] Francis H. Harlow and William E. Pracht, "A Theoretical Study of Geothermal Energy Extration," *J. Geophys. Res.* **77**, 7038 (1972).

Mathew E. Maltrud, Elizabeth C. Hunke, Philip W. Jones, Richard D. Smith, John K. Dukowicz, and William H. Lipscomb, "Climate, Ocean, and Sea Ice Modeling"

[1] S. W. Matthews, "What's Happening to Our Climate?" *National Geographic* **150**, 576 (November 1976).
[2] J. K. Dukowicz, R. D. Smith, and R. C. Malone, "A Reformulation and Implementation of the Bryan-Cox-Semtner Ocean Model on the Connection Machine," *J. Atmos. Ocean. Tech.*, **10**, 195 (1993).
[3] E. C. Hunke and J. K. Dukowicz, "An Elastic-Viscous-Plastic Model for Sea Ice Dynamics," *J. Phys. Oceanogr.* **27**, 1849 (1997).
[4] G. H. Denton, "Does an Asymmetric Thermohaline-Ice-Sheet Oscillator Drive 100,000-Year Glacial Cycles?" *J. Quaternary Sci* **15**, 301 (May 2000).

T. Daniel Butler, Larry Cook, Frank H. Harlow, C. W. Hirt, and Robert S. Hotchkiss, "Air Pollution Transport in Street Canyons"

[1] R. S. Hotchkiss and F. H. Harlow, "Air Pollution Transport in Street Canyons," Office of Research and Monitoring, U.S. Environmental Protection Agency report EPA-R4-73-029 (June 1973).
[2] C. W. Hirt and J. L. Cook, "Calculating Three-Dimensional Flows around Structures and over Rough Terrain," Los Alamos Scientific Laboratory report LA-DC-13289 (November 12, 1971).

Fluid Dynamics

Darryl D. Holm, "Navier-Stokes-alpha Model of Fluid Turbulence"

[1] C. Foias, D. D. Holm, and E. S. Titi, *Physica D* **152**, 505 (2001); C. Foias, D. D. Holm, and E. S. Titi, *J. Dyn. and Diff. Eqns*. **14,** 1 (2002).
[2] D. D. Holm, J. E. Marsden, and T. S. Ratiu, *Adv. in Math*. **137**, 1 (1998); D. D. Holm, J. E. Marsden, and T. S. Ratiu, *Phys.Rev. Lett*. **80**, 4173 (1998).
[3] D. G. Andrews and M. E. McIntyre, *J. Fluid Mech*. **89,** 609 (1978).
[4] G. I. Taylor, *Proc. R. Soc. London Ser. A* **164,** 476 (1938).
[5] S. Y. Chen, et al., *Physica D* **133,** 66 (1999).
[6] M. V. Zagarola, "Mean-Flow Scaling of Turbulent Pipe Flow," Ph.D. Thesis, Princeton University, 1996.
[7] S. Y. Chen, et al., *Phys. Rev. Lett.* **81**, 5338 (1998); S. Y. Chen, et al., *Physica D* **133**, 49 (1999); S. Y. Chen, et al., *Phys. Fluids* **11**, 2343 (1999).

References

[8] B. J. Geurts and D. D. Holm, "Alpha-Modeling Strategy for LES of Turbulent Mixing," in *Turbulent Flow Computation*, D. Drikakis and B. G. Geurts, Eds. (Kluwer, London, 2002), p. 237.

Industrial Collaborations

A. Amsden, P. J. O'Rourke, T. D. Butler, "KIVA: Working for Clean Air and Fuel Efficiency

[1] J. K. Dukowicz, "Particle-Fluid Numerical Model for Liquid Sprays," *J. Comput. Phys.* **35**, 229 (1980).
[2] P. J. O'Rourke, "Collective Drop Effects in Vaporizing Liquid Sprays," Ph.D. Thesis 1532-T, Princeton University (1981).
[3] A. A. Amsden, et al., "KIVA—A Comprehensive Model for 2-D and 3-D Engine Simulations," Society of Automotive Engineers Technical Paper 850554 (1985).
[4] A. A. Amsden, et al., "KIVA: A Computer Program for Two- and Three-Dimensional Fluid Flows with Chemical Reactions and Fuel Sprays," Los Alamos National Laboratory report LA-10245-MS (February 1985).
[5] A. A. Amsden, T. D. Butler, and P. J. O'Rourke, "The KIVA-II Computer Program for Transient Multidimensional Chemically Reactive Flows with Sprays," Society of Automotive Engineers technical paper 872072 (1987).
[6] A. A. Amsden, P. J. O'Rourke, and T. D. Butler, "KIVA-II: A Computer Program for Chemically Reactive Flows with Sprays," Los Alamos National Laboratory report LA-11560-MS (May 1989).
[7] A. A. Amsden, "KIVA-3: A KIVA Program with Block-Structured Mesh for Complex Geometries," Los Alamos National Laboratory report LA-12503-MS (March 1993).
[8] A. A. Amsden, "KIVA-3V: A Block-Structured KIVA Program for Engines with Vertical or Canted Valves," Los Alamos National Laboratory report LA-13313-MS (July 1997).

Material Properties

John M. Wills, "The Electronic Structure of the Actinides"

[1] E. A. Kmetko and J. T. Waber, "Plutonium," in *Proceedings of the 3rd International Conference* (Chapman & Hall, London, 1965).
[2] H. H. Hill and E. A. Kmetko, *Nucl. Met.* **17**, 223 (1970), W. N. Miner, Ed.
[3] J. M. Wills and O. Eriksson, *Phys. Rev. B* **45**, 13,879 (1992).
[4] P. Soderlind, et al., *Phys. Rev. B* **55**, 1997 (1997).
[5] L. Fast, et al., *Phys. Rev. B* **81**, 2978 (1998).
[6] P. Soderlind, et al., *Nature* **374**, 524 (1995).
[7] O. Eriksson, et al., *J. Alloy Compd.* **287**, 1 (1999).

B. E. Clements, J. D. Johnson, and D. C. Wallace, "Equations of State Research and the SESAME Library"

[1] R. P. Feynman, N. Metropolis, and E. Teller, "Equations of State of Elements Based on the Generalized Fermi-Thomas Theory," *Phys. Rev.* **75**, 1561 (1949).
[2] R. D. Cowan and J. Ashkin, "Extensions of the Thomas-Fermi-Dirac Statistical Theory of the Atom to Finite Temperature," *Phys. Rev.* **105**, 144 (1957).
[3] R. M. More, et al., "A New Quotidian Equation of State (QEOS) for Hot Dense Matter," *Phys. Fluids* **31**, 3059 (1988).
[4] J. Barnes, "Statistical Atom Theory and Equation of State of Solids," *Phys. Rev.* **153**, 269 (1967).
[5] G. Kerley, "Perturbation Theory and the Thermodynamic Properties of Fluids: 2. The CRIS Model," *J. Chem. Phys.* **73**, 478 (1980).
[6] D. A. Liberman, "Self-Consistent Field Model for Condensed Matter," *Phys. Rev. B* **20**, 4981 (1979).
[7] D. C. Wallace, *Statistical Physics of Crystals and Liquids: A Guide to Highly Accurate Equations of State* (World Scientific Publishing Company, Singapore, 2003).
[8] S. P. Lyon and J. D. Johnson, "T-1 Handbook of the SESAME Equation of State Library, Los Alamos National Laboratory document LA-CP-98-100 (1998).

[9] S. P. Lyon and J. D. Johnson, Eds., "SESAME: The Los Alamos National Laboratory Equation-of-State Database," Los Alamos National Laboratory document LA-UR-92–3407 (1992).
[10] K. S. Holian, Ed., "T-1 Handbook of Material Properties Data Bases, Vol. Ic: Equations of State," Los Alamos National Laboratory document LA-10160-MS (1984).
[11] Equation-of-State and Opacity Group, Theoretical Division, "SESAME '83: Report on the Los Alamos Equation-of-State Library," Los Alamos National Laboratory report LA-LP-83-4 (1983).

J. N. Johnson, "Void Growth and Fracture in Ductile Solids"

[1] A. K. Zurek, MST-8, Los Alamos National Laboratory, personal communication.
[2] M. M. Carroll and A. C. Holt, *J. Appl. Phys*. **43**, 1626 (1972).
[3] J. N. Johnson, "Dynamic Fracture and Spallation in Ductile Solids," *J. Appl. Phys*. **52**, 2812 (1981).
[4] J. N. Johnson and F. L. Addessio, "Tensile Plasticity and Ductile Fracture," *J. Appl. Phys*. **64**, 6699 (1988).

Mathematics Science

A. Patrascioiu, "The Ergodic Hypothesis: A Complicated Problem in Mathematics and Physics"

[1] J. S. Bell, "On the Einstein-Podolsky-Rosen Paradox," *Physics* **1**, 195 (1964). A simple discussion of Bell's inequality and its experimental verification can he found in "The Quantum Theory and Reality," by Bernard d'Espagnat, Sc*ientific American,* November 1979, p. 128.
[2] Alain Aspect, Philippe Grangier, and Gerard Roger, "Experimental Realization of Einstein-Podolsky-Rosen-Bohm *Gedankenexperiment*: A New Violation of Bell's Inequalities," *Phys. Rev. Lett.* **49**, 91 (1982).
[3] L. Boltzmann, "Studien über des Gleichgewicht der lebendigen Kraft zwichen bewegten materiellen Punkten," *Wiener Berichte* **58**, 517 (1868).
[4] L. Boltzmann, "Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen," *Wiener Berichte* **06**, 275 (1872).
[5] J. C. Maxwell, "Illustrations of the Dynamical Theory of Gases. Part I: On the Motions and Collisions of Perfectly Elastic Spheres," *Philosophical Magazine and Journal of Science,* Fourth Series **19**, 19 (1860).
[6] J. C. Maxwell. "Illustrations of the Dynamical Theory of Gases. Part II. On the Process of Diffusion of Two or More Kinds of Moving Particles Among One Another," *Philosophical Magazine and Journal of Science,* Fourth Series **20**, 21 (1860).
[7] J. C. Maxwell, "Illustrations of the Dynamical Theory of Gases. Part III. On the Collision of Perfectly Elastic Bodies of Any Form," *Philosophical Magazine and Journal of Science,* Fourth Series **20**, 33 (1860).
[8] J. Clerk Maxwell, "On the Dynamical Theory of Gases," *Philosophical Transactions of the Royal Society of London* **157**, 49 (1867).

Alan S. Lapedes, "Neural Networks"

[1] A. Lapedes and R. Farber, "Nonlinear Signal Processing Using Neural Networks," Los Alamos National Laboratory report LA-UR-2662 (1987).
[2] A. Lapedes and R. Farber, "How Neural Nets Work," in *Proceedings of the First IEEE Conference in Neural Information Processing Systems* (American Institute of Physics, Melville, New York, 1988), p. 442.

Darryl D. Holm, "Solitons, Peakons, and Nonlinear Integrability"

[1] S. Ulam, *Collected Papers of Enrico Fermi*, Vol. 2 (University of Chicago Press, Chicago, 1965).
[2] S. Ulam, *Adventures of a Mathematician* (University of California Press, Berkeley, California, 1976).

References

[3] E. Fermi, J. Pasta, and S. Ulam, "Studies of Nonlinear Problems," originally in Los Alamos Scientific Laboratory report LA-1940 (1955); later in S. Ulam, *Stanislaw Ulam: Sets, Numbers, and Universes; Selected Works* (Massachusetts Institute of Technology Press, Cambridge, Massachusetts, 1974), p. 491.
[4] P. D. Lax, "Almost Periodic Behavior of Nonlinear Waves," in *Surveys in Applied Mathematics: Essays Dedicated to S. M. Ulam*, N. Metropolis, S. Orszag, and G.-C. Rota, Eds. (Academic Press, Inc., New York, 1976), p. 259.
[5] J. S. Russell, *Report of the Fourteenth Meeting of the British Association for the Advancement of Science*, York, September 1844 (London, 1845), p. 311, Plates XLVII–LVII.
[6] R. K. Bullough, "The Wave 'Par Excellence,' the Solitary, Progressive Great Wave of Equilibrium of the Fluid—an Early History of the Solitary Wave," *Solitons*, M. Lakshmanan, Ed. (Springer Series in Nonlinear Dynamics, 1988), p. 150. See http://www.ma.hw.ac.uk/solitons/ for additional background.
[7] R. Camassa and D. D. Holm, "An Integrable Shallow Water Equation with Peaked Solitons," *Phys. Rev. Lett.* **71**, 1661 (1993).
[8] H. Dullin, G. Gottwald, and D. D. Holm, "An Integrable Shallow Water Equation with Linear and Nonlinear Dispersion," *Phys. Rev. Lett.* **87**, 194501-1-4 (2001).
[9] M. J. Ablowitz and H. Segur, *Solitons and the Inverse Scattering Transform* (Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1981).
[10] P. G. Drazin and R. S. Johnson, *Solitons: An Introduction* (Cambridge University Press, Cambridge, United Kingdom, 1989).
[11] D. D. Holm, J. E. Marsden, and T. S. Ratiu, "The Euler-Poincaré Equations and Semidirect Products with Applications to Continuum Theories," *Adv. in Math*. **137**, 1 (1998).
[12] C. Foias, D. D. Holm, and E. S. Titi, "The Navier-Stokes-alpha Model of Fluid Turbulence," *Physica D* **152**, 505 (2001).

Joel E. Dendy, Jr., "Solution of Linear and Nonlinear Equations"

[1] T. A. Oliphant, "An Extrapolation Procedure for Solving Linear Systems," *Q. Appl. Math*. **20**, 257 (1962).
[2] Y. Saad and H. A. van de Vorst, "Iterative Solution of Linear Systems in the 20th Century," *J. Comp. Appl. Math.* **123**, 1 (2000).
[3] J. A. Meijerink and H. A. van de Vorst, "An Iterative Solution Method for Linear Systems of which the Coefficient Matrix is a Symmetric M-Matrix," *Math. Comput.* **31,** 137, 148 (1977).
[4] R. E. Alcouffe, "Diffusion Synthetic Acceleration Methods for the Diamond-Differenced Discrete-Ordinates Equations," *Nucl. Sci. Eng.* **64**, 344 (1977).
[5] R. E. Alcouffe, et al., "The Multigrid Method for the Diffusion Equation with Strongly Discontinuous Coefficients," *Society for Industrial and Applied Mathematics J. Sci. Statist. Comput.* **2**, 430 (1981).
[6] J. E. Dendy, Jr., "Black Box Multigrid," *J. Comp. Phys.* **48** , 366 (1982).
[7] J. E. Morel, et al., "A Cell-Centered Lagrangian-Mesh Diffusion Difference Scheme," *J. Comput. Phys*. **103**, 286 (1992).
[8] V. A. Mousseau, D. A. Knoll, and W. J. Rider, "Physics-Based Preconditioning and the Newton-Krylov Method for Non-Equilibrium Radiation Diffusion," *J. Comput. Phys.* **160**, 743 (2000).
[9] D. A. Knoll, et al., "Preconditioning Newton-Krylov Methods in Solidifying Flow Applications," *Society for Industrial and Applied Mathemataics J. Sci. Comput.* **23**, 381 (2001).

National Security

Francis H. Harlow, "Developing Nuclear Weapons"

[1] L. Hoddeson, et al., *Critical Assembly, A Technical History of Los Alamos during the Oppenheimer Years, 1943–1945* (Cambridge University Press, Cambridge, United Kingdom, 1993).

Nuclear Physics

Peter Möller and Arnold J. Sierk, "A Unified Theory of Nuclear Masses, Fission Barriers, and Superheavy Nuclei"

[1] S. G. Nilsson, *Kgl. Danske Videnskab. Selskab. Mat.-Fys. Medd.* **29**, 16 (1955).
[2] W. J. Swiatecki, *Phys. Rev.* **100**, 937 (1955).
[3] S. G. Nilsson, et al., *Nucl. Phys. A* **115**, 545 (1968).
[4] S. G. Nilsson, et al., *Nucl. Phys. A* **131**, 1 (1969).
[5] M. Bolsterli, et al., *Phys. Rev. C* **5**, 1050 (1972).
[6] E. O. Fiset and J. R. Nix, *Nucl. Phys. A* **193**, 647 (1972).
[7] H. J. Krappe and J. R. Nix, *Proceedings of the Third IAEA Symposium on the Physics and Chemistry of Fission*, Rochester, 1973, Vol. I (International Atomic Energy Agency, Vienna, Austria, 1974), p. 159.
[8] H. J. Krappe, J. R. Nix, and A. J. Sierk, *Phys. Rev. C* **20**, 992 (1979).
[9] P. Möller, S. G. Nilsson, and J. R. Nix, *Nucl. Phys. A* **229**, 292 (1974).
[10] P. Möller and J. R. Nix, *Nucl. Phys. A* **272**, 502 (1976).
[11] P. Möller and J. R. Nix, *Nucl. Phys. A* **281**, 354 (1977).
[12] T. Kodama, et al., *Phys. Rev. C* **17**, 111 (1978).
[13] P. Möller and J. R. Nix, *Nucl. Phys. A* **361**, 117 (1981).
[14] P. Möller and J. R. Nix, *Atomic Data Nucl. Data Tables* **26**, 165 (1981).
[15] P. Möller, J. R. Nix, and W. J. Swiatecki, *Proc. Winter Workshop on Nuclear Dynamics V* (Sun Valley, Idaho, 1988), Michigan State University report (1989), p. 5.
[16] P. Möller and J. R. Nix, *Atomic Data Nucl. Data Tables* **39**, 213 (1988).
[17] P. Möller, et al., *Atomic Data Nucl. Data Tables* **59**, 185 (1995).
[18] P. Möller, J. R. Nix, and K.-L. Kratz, *Atomic Data Nucl. Data Tables* **66**, 131 (1997).
[19] G. A. Leander, et al., *Nucl. Phys. A* **388**, 452 (1982).
[20] W. Nazarewicz, et al., *Nucl. Phys. A* **429**, 269 (1984).
[21] R. Bengtsson, et al., *Phys. Scr.* **29**, 402 (1984).
[22] P. Möller and J. R. Nix, *Nucl. Phys. A* **536**, 20 (1992).
[23] P. Möller, et al., *Nature* **409**, 785 (2001).

Joseph Carlson, "From Nuclear Forces to Nuclei"

[1] J. Carlson, *Phys. Rev. C* **38**, 1879 (1988).
[2] S. C. Pieper, et al., *Phys. Rev. C* **64**, 014001 (2001).
[3] J. Carlson and R. Schiavilla, *Rev. Mod. Phys*. **70**, 743 (1998).

Gerald M. Hale and Donald C. Dodder, "R-Matrix Theory of Nuclear Reactions in T-Division"

[1] E. P. Wigner and L. Eisenbud, *Phys. Rev.* **72**, 29 (1947).
[2] A. M. Lane and R. G. Thomas, *Rev. Mod. Phys*. **30**, 257 (1958).
[3] G. M. Hale and D. C. Dodder, *Proceedings of the International Conference on Nuclear Cross Sections for Technology*, Knoxville, Tennessee, October 22–26, 1979 (National Bureau of Standards Special Publication **594**, 1980), p. 650.
[4] G. M. Hale, R. E. Brown, and N. Jarmie, *Phys. Rev. Lett.* **59**, 763 (1987).

Leon Heller, "The Nucleon-Nucleon Interaction"

[1] L. Heller, *Phys. Rev*. **120**, 627 (1960).
[2] M. L. Gursky and L. Heller, *Phys. Rev. B* **136**, 1693 (1964).
[3] L. Heller, P. Signell, and N. R. Yoder, *Phys. Rev. Lett.* **13**, 577 (1964).
[4] L. Heller, "Off-Shell Effects in the Nucleon-Nucleon System," in *Few Body Systems and Nuclear Forces II*, H. Zingl, M. Haftel, and H. Zankel, Eds. (Springer-Verlag, Berlin, West Germany, 1978).

References

David G. Madland and J. Rayford Nix, "The Los Alamos Model of Neutron Emission in Fission"

[1] B. E. Watt, *Phys. Rev.* **87**, 1037 (1952).
[2] N. Feather, *Emission of Neutrons from Moving Fission Fragments*, BM-148 (British Mission, 1942).
[3] F. Manero and V. A. Konshin, *Atomic Energy Rev*. **10**, 637 (1972).
[4] D. G. Madland and J. R. Nix, *Nucl. Sci. Eng*. **81**, 213 (1982).
[5] V. F. Weisskopf, *Phys. Rev.* **52**, 295 (1937).
[6] D. G. Madland, *Proceedings of the Conference 50 Years with Nuclear Fission*, Vol. 1, Washington, D.C., April 25–28, 1989 (American Nuclear Society, La Grange Park, Illinois, 1989), p. 429.

J. N. Ginocchio, "Dynamics of Atomic Nuclei"

[1] F. Iachello and A. Arima, "The Interacting Boson Model"(Cambridge University Press, Cambridge, United Kingdom, 1987).
[2] J. N. Ginocchio and M. Kirson, *Nucl. Phys A* **350**, 31 (1980).
[3] J. N. Ginocchio, *Ann. Phys.* **126,** 234 (1980).
[4] W. C. Haxton and G. J. Stephenson, Jr., *Prog. Part. Nucl. Phys.*, **12**, 409 (1984).
[5] N. Auerbach, et al., *Phys. Rev. C* **38**, 1277 (1988).
[6] J. N. Ginocchio, *Phys. Rev. Lett*. **78**, 436 (1997).
[7] J. N. Ginocchio, *Phys. Rep.* **315**, 231 (1999).

Arnold J. Sierk, "Macroscopic Dynamics of Nuclear Fission and Fusion"

[1] J. R. Nix, "Further Studies in the Liquid-Drop Theory of Nuclear Fission," *Nucl. Phys. A* **130**, 241 (1969).
[2] A. J. Sierk and J. R. Nix, "Dynamics of Fission and Fusion with Applications to the Formation of Superheavy Nuclei," *Proceedings of the Third IAEA Symposium on the Physics and Chemistry of Fission*, Vol. II, Rochester, New York, 1973 (International Atomic Energy Agency, Vienna, Austria, 1974), p. 273.
[3] K. T. R. Davies, J. R. Nix, and A. J. Sierk, "Effect of Viscosity on the Dynamics of Fission," *Phys. Rev. C* **13,** 2385 (1976).
[4] J. Blocki, et al., "One-Body Dissipation and Superviscidity of Nuclei," *Ann. Phys.* (New York) **113**, 330 (1979).
[5] H. J. Krappe, J. R. Nix, and A. J. Sierk, "Unified Nuclear Potential for Heavy-Ion Elastic Scattering, Fusion, Fission, and Ground-State Masses and Deformations," *Phys. Rev. C* **20,** 992 (1979).
[6] J. R. Nix and A. J. Sierk, "Macroscopic Description of Isoscalar Giant Multipole Resonances," *Phys. Rev. C* **21,** 396 (1980); erratum: *Phys. Rev. C* **25**, 1068 (1982).
[7] J. R. Nix and A. J. Sierk, "New Picture of Dissipation in Heavy-Ion Reactions and Other Collective Phenomena," *Proceedings of the International Symposium on Perspectives in Nuclear Physics* (Madras, India, 1987), *J. Madras University* **B50**, 38 (1987).
[8] A. J. Sierk, "Macroscopic Model of Rotating Nuclei," *Phys. Rev. C* **33**, 2039 (1986).

Numerical Analysis and Algorithms

Bryan Kashiwa, "The CFDLib Project"

[1] H. E. Edgerton, *Massachusetts Institute of Technology Technology Review* **34**, 376 (1932).
[2] F. H. Harlow and A. Amsden, "Fluid Dynamics–A LASL Monograph," Los Alamos Scientific Laboratory report LA-4700 (1971).
[3] J. Brackbill, D. Kothe, and C. Zemach, *J. Comput. Phys.* **100**, 335 (1992).

Michael C. Cline, "Analysis of Engineering Flow Applications"

[1] M. C. Cline, "The Computation of Steady Nozzle Flow by a Time-Dependent Method," *AIAA J.* **12**, 419 (1974).
[2] M. C. Cline, "Computation of Two-Dimensional, Viscous Nozzle Flow," *AIAA J.* **14**, 295 (1976).
[3] M. C. Cline and R. G. Wilmoth, "Computation of High Reynolds Number Internal/External Flows," *AIAA J.* **21**, 172 (1983).

Particle Physics

T. Bhattacharya and R. Gupta, "The Standard Model and Its Validation on the Computer"

[1] C. Quigg, *Theories of Strong, Weak and Electromagnetic Interactions* (Benjamin-Cummings, Menlo Park, California, 1983).
[2] N. Cooper and G. West, "Particle Physics: A Los Alamos Primer," *Los Alamos Science* Number **11**, and (Cambridge University Press, Cambridge, United Kingdom, 1984).
[3] K. Wilson, *Physical Review D* **10**, 2445 (1974).
[4] M. Creutz, *Quarks, Gluons, and Lattices* (Cambridge University Press, Cambridge, United Kingdom, 1983).

***Terrance Goldman, Peter Herczeg, and John Terning,* "Investigations in Physics Beyond the Standard Model: A Review"**

[1] M. Gell-Mann, P. Ramond, and R. Slansky, in *Supergravity*: *Proceedings of the Supergravity Workshop at Stony Brook*, September 27–29, 1979, P. van Nieuwenhuizen and D. Z. Friedman, Eds. (North-Holland Publishing Company, Amsterdam, 1979).
[2] T. Yanagida, in *Proceedings of the Workshop on Unified Theory and Baryon Number in the Universe*, O. Sawada and A. Sugamoto, Eds. (KEK, Tsukuba, Japan, 1979); R. N. Mohapatra and G. Senjanovic, *Phys. Rev. Lett.* **44**, 912 (1980); *Phys. Rev. D* **23**, 165 (1981).
[3] G. J. Stephenson, private communication.
[4] S. P. Rosen and J. M. Gelb, "Mikheev-Smirnov-Wolfenstein Enhancement of Oscillations as a Possible Solution to the Solar Neutrino problem," *Phys. Rev. D* **34,** 969 (1986).
[5] G. J. Stephenson, Jr., and T. Goldman, "A Possible Solution to the Tritium Endpoint Problem," *Phys. Lett. B* **440,** 89 (1998); G. J. Stephenson, Jr., T. Goldman, and B. H. J. McKellar, "Tritium Beta Decay, Neutrino Mass Matrices and Interactions beyond the Standard Model," *Phys. Rev. D* **62**, 093013 (2000).
[6] P. Herczeg, "Exotic Muon Decays and Searches for Neutrino Oscillations," in *Beyond the Desert 1997—Accelerator and Non-Accelerator Approaches,* Proceedings of the First International Conference on Particle Physics beyond the Standard Model, Castle Ringberg, Tegernsee, Germany, June 8–14, 1997, H. V. Klapdor-Kleingrothaus and H. Päs, Eds. (Institute of Physics Publishing, Bristol, United Kingdom, 1998), p. 124.
[7] A. C. Hayes and I. S. Towner, "Shell Model Calculations of Neutrino Scattering from C-12," *Phys. Rev. C* **61**, 044603 (2000).
[8] M. Gell-Mann, P. Ramond, and R. Slansky, "Color Embeddings, Charge Assignments, and Proton Stability in Unified Gauge Theories," *Rev. Mod. Phys.* **50**, 721 (1978).
[9] C. Csaki, G. D. Kribs, and J. Terning, "4-D Models of Scherk-Schwarz GUT Breaking via Deconstruction," *Phys. Rev. D* **65**, 015004 (2002).
[10] T. J. Goldman and D. A. Ross, "A New Estimate of the Proton Lifetime," *Phys. Lett.* **84B**, 208 (1979); "How Accurately Can We Estimate the Proton Lifetime in an SU(5) Grand Unified Model?" *Nucl. Phys. B* **171**, 273 (1980); A. S. Goldhaber, T. Goldman, and S. Nussinov, "Meson Field Quenching of Proton Decay," *Phys. Lett.* **142**, 47 (1984).
[11] S. Dimopoulos, S. Raby, and F. Wilczek, "Proton Decay in Supersymmetric Models," *Phys. Lett. B* **112**, 133 (1982).
[12] W. C. Haxton, G. J. Stephenson, Jr., and D. Strottman, "Double Beta Decay and the Majorana Mass of the Electron Neutrino," *Phys. Rev. Lett.* **47**, 698 (1981); "Lepton Number Conservation and the Double Beta Decay of 128Te and 130Te," *Phys. Rev. D* **25**, 2360 (1982).
[13] T. Goldman, E. W. Kolb, and G. J. Stephenson, Jr., "Majorons and Muon Decay," *Phys. Rev. D* **26**, 2503 (1982).
[14] L. J. Hall, V. A. Kostelecky, and S. Raby, "New Flavor Violations in Supergravity Models," *Nucl. Phys. B* **267**, 415 (1986).

References

[15] T. Goldman and G. J. Stephenson, Jr., "Limits on the Mass of the Muon Neutrino in the Absence of Muon Lepton Number Conservation," *Phys. Rev. D* **16**, 2256 (1977).
[16] T. Goldman, et al., "Light-Boson Emission in the Decay of the μ^+," *Phys. Rev. D* **36**, 1543 (1987).
[17] P. Herczeg and T. Oka, "Study of Muon-Number Violating Hyperon Decays," *Phys. Rev. D* **29**, 475 (1984).
[18] P. Herczeg and R. N. Mohapatra, "Muonium to Antimuonium Conversion and the Decay $\mu^+ \rightarrow e^+ \bar{\nu}_e \nu_\mu$ in Left-Right Symmetric Models," *Phys. Rev. Lett.* **69**, 2475 (1992).
[19] P. Herczeg, "On the Mohapatra-Pati Model of CP-Violation," *Phys. Rev. D* **28**, 200 (1983); "Beta Decay Beyond the Standard Model," in *Fundamental Symmetries in Nuclei and Particles*, H. Henrikson and P. Vogel, Eds. (World Scientific Publishing Company, Singapore, 1990), p. 46.
[20] W. C. Haxton and E. M. Henley, "Enhanced T-Violating Nuclear Moments," *Phys. Rev. Lett.* **51**, 1937 (1983).
[21] P. Herczeg, "T-Violation in Nuclear Interactions—An Overview," *Hyperfine Interact.* **43**, 61 (1988).
[22] J. Engel, J. L. Friar, and A. C. Hayes, "Nuclear Octupole Correlations and the Enhancement of Atomic Time-Reversal Violation," *Phys. Rev. C* **61**, 035502 (2000).
[23] P. Herczeg, "P-,T-Violating Electron-Nucleon Interactions in the R-Parity Violating Minimal Supersymmetric Standard Model," *Phys. Rev. D* **61**, 095010-1 (2000).
[24] T. Goldman and C. Hoffman, "Will the Axion Be Found Soon?," *Phys. Rev. Lett.* **40**, 220 (1978).
[25] S. Dimopoulos and S. Raby, "Geometric Hierarchy," *Nucl. Phys. B* **219**, 479 (1983).
[26] C. Csáki, et al., "Universal Aspects of Gravity Localized on Thick Branes," *Nucl. Phys. B* **581**, 309 (2000).
[27] C. Csáki, J. Erlich, and J. Terning, "The Effective Lagrangian in the Randall-Sundrum Model and Electroweak Physics," *Phys. Rev. D* **66**, 064021 (2002).
[28] C. Csáki, J. Erlich, and C. Grojean, "Gravitational Lorentz Violations and Adjustment of the Cosmological Constant in Asymmetrically Warped Space-Times," *Nucl. Phys. B* **604,** 312 (2001).
[29] C. Csáki, et al., "Cosmology of Brane Models with Radion Stabilization," *Phys. Rev. D* **62**, 045015 (2000).
[30] C. Csáki, N. Kaloper, and J. Terning, "Dimming Supernovae without Cosmic Acceleration," *Phys. Rev. Lett.* **88**, 161302 (2002).
[31] I. Antoniadis and E. Mottola, "4-D Quantum Gravity in the Conformal Sector," *Phys. Rev. D* **45**, 2013 (1992).

M. M. Nieto, "Tests of Gravity"

[1] M. M. Nieto and T. Goldman, *Phys. Rep.* **205,** 221 (1991); **216,** 343 (1992).
[2] M. Amoretti, et al., "Production and Detection of Cold Antihydrogen Atoms," *Nature* **419**, 456 (2002).
[3] J. D. Anderson, et al., *Phys. Rev. D* **65**, 082004 (2002).

Plasma Physics

Jeremiah U. Brackbill, Dana A. Knoll, and Giovanni M. Lapenta, "A Brief History of Plasma Modeling in Theoretical Division"

[1] D. W. Forslund and J. P. Freidberg, "Theory of Laminar Collisionless Shocks," *Phys. Rev. Lett.* **27**, 1189 (1971).
[2] J. U. Brackbill and B. I. Cohen, Eds., *Brackbill 1985 Multiple Time Scales* (Academic Press, New York, 1985).
[3] H. X. Vu and J. U. Brackbill, "CELEST1D: An Implicit, Fully Kinetic Model for Low-Frequency, Electromagnetic Plasma Simulation," *Comput. Phys. Commun.* **69**, 25 (1992).
[4] G. Lapenta and J. U. Brackbill, "Nonlinear Evolution of the Lower Hybrid Drift Instability: Current Sheet Thinning and Kinking," *Phys. Plasmas* **9**, 1544 (2002).

[5] J. U. Brackbill and W. E. Pracht, *J. Comput. Phys.* **13**, 455 (1973).
[6] D. C. Barnes, J. U. Brackbill, and W. O. Schneider, "A Numerical Study of High-Beta Stellarator Equilibria," *Nucl. Fusion* **21**, 537 (1981).
[7] J. U. Brackbill, "Numerical Magnetohydrodynamics for High-Beta Plasmas," *Methods in Computational Physics* **16**, 1 (1976).
[8] J. U. Brackbill and D. A. Knoll, "Transient Magnetic Reconnection and Unstable Shear Layers," *Phys. Rev. Lett.* **86**, 2329 (2001).
[9] "A Multilevel Iterative Field Solver for Implicit, Kinetic, Plasma Simulation," *J. Comput. Phys.* **149**, 377 (1999).
[10] D. A. Knoll and L. Chacon, "Magnetic Reconnection in Two-Dimensional Kelvin-Helmholtz Instability," *Phys. Rev. Lett.* **8821**, 5003 (2002).
[11] P. Ricci, G. Lapenta, and J. U. Brackbill, "A Simplified Implicit Maxwell Solver," *J. Comput. Phys.* (accepted for publication in 2003).

Donald F. DuBois and Harvey A. Rose, "Nonlinear Coupling of Coherent Electromagnetic Waves with Plasma"

[1] D. F. DuBois, et al., *Phys. Plasmas* **8**, 791 (2001)
[2] P. Y. Cheung, et al., *Phys. Plasmas* **8**, 802 (2001).
[3] D. A. Russell, D. F. DuBois, and H. A. Rose, *Phys. Plasmas* **6**, 1294 (1999).
[4] H. X. Vu, D. F. DuBois, and B. Bezzerides, *Phys. Rev. Lett.* **86**, 4306 (2001).
[5] D. S. Montgomery, et al., *Phys. Rev. Lett.* **87**, 155001 (2001).
[6] D. S. Montgomery, et al., *Phys. Rev. Lett.* **84,** 678 (2000).

Appendix B: Acronym List

1-D	one-dimensional
2-D	two-dimensional
3-D	three-dimensional
3QS	three-quadratic-surface
AAA	Advanced Accelerator Applications Program
ACL	Advanced Computing Laboratory (Los Alamos)
AEC	Atomic Energy Commission
ALE	Arbitrary Eulerian-Lagrangian (CFD method)
ANFO	ammonium nitrate/fuel oil
ANL	Argonne National Laboratory
ARPAnet	Advanced Research Projects Agency Network
ASC	Advanced Simulation and Computing Program (formerly ASCI)
ASCI	Accelerated Strategic Computing Initiative
ATW	Accelerator Transmutation of Waste Initiative
AU	astronomical unit
BBN	Bolt, Beranek and Newman
bcc	body-centered cubic atomic structure
BES	Office of Basic Energy Sciences (DOE)
BHR model	Besnard-Harlow-Rauenzahn model
BP	British Petroleum
C-Division	Chemistry Division (Los Alamos, 1940s)
C-Division	Computing Sciences and Services Division (Los Alamos, 1960s and 1970s)
CCN Division	Computing, Communications, and Networking Division (Los Alamos, current)
CCSM	community climate system model
CDC	Center for Disease Control and Prevention
CE	coronal equilibrium
CEA	Commissariat a L'Energie Atomique (France)
CERN	European Center for Nuclear Research
CFD	computer fluid dynamics
CFDLib	CFD code library
CH	Camassa-Holm
CHAMMP	Computer Hardware, Advanced Mathematics and Model Physics Program (DOE)
CICE	Sea Ice Model
CNO	carbon-nitrogen-oxygen
COM	center of mass

COSIM	Climate, Ocean, and Sea Ice Simulation Program (Los Alamos)
CRADA	Cooperative Research and Development Agreement
CRI	Cray Research, Inc.
CSMS	Center for Semiconductor Modeling
CTCP	computational turbulence closure problem
CTR	controlled thermonuclear research
CTR Division	Controlled Thermonuclear Research Division (Los Alamos)
DARPA	Defense Advanced Research Projects Agency
DCA	detailed configuration accounting
DFT	density functional theory
DI	direct injection (design)
DISC	direct fuel injection stratified charge
DNA	dioxyribonucleic acid
DNS	direct numerical stimulation
DOD	U.S. Department of Defense
DOE	U.S. Department of Energy
DOE/DP	DOE Office of Defense Programs
DOE/EERE/OIT	DOE Office of Energy Efficiency and Renewable Energy/Office of Industrial Technologies
DOE/OFES	DOE Office of Fusion Energy Sciences
DOT	U.S. Department of Transportation
DWA	distorted wave approximation
ECA	Earth-crossing asteroid
EDA	Energy Dependent Analysis
EMBL	European Molecular Biological Laboratory
ENDF	evaluated nuclear data file
ENIAC	electronic numerical integrator and calculator
EOS	equation of state
EPA	Environmental Protection Agency
ERDA	Energy Research and Development Administration
ETC*	European Center for Theoretical Studies in Nuclear Physics and Related Areas (Italy)
EUV	extreme ultraviolet radiation
eV	electron volt(s)
fcc	face-centered cubic
FD	finite difference
FE	finite element
FFB	fast-fluidized-bed reactor
FLC	Federal Laboratory Consortium
FLIP	fluid-implicit-particle (CFD method)
FPU	Fermi-Pasta-Ulam
GBEC	gravitational Bose-Einstein condensation

GCM	general circulation model
GDI	gasoline-direct-injection (design)
GK	Green-Kubo
GL	Ginzburg-Landau (theory)
GLM	generalized Lagrangian mean (theory)
GMC	General Motors Corporation
GTA	ground test accelerator
Gyr	gigayear(s)
hcp	hexagonal close-packed structure
HEO	heavy element opacity
HEVR	high-explosive violent reaction
HIV	human immunodeficiency virus
HPCC	high-performance computing and communication
IAEA	International Atomic Energy Agency
IBM	International Business Machines
ICE	implicit-continuous-fluid Eulerian (CFD method)
ICF	inertial confinement fusion
IDL	interactive data language
IEC	inertial-elecrostatic confinement
IEEE/AIP	Institute of Electrical & Electronics Engineers American Institute of Physics
IGPP	Institute of Geophysical and Planetary Physics (Los Alamos)
IMB	Irvine-Michigan-Brookhaven
IPCC	Intergovernmental Panel on Climate Change (United Nations)
IST	inverse scattering transform
J-Division	Field Test Division (Los Alamos)
JAN	Joint-Army-Navy
JET	Joint European Torus
JINR	Joint Institute for Nuclear Research (Russia)
JPL	Jet Propulsion Laboratory
K	kelvin
KdV	Korteweg-de Vries
KEK	National Laboratory for High-Energy Physics (Japan)
kg	kilogram(s)
KH theorem	Karman-Howarth theorem
km	kilometer(s)
LaGriT	Los Alamos grid toolbox
LAMPF	Los Alamos Meson Physics Facility
LANL	Los Alamos National Laboratory (name after 1980)
LANS-alpha	Lagrangian-averaged Navier-Stokes-alpha
LAPD	large plasma device

LASL	Los Alamos Scientific Laboratory (name before December 1980)
LDA	local density approximation
LDRD	Laboratory-Directed Research and Development
LDRD/DR	Laboratory-Directed Research and Development/ Directed Research
LEAR	Low-Energy Antiproton Ring
LEO	light-element opacity code
LLNL	Lawrence Livermore National Laboratory
LNPI	Leningrad Nuclear Physics Institute (Union of Soviet Socialist Republics)
LTE	local thermodynamic equilibrium
M	solar mass(es)
MAC	marker-and-cell (CFD method)
MANIAC	mathematical analyzer, numerical integrator, and computer
MATLAB	software compiler developed by The MathWorks, Inc. (Trademarked)
MCISDC	Multicharged Ion Spectral Data Center (Russia)
MD	molecular dynamics
MED	Manhattan Engineering District
MeV	million electron volt(s)
MFDRC	Multiphase Fluid Dynamics Research Consortium
MFE	magnetic fusion energy
MHD	magnetohydrodynamics
mCF	muon-catalyzed fusion
MIT	Massachusetts Institute of Technology
MP Division	Meson Physics Division (Los Alamos)
MTA	(Cray MTA) multithreaded architecture system
NAG-FRAG	"Seaman's NAG-FRAG concepts"
NALM	nonlinear amplifying loop mirror
NASA	U.S. National Aeronautics and Space Administration
NASP	National Aerospace Plane Program
NCAR	National Center for Atmospheric Research
NCBI	National Center for Biotechnology Information
NEA	near-Earth asteroid
NEA	Nuclear Energy Agency
NEMD	nonequilibrium molecular dynamics
NERSC	National Energy Research Scientific Computing Center
NIAID	National Institute of Allergy and Infectious Diseases
NIGMS	National Institute of General Medical Sciences

NIH	National Institutes of Health
NIST	National Institute of Standards and Technology
NRC	Nuclear Regulatory Commission
NSF	National Science Foundation
NTRS	National Technology Roadmap for Semiconductors
NUCOR	New Mexico Universities Collaborative Research
NWT	Nuclear Weapons Technology Programs
O-Division	Ordnance Division (Los Alamos, 1940s)
OCGCNR	open-cycle gas-core nuclear rocket
OSU	Ohio State University
P-Division	Physics Division (Los Alamos)
PBX	plastic-bonded explosive
PC	personal computer
PDF	Portable Document Format
PHERMEX	pulsed high-energy radiographic machine emitting x-rays
PIC	particle-in-cell
POP	Parallel Ocean Program
PRONTO	"finite-element computer code"
PS	PostScript
PSI	Paul Scherrer Institut (Switzerland)
PWR	pressurized water reactor
QCD	quantum chromodynamics
QED	quantum electrodynamics
QFT	quantum field theory
RAL	Rutherford Appleton Laboratory
RFP	reversed-field pinch
RVE	representative volume element
s	second(s)
SAE	Society of Automotive Engineers
SAIC	Scientific Applications International Corporation
SCRAM	statistical crack mechanics
SCRIP	spherical coordinate remapping and interpolation package
SDI	Strategic Defense Initiative
SEAS	stimulated electron acoustic scatter
SESAME	equation of state library
SIMD	single instruction multiple data
SIN	Scwerezeresches Institut für Nuklearforschung (Switzerland)
SLAC	Stanford Linear Accelerator Center
SOLA	solution-algorithm (CFD method)
SPaSM	scalable portable short-range molecular dynamics

SRB	solid rocket booster
SRC	Semiconductor Research Corporation
SRI	Stanford Research Institute
SSPT	solid-solid phase transformations
SWIG	simplified wrapper and interface generator
TA	transition array
TAD	temperature-accelerated dynamics
TATB	triamino trinitro benzene
TDL	target detection limit
TE	thermodynamic equilibrium
TFD	Thomas-Fermi-Dirac (theory)
TRIUMF	Tri-Universities Meson (Physics) Facility/Factory (British Columbia, Canada)
TST	transition state theory
UC Berkeley	University of California, Berkeley
USSR	Union of Soviet Socialist Republics
VACF	velocity autocorrelation factor
VAX	Virtual Address Extension
VNIIEF	Russian Federation Nuclear Center (Russia)
VNIIFTRI	All-Russian Research Institute of Physical-Technical Radio-Technical Measurement (Russia)
VOF	volume-of-fluid (CFD method)
ZECA	Zero Emission Coal Alliance

Appendix C: Index of Researchers

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

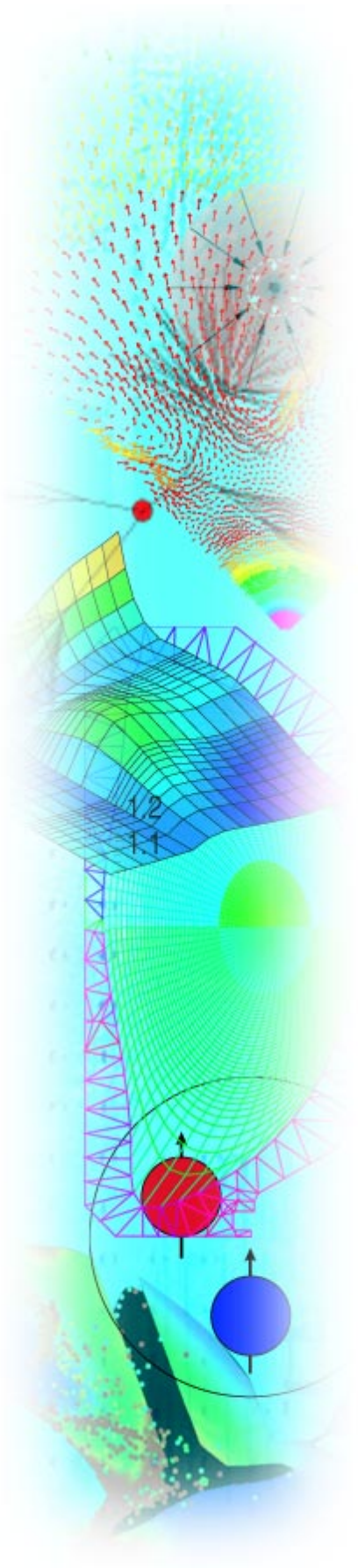
V

W

Y

Z

LA-14000-H



Los Alamos
NATIONAL LABORATORY

LOS ALAMOS NEW MEXICO 87544


1943 - 2003
Los Alamos
NATIONAL LABORATORY
Ideas That Change the World