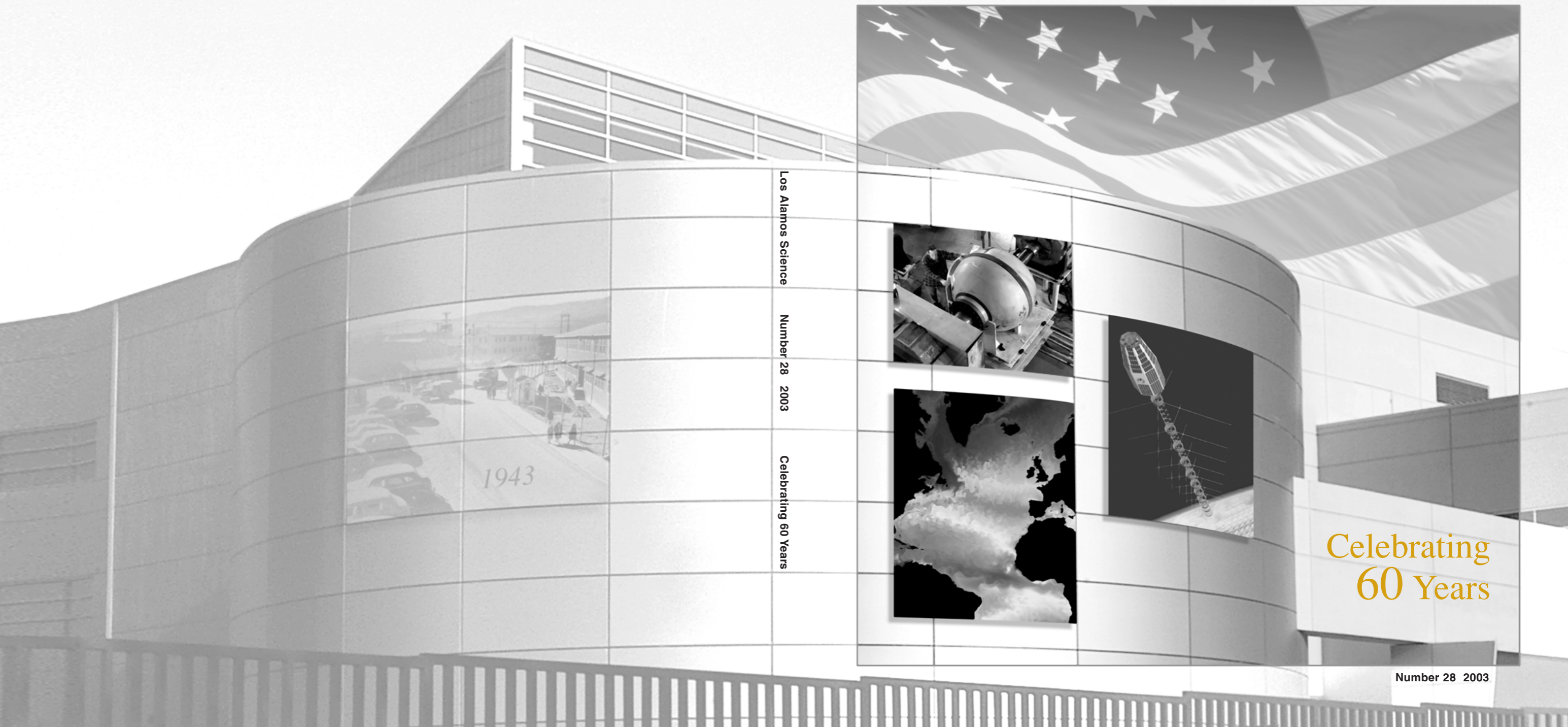


1943 - 2003  
• **Los Alamos**  
NATIONAL LABORATORY  
*Ideas That Change the World*

# Los Alamos Science

LOS ALAMOS NATIONAL LABORATORY



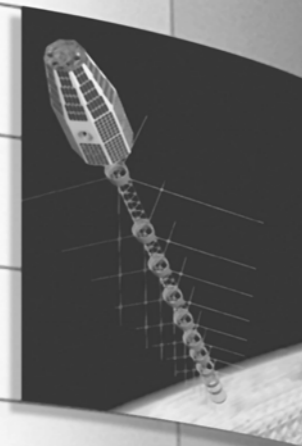
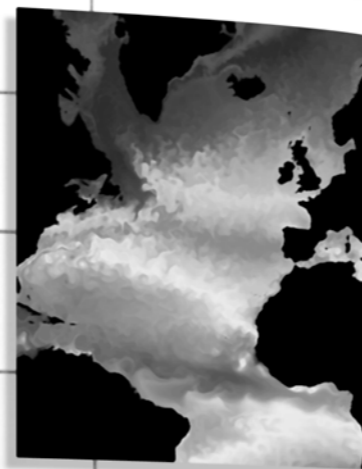
Los Alamos Science

Number 28 2003

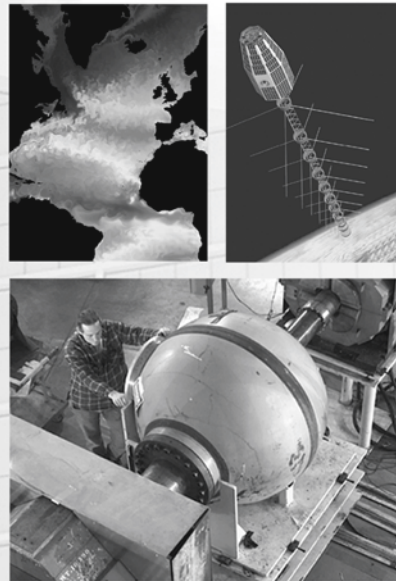
Celebrating 60 Years



1943



Celebrating  
60 Years



*About the Cover*  
*The technical area during the Manhattan Project (back cover) and the sparkling new Nicholas C. Metropolis Center for Modeling and Simulation (front cover) are symbols of reflection and renewal. These themes reverberate through this anniversary volume, celebrating 60 years of service to the nation. The volume presents a portrait of the Laboratory's main activities—ensuring the safety and reliability of the nation's nuclear stockpile; reducing the threat of weapons of mass destruction, proliferation, and terrorism; solving national problems in defense, energy, environment, and infrastructure; and providing the intellectual foundations for our missions through strategic investments in cutting-edge research. (Upper left) High-resolution modeling of ocean circulation, including the mesoscale eddies that redistribute energy, is providing the first accurate models of the Gulf Stream and other critical flows that determine climate. (Upper right) The FORTÉ satellite, developed at Los Alamos as a test bed for the next generation of broadband radio-frequency (rf) sensors, was launched in 1997 and is still in orbit. The Los Alamos broadband sensors not only monitor the sky for possible signals from nuclear detonations but detect the rf signatures of terrestrial lightning. (Bottom) The containment vessel in the proton radiography line at the Los Alamos Neutron Science Center confines dynamic experiments relating to the early stages of nuclear weapons performance. High-precision radiographs of the experiment are made with energetic protons instead of x-rays.*

We would like to thank the following people for their help on this volume:

- Alan Bishop
- John C. Browne
- Gary Dean Doolen
- Dennis J. Erickson
- William J. Feiereisen
- Daniel J. Gerth
- Marie F. Harper
- Allen Hartford, Jr.
- Houston T. Hawkins
- Edward A. Heighway
- Mary Y. Hockaday
- John D. Immele
- Raymond J. Juzaitis
- Richard Mah
- Carolyn A. Mangeng
- John B. McClelland
- Roger A. Meade
- Ping Lee
- James C. Porter, Jr.
- Susan J. Seestrom
- William L. Thompson



Los Alamos National Laboratory, an affirmative action/equal opportunity employer, is operated by the University of California for the U.S. Department of Energy under contract W-7405-ENG-36. All company names, logos, and products mentioned herein are trademarks of their respective companies. Reference to any specific company or product is not to be construed as an endorsement of said company or product by the Regents of the University of California, the United States Government, the US Department of Energy, nor any of their employees. The Los Alamos National Laboratory strongly supports academic freedom and a researcher's right to publish; as an institution, however, the Laboratory does not endorse the viewpoint of a publication or guarantee its technical correctness.



LA-UR-03-5704

# *Celebrating* **60** *Years*

*Editor*  
Necia Grant Cooper

*Managing Editor*  
Ileana G. Buican

*Science Writer*  
Jay A. Schecker

*Designer*  
Gloria E. Sharp

*Illustrators*  
Andrea J. Kron  
Christopher D. Brigman  
David R. Delano  
Ward G. Zaelke

*Editorial Support*  
Faith J. Harp  
Charmian O. Schaller

*Composition Support*  
Joy E. Baker

*Photographers*  
John A. Flower  
Richard C. Robinson  
Michael D. Greenbank

*Printing Coordination*  
Guadalupe D. Archuleta

Address mail to  
*Los Alamos Science*  
Mail Stop M711  
Los Alamos National Laboratory  
Los Alamos, NM 87545

lascience@lanl.gov  
Fax: 505-665-4408  
Tel: 505-667-1447

Inspiration from Our Past  
The Laboratory Today  
Building the Future

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

This volume of *Los Alamos Science* commemorates six decades of service to the nation by the Los Alamos National Laboratory staff and by the University of California.

Over the years, the freedom to explore new ideas has been protected by the traditions and prestige of the University of California and has made Los Alamos one of the great scientific organizations in the world. The connection between Los Alamos and the University began in 1942 with the Berkeley summer study on the building of the atomic bomb. Soon thereafter, in February 1943, the University and the government signed an agreement “for certain investigations to be directed by Dr. J. R. Oppenheimer.” Since then, the University’s connection to Los Alamos has been uninterrupted. In 1946, when the Laboratory’s future was in question, Director Norris Bradbury represented Los Alamos as an institution with a “very definite academic tradition in spite of the fact that we are only about three years old. The entire staff of the Laboratory has been drawn almost without exception from the staffs of academic institutions and from their graduate students.”

The intellectual leadership and diversity of the staff, constantly renewed through collaborations with the international community, have ensured the unmatched strength of the nation’s nuclear deterrent and have produced new ideas and technologies, many of which are applicable to issues of national security. Because the interior workings of a nuclear bomb involve temperatures and pressures that could never be reached in the laboratory, theoretical physics, mathematics, and new diagnostic techniques played an essential role in filling that gap at the founding of the Laboratory. Those capabilities continue to be central today, as we work to maintain the nuclear deterrent in the absence of nuclear testing. Because Oppenheimer, himself a theoretical physicist and a master at managing creative people, believed that openness among all levels of the scientific staff was essential to achieving the goal, the staff were quick to learn, adapt, and respond as new facts presented themselves. That heritage serves us well right now, as we adapt to evolving issues of national security. Given our shared experience on uniquely difficult problems, we strongly believe that a continuing relationship between the University and the Laboratory is in the best interest of the nation.

In the last six months, many of us have participated in the anniversary activities inspired by the theme “Ideas That Change the World.” Publication of this special volume is a fitting close to those activities. *Los Alamos Science* typically presents the excellence of our science to the international scientific community, but this volume was produced with a different purpose in mind. The idea was to create a forward-looking portrait of the Laboratory, from which we could learn more about ourselves and about the tough problems we face in stewardship, threat reduction, and national security in the broadest sense. The volume begins by taking us back to the Laboratory’s first decade through Harris Mayer’s personal reflection “People of the Hill,” and then it turns the spotlight on our present and future national security missions. It gives presence to both older and younger staff, voice to fears and hopes, and welcome to the enthusiasm, dedication, and can-do spirit that continue to motivate this institution.

In my vision of our Laboratory, all members of the staff learn about the technical issues we face and become actively engaged in their solution. The articles in the section on nuclear stewardship are a step in that direction. Both management and research staff share their varying views on the scientific challenges of certifying the safety and reliability of the nuclear weapons stockpile without the benefit of nuclear testing. In “How Archival Test Data Contribute to Certification,” two of our most experienced scientists, in collaboration with a young designer, give us a palpable description of the workings of a nuclear bomb and the complex physics experiments that were performed during more than a thousand nuclear tests to record and analyze the time history of events. After 60 years, that legacy of research is still the bedrock of knowledge for the present certification of the stockpile and a training ground for new staff. That intellectual legacy derives from the excellence of our staff and attests to our University of California heritage. At the same

time, the staff work on new theory, computational techniques, and experimental measurements and from them construct high-fidelity computer simulations of weapons performance. Those simulations, we hope, will fill the gap left by the moratorium on testing. I am encouraged by advancements in the key diagnostic of quantitative radiography, including the new technique of proton radiography. That diagnostic should help us image the early stages of weapon assembly and resolve important uncertainties (“The Development of Flash Radiography”). Another significant development is a new, efficient computational model for simulating the detonation of high explosives (“High-Explosives Performance”). That model is adding certainty into the simulations of the initial stages of weapon performance.

The tragic day of September 11, 2001, was a devastating realization of some of our worst forebodings. After the collapse of the Soviet Union, the Laboratory had begun to focus on the prospect that nuclear weapons or materials could find their way into the hands of dangerous proliferants or terrorists. The second major mission of the Laboratory became to prevent, deter, detect, respond to, and reverse the threat of weapons of mass destruction, proliferation, and terrorism. Some of our most innovative scientists and engineers, working in multidisciplinary teams, are applying their talents and best ideas to these highly complex problems. In our work on biothreat reduction, for example, spinoffs from the Laboratory’s work on the Human Genome Project helped us penetrate the secrets of the bioweapons programs in the former Soviet Union and in Iraq (“Reducing the Biological Threat”). Our long history of using satellites to verify nuclear nonproliferation treaties has prepared us to develop the types of remote sensing we need today (“Eyes in Space”). We are now working with universities, research laboratories, the very best high-technology companies, and scientists worldwide to prevent the illegal migration of nuclear materials and technologies. Most of all, because our work is increasing the safety and security of freedom-seeking people everywhere, the men and women of this Laboratory are proud of their role in the post-9/11 world.

In “Six Decades of Reducing Threats and Allaying Fears,” Terry Hawkins gathers in a few short pages the story of Los Alamos from the dark days of World War II to our present contribution to the war on terror. Should you ever doubt the role of the Laboratory in maintaining freedom and democracy around the world, let that article remind you of how the silent presence of our strategic and tactical nuclear weapons and the visible power of our surveillance have brought caution and sanity when there might have been none.

Our Laboratory’s “self-portrait” closes with the section “Strategic Investments,” that is, research in fundamental science and in technologies that can provide the intellectual foundations of our mission and help sustain the health and well-being of life on our planet. We are preparing for the national security issues of the future by investing broadly in advancing the frontiers of science and in nurturing the talented scientists whose ideas will change the world.

In the years ahead, the world is likely to become more complicated, and the challenges to national security, more diverse. To perform its missions effectively, the Laboratory is renewing itself inside and out. We are creating business and process systems that mirror and support our excellence in science. We are developing the leadership qualities to meet the growing demands of our complex society. We are planning every leg of this journey with purpose and deliberation. As the new director, I am bullish about the future before us. We will continue to be the best because of our creativity, diversity, and unswerving dedication. ■



G. Peter Nanos, Laboratory Director

Preface *G. Peter Nanos, Laboratory Director* . . . . .iv

---

## INSPIRATION FROM OUR PAST

---

People of the Hill—The Early Days . . . . . 2  
*Harris Mayer*

---

## THE LABORATORY TODAY

---

### Nuclear Stewardship in the 21st Century

Science-Based Stockpile Stewardship—An Overview . . . . . 32  
*Raymond J. Juzaitis*

How Archival Test Data Contribute to Certification . . . . . 38  
*Fred N. Mortensen, John M. Scott, and Stirling A. Colgate*

QMU and Nuclear Weapons Certification—What’s under the Hood? . . . . . 47  
*David H. Sharp and Merri M. Wood-Schultz*

Weapon Certification—A Personal View . . . . . 54  
*Donald R. McCoy*

The Pit Production Story . . . . . 58  
*Douglas D. Kautz, David B. Mann, Richard G. Castro, Lawrence E. Lucero, and  
Steven M. Dinehart*

Strategy for Small-Lot Manufacturing—In-Process Monitoring and Control . . . . . 63  
*Vivek R. Davé, Daniel A. Hartman, William H. King, Mark J. Cola,  
and Rajendra U. Vaidya*

The New World of the Nevada Test Site . . . . .	68
<i>Ghazar R. Papazian, Robert E. Reinovsky, and Jerry N. Beatty</i>	
The Development of Flash Radiography . . . . .	76
<i>Gregory S. Cunningham and Christopher Morris</i>	
The DARHT Camera . . . . .	92
<i>Scott A. Watson</i>	
High-Explosives Performance—Understanding the Effects of a Finite-Length Reaction Zone . . . . .	96
<i>John B. Bdzil with Tariq D. Aslam, Rudolph Henninger, and James J. Quirk</i>	
Damage Evolution in Ductile Metals—Developing a Quantitative and Predictive Understanding . . . . .	111
<i>Anna K. Zurek, W. Rich Thissell, Carl P. Trujillo, Davis L. Tonks, Benjamin L. Henrie, and Rhonald K. Keinigs</i>	
Shock Compression Techniques for Developing Multiphase Equations of State . . . . .	114
<i>Robert S. Hixson, George T. Gray III, and Dennis B. Hayes</i>	
ASCI—Big Engineering in the Field of Scientific Computing . . . . .	120
<i>James S. Peery</i>	
Massively Parallel Multiphysics Code Development . . . . .	128
<i>Jim E. Morel</i>	
High-Resolution Methods for Hydrodynamics . . . . .	132
<i>William J. Rider</i>	
A Vision of Hidden Worlds . . . . .	135
<i>Robert J. Kares</i>	
The LANSCE National User Facility . . . . .	138
<i>Thomas Wangler and Paul W. Lisowski</i>	

## THE LABORATORY TODAY

---

### Threat Reduction and Homeland Security Developing Strategies and Tools for Protecting Our Nation and the World

Six Decades of Reducing Threats and Allaying Fears . . . . .	144
<i>Houston T. Hawkins</i>	
Eyes in Space—Sensors for Treaty Verification and Basic Research . . . . .	152
<i>William C. Friedhorsky, Richard D. Belian, Steven P. Brumby, D. Roussel-Dupre, David J. Lawrence, Edward E. Fenimore, Maya Gokhale, John T. Gosling, Cheng Ho, Stephen O. Knox, Geoffrey D. Reeves, David M. Suszcynsky, W. Thomas Vestrand, and Jay A. Schecker</i>	
The Little Satellite That Could . . . . .	155
<i>Diane Roussel-Dupré</i>	
Gotcha! You Blinked! . . . . .	161
<i>W. Thomas Vestrand</i>	
Near Space and ENA Imaging . . . . .	162
<i>John T. Gosling and Geoffrey D. Reeves</i>	
Geochemical Studies of the Moon and Planets . . . . .	166
<i>William C. Feldman</i>	
Reducing the Biological Threat—Detection, Characterization, and Response . . . . .	168
<i>Paul J. Jackson and Jill Trehwella</i>	
National Health Security . . . . .	177
<i>I. Gary Resnick</i>	
Fluorobodies—Mixing Antibodies and the Green Fluorescent Protein to Unravel the Genomic Revolution . . . . .	178
<i>Andrew M. Bradbury, Geoffrey S. Waldo, and Ahmet Zeytun</i>	
Analyzing Pathogen DNA Sequences . . . . .	182
<i>Thomas S. Brettin</i>	
Understanding <i>Why</i> —Dissecting Radical Islamist Terrorism with Agent-Based Simulation . . . . .	184
<i>Edward P. MacKerrow</i>	
The Los Alamos Center for Homeland Security . . . . .	192
<i>Thomas W. Meyer, J. Wiley Davidson, I. Gary Resnick, Ray C. Gordon III, Brian W. Bush, Cetin Unal, Gaspar L. Toole, L. Jonathan Dowell, and Sara C. Scott</i>	



---

## BUILDING THE FUTURE

---

### Strategic Investments

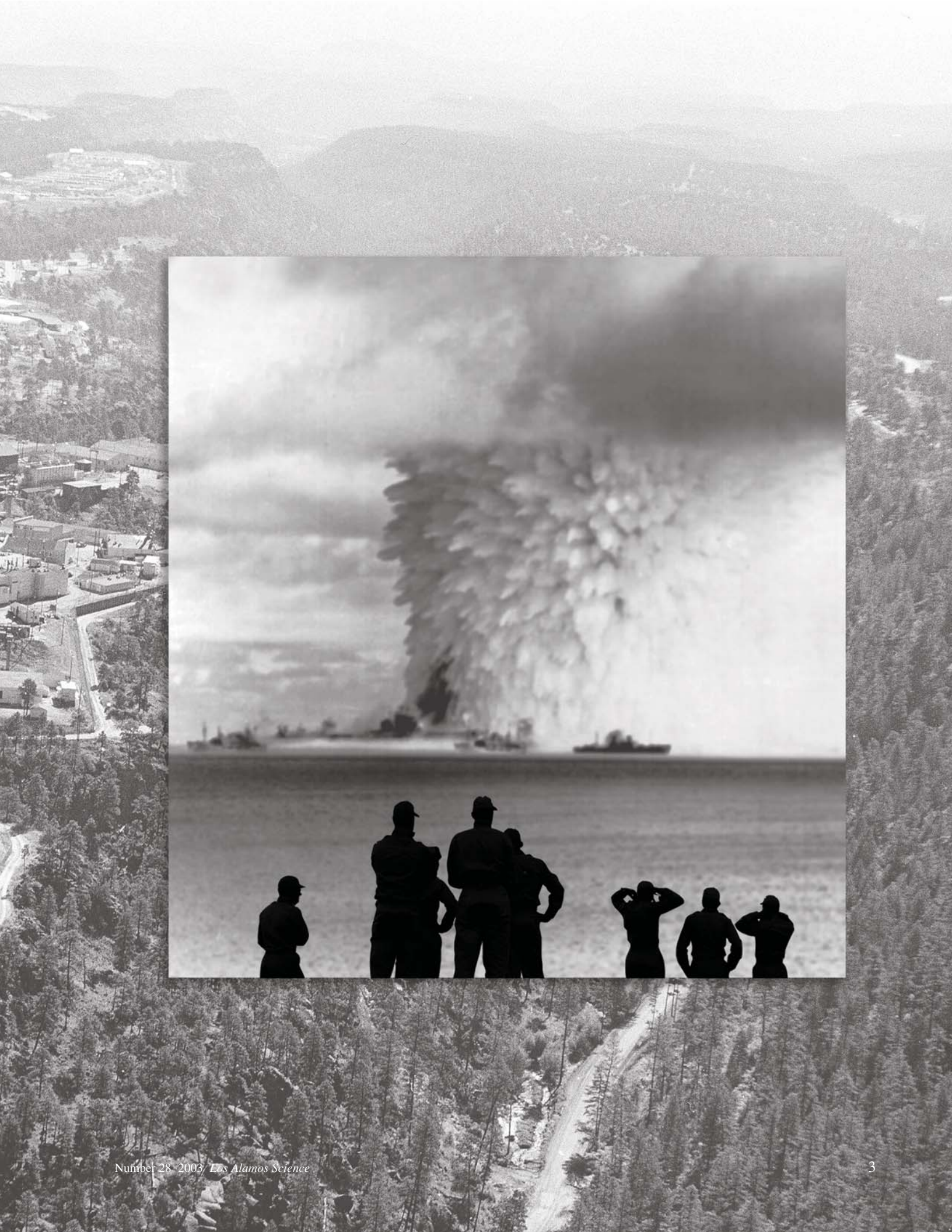
Strategic Research at Los Alamos . . . . .	200
<i>Rajan Gupta and David E. Watkins</i>	
Resolving the Solar Neutrino Problem. . . . .	201
<i>Andrew Hime</i>	
A New Source of Ultracold Neutrons . . . . .	202
<i>Chen-Yu Liu, Steve K. Lamoreaux, Thomas J. Bowles, and Christopher Morris</i>	
The National High Magnetic Field Laboratory—A User Facility. . . . .	203
<i>Alex H. Lacerda and Gregory S. Boebinger</i>	
EpiSims . . . . .	204
<i>Stephen Eubank</i>	
The Magnetized Universe. . . . .	205
<i>Hui Li and Stirling A. Colgate</i>	
Computational Tools to Battle HIV . . . . .	208
<i>Bette T. M. Korber and Alan S. Perelson</i>	
Nanocrystal Quantum Dots—From Fundamental Photophysics to Multicolor Lasing . . . . .	214
<i>Victor I. Klimov</i>	
Nanoscience and the Center for Integrated Nanotechnologies . . . . .	221
<i>Don Merrill Parkin</i>	
Eddy-Resolving Ocean Modeling. . . . .	223
<i>Robert C. Malone, Richard D. Smith, Mathew E. Maltrud, and Matthew W. Hecht</i>	
Virtual Watershed—Simulating the Water Balance of the Rio Grande Basin . . . . .	232
<i>Larry Winter and Everett P. Springer</i>	
Water for Energy—A Critical Piece of the Energy Sustainability Puzzle . . . . .	238
<i>Anthony Mancino and Charryl L. Berger</i>	
Toward a Sustainable Energy Future—Fuel Cell Research at Los Alamos . . . . .	240
<i>Kenneth R. Stroh</i>	
Highlights of the Laboratory’s Anniversary Celebration. . . . .	246
<i>Written and designed by Dennis J. Erickson and Andrea M. Gaskey</i>	

*Inspiration from the Past*

# People of the Hill

## *The early days*

*Harris Mayer*



## Preface

In the first decade of its existence, 1943 to 1953, the Los Alamos Laboratory developed the fission weapon and the thermonuclear fusion weapon, popularly known as the atomic bomb and the hydrogen bomb. This memoir of that early period is one person's viewpoint, the view of a man now over 80 years old, looking back on a golden time when he first arrived in Los Alamos with his new bride in March 1947. It is my recall, seasoned with the knowledge of a lifetime, of a new town and a new laboratory.

Most of the scientists in this story were known to me personally. Others, I knew through the eyes of my young close friends. But my knowledge is only that of a student, blooming into scholarship in the presence of some of the master scientists of the era. That there is wonder and worship is no accident; these are my personal impressions, not the complete view of a skilled biographer. Of course, these people are far more complex than revealed to me by the professor-student relation. Also, I have stayed entirely within the period of that first decade, before the Oppenheimer security investigation, which polarized the scientific community and profoundly altered its relationships. I have not permitted that tragic affair to rewrite the sentiments of the earlier time.

So this account is not meant to be history's dispassionate catalog of events. On the contrary, it is an attempt to give my personal impassioned interpretation of events as I perceived them, played out by people as I have known them. In this process, I have tried to capture some of the essential spirit of the Laboratory at that time. Here is one play that I have written among the many that others could write. I have much enjoyed this scripting. I hope the reader will enjoy the production.

## Part I: The Fission Weapons (1943–1946)

### The Town, Security, the People

During the war, as a newly hired Los Alamos staff member, you checked in at 109 Palace Avenue, in Santa Fe, where Dorothy McKibben welcomed you warmly and processed your initial paperwork. You would still be unaware of what your job would be like. Then a no-nonsense WAC in a dusty Jeep drove you up a dirt mountain road, fit only for a wagon trail, that led to the top of a mesa shouldering the east slope of the Jemez mountain range. This range is the rim of the caldera formed over a million years ago when an ancient volcano blew its top. The result of the explosion was the Valle Grande, a beautiful meadowed valley measuring some 15 by 20 miles, a favorite visiting site of the Los Alamos residents.

Any arriving young city dweller was in for a culture shock. The streets were unpaved. They were a beaded string of mud puddles after a rain, a rutted obstacle course when dried out. The temporary huts and barracks that served as residences were scattered apparently randomly about the mesa top. That arrangement was one of the two great administrative accomplishments of J. Robert Oppenheimer, the Laboratory director. He had convinced General Groves, the leader of the entire Manhattan nuclear weapons project, to spare the pine trees that dotted the mesa. In typical military fashion, Groves had planned to bulldoze the area flat to facilitate construction. At Oppenheimer's urgent insistence, however, the new houses were not placed monotonously side by side along barren checkerboard streets but were angled higgledy-piggledy among

the trees. Only the few houses once occupied by the faculty and owners of the Los Alamos Ranch School, which had been taken over by the Army, were substantial. The only houses that had bathtubs, they still stand today along the appropriately named Bathtub Row. Oppie, as every one called him, and his family occupied one of them. But all in all, what the newly hired saw was an ugly shanty town, mud in the streets, wash drying on the outside clothes lines, babies bawling in a place secluded, unknown to the outside world. Yet this was the town, under the startlingly clear air of the northern New Mexico mountains, where clear-eyed young men were busy discerning nature, defying technical difficulties, striving against time to alter a foreboding history.

At the center of town was Ashley Pond, about the size of a baseball field. Buildings from the Los Alamos

Ranch School were retained to the north: Fuller Lodge with its dramatic big log construction; the Big House, also rustic; and of course, a few hundred yards farther along, the Bathtub Row homes. Also in this area, the Army had built a cafeteria and a commissary run by the military. In a time of war rationing, everyone could get great buys at the commissary, which was like a combined general store and small supermarket. This was one of many fringe benefits to ease the otherwise Spartan living conditions on The Hill. To the south and east, fronting on the pond, was an elbow of wood construction technical (tech) area buildings. On the reverse side of these buildings, an unpaved road ran, later named Trinity Drive after the site of the first test of a nuclear explosion. Two enclosed pedestrian bridges over Trinity formed convenient passageways to other similar office buildings and to Laboratory facilities. The tech area was isolated from the general town site by security fences. Military guards, pistols at hip, checked one's tech area badge at entrances on either side of Trinity Drive. Designed to be temporary, the buildings outlasted the war and eventually even their economic utility. Starting in 1956, as the new tech area was built on the adjacent mesa to the south of town, the old buildings were torn down. Around Ashley Pond today only Fuller Lodge remains, modernized with additions in keeping with its original style. Nowadays, the Los Alamos Inn stands on land cleared of the old Lab buildings. Musing there, one can sense the spirit of that dynamic, vibrant old tech area. The meandering winds seem to whisper in one's ear, recounting the wondrous secrets of the atomic bomb that they long ago overheard here.

This was the city on the mesa. But the people living there never called themselves mesa dwellers, hardly ever said, "I live in Los Alamos." Los Alamos was usually referred to as

"The Hill." If one visited Santa Fe, it was "I've got to get back up The Hill." A scheduled bus ran between Los Alamos and Santa Fe. It was considered inappropriate by security to miss the return trip. Underlying this reticence to mention Los Alamos was the pervading atmosphere of security. Round the clock, sentries manned guard gates on all approaches to the town. Official badges were inspected when people entered and left the area. Personal guests had to be approved by the security office merely to enter the town. Off-site, the famous scientists had pseudonyms to conceal their true identities. Neils Bohr was Nicholas Baker; Enrico Fermi was Earnest Farmer.

Everywhere the Army presence was apparent. Soldiers in uniform manned the guard posts and ran the motor pool and the communications facilities. The Army engineers did all the heavy construction. But the Army was also a valuable personnel resource for the Laboratory. Young scientists summoned for service by their local draft boards were diverted into the Special Engineering Detachment (SED), some ending up at Los Alamos. Although they were housed in barracks and sometimes mustered for parade, they worked on a par with their brother scientists inside the fences at the Lab.<sup>1</sup>

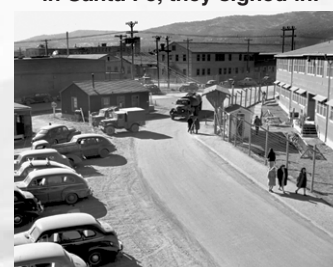
To the scientists working at the Lab, their quasi confinement was an understood security necessity. Besides, they were too busy working to be much concerned by the restrictions. But for wives who had no employment at the Lab, it was different. Some developed acute cases of



**Newly hired Los Alamos staff arrived at Lamy, New Mexico.**



**At 109 Palace Avenue in Santa Fe, they signed in.**



**Once at the tech area in Los Alamos, they spent long hours applying the new concepts of nuclear physics to a usable weapon.**

"Hill fever," a condition usually not part of the medical lexicon and one not always appreciated by a husband mired in a resistant technical analysis. People were sometimes under stress, and release by drinking was a common indulgence. Although alcohol was off-limits on a military base, that was no barrier to the ingenuity of the

<sup>1</sup> After the war, some of the SEDs returned as staff members—for example, Jay Wechsler helped greatly in the engineering of the first hydrogen bomb test. In the Theoretical Division (T Division), Bill Lane was an SED holdover. He was the slender living memory of the group that had done detailed numerical calculations of the fission bomb using the most modern IBM equipment of the time. From the higher military ranks, Colonel Ralph Carlisle Smith, Smitty in his civilian reincarnation, was a welcome holdover. Although on the organization charts he headed the newly formed legal and document division, he was essentially the unofficial chief of staff of the new Laboratory director, Norris Bradbury, who had replaced Oppie.

staff. It was reliably reported that one of the well-known landmarks on The Hill was useful in giving directions as in “To get to the new cafeteria, go two blocks past the bootlegger’s and turn left.”

After he had saved the pines on the mesa from the general’s bulldozers, Oppenheimer achieved his second great administrative accomplishment: the particular security structure within the Laboratory, quite different from the standard need-to-know compartmentalization desired by General Groves. Oppie insisted that there be no security barriers within the Lab. Essentially, every staff member had “need to know.” Groves gave in; Oppie won out. It was a wise victory. It fostered an essential spirit of camaraderie, which resulted in cooperation even under the trying work conditions at the Lab and the daunting challenges of nature. All that made for a great laboratory.

## To Build a Laboratory

How did the great laboratory that made the atomic bomb come into being on an isolated mesa on the eastern shoulder of New Mexico’s Jemez mountain range? From his early years of vacationing in the area, Robert Oppenheimer knew of the exclusive Los Alamos Ranch School on that mesa. In the site selection process, the Ranch School was probably his secret choice for his nuclear weapons laboratory. So on a brisk day in November 1942, Oppie himself, accompanied by Ed McMillan from Berkeley and Major John H. Dudley of the Manhattan Engineering District, visited the site incognito. The residents at the school wondered what these strange visitors were about. It was wartime; perhaps some of the faculty had a potent premonition. In fact, this visit meant that the days of the school were at an end. The school was to be

taken over for the war effort.

Oppie made the decision. Here was the place. It was a felicitous choice. The sun sank slowly behind the Jemez range. The shadow of its ridge crept sedately eastward along the finger of the mesa, dulling first the base of the ponderosas while their crowns for a while longer remained enflamed in orange gold. In the years ahead, this daily ritual was to contribute much to the essential spirit of the Laboratory.

Oppie’s choice was ratified by General Groves. He arranged the purchase of the Ranch School for \$440,000. A contract was drawn up with the University of California as the legal entity responsible for the Lab. Initially, it was a responsibility in name only, but as the years went on, the University became an important element in the workings of the Laboratory. The contract continues in force to this day.

The Ranch School was no more. Now it was time to build the Laboratory on the site. The first priority was to assemble the staff. Originally, Oppie thought he would need three hundred—he ended with about three thousand. With Groves’s inspired selection of J. Robert Oppenheimer as Laboratory director, staffing of the Laboratory clearly was to be successful. A sure accession was Edward Teller. He had been captivated by Oppie’s quick mind, one as fast as his own, and he even envied Oppie’s universal erudition, more varied than his. Teller hoped to be the leader of T Division at the Laboratory. He also hungered to continue exploration of the thermonuclear bomb, the Super. It was Teller who had brought that possibility to Oppenheimer’s attention during the 1942 summer study at Berkeley. There, Oppie and Hans Bethe had enthusiastically joined Teller in divining the initial concept of that weapon, so fascinating in its scientific complexity. Edward was to be disappoint-

ed in both these hopes.

The next and the essential convert was Bethe. He was at the Massachusetts Institute of Technology (MIT) Radiation Laboratory in Cambridge, working on the radar devices that had won the air war over Britain. Oppie’s magnetic aura captured Bethe. Other senior scientists were not so easy to recruit. Oppie’s great personal magnetism was not capable of catching the pragmatic experimentalist, Isador Isaac Rabi. That Columbia University Nobel laureate to be (1944) reasoned that the nuclear bomb would never be ready in time to be a factor in the war, whereas radar had already demonstrated its potential to make a meaningful difference. Rabi only consented to be a part-time consultant, an elder statesman to the younger master scientists, and a valuable representative of the Lab on the national scene. But many other people from the MIT Radiation Laboratory, particularly Victor Weisskopf, came to Los Alamos. Bethe and Weisskopf were to be leader and deputy of T Division.

With Bethe and Teller aboard, recruitment snowballed. Oppie could not discuss the nature of the project at Los Alamos, but he could promise association with those two scientists. It was wartime, and the enemy was real, powerful, perceived as evil, and threatening. Not only patriotism, but also concern for the concepts of civilization itself, motivated the scientists to accept without question the unrevealed and highly secret work. But they could not have dreamed of its eventual worldwide impact.

Personnel and equipment were gathered from the university campuses and industrial laboratories around the country. First to arrive at Los Alamos was Robert Wilson’s Princeton group because their project, the isotron for separating uranium isotopes, had been abandoned. Richard P. Feynman was one of the group.

Feynman's addition to the later formed T Division, with Bethe and Weisskopf, made a remarkable triplet combination.

It was Churchill himself who approved the transfer of a star group of scientists from the British atomic project to the safety of the United States to aid in the Manhattan Project. They arrived in November 1943. The leader of the British mission was James Chadwick, that careful nuclear experimentalist who patiently pursued the neutron. For that discovery he had received the Nobel Prize in 1935.

The British mission included theorists Rudolf Peierls, Geoffrey Taylor, T. H. R. Skyrmes, and Klaus Fuchs. Peierls, a well-rounded theoretical physicist, was a Berliner, assistant to Heisenberg, working in England when Hitler expelled the Jews from the universities. Peierls wisely stayed in England. When the war came, he was eager to help his adopted country. Working on the British atomic bomb project, he had already correctly estimated the order of magnitude of the critical mass of uranium-235 for a fast fission weapon, which Heisenberg had very badly missed. That error from so capable a physicist and his associates passeth understanding. It precluded the possibility of success in the German atomic weapon program.

Geoffrey (G. I.) Taylor, the ace hydrodynamics professor, combined experimental experience, intuitive theory, and clever classical mathematics to achieve a more thorough understanding of this important aspect of weapon design than any of his Los Alamos colleagues. Klaus Fuchs, of course, was the perfect espionage agent of the scientific era—capable in his science, universal in his technical interest, and morally motivated to aid the communist cause. He became involved in every aspect of weapon design, the thermonuclear Super, as well as the two types of fission weapons. He was easily accepted as a

valuable coworker, which of course he was; no one ever dreamed that he was a spy. Actually, the United States had accepted the U.K. security clearances for the British mission, and so made no independent investigations. It is doubtful, however, that they would have detected the potential spying of Klaus Fuchs.

Other experimentalists rounded out the British mission: William Penney, P. B. Moon, Egon Bretscher, Otto Frisch, and James Tuck, among others.<sup>2</sup> It was a year later that the Canadian group, George Plazek, Carson Mark, Bengt Carlson, and Max Goldstein appeared at Los Alamos. They arrived too late to take part in the main work on the fission bombs, but they stayed on after the war and made important contributions in neutronics and in the computational capability of the Laboratory.

In building up the Los Alamos staff, physicists and chemists mysteriously disappeared from their universities. Bewildered colleagues they had left behind sometimes later joined them on the mesa top. Some of the recruits were the following: from the Columbia University physics department, Norman Ramsey, Julian Askin, and Bernie Feld; from the University of California, Eldred Nelson, Robert Christy, and Robert Serber; from the University of Chicago, Nick Metropolis, Stan Frankel, Harold Agnew, and Harold Argo. Somewhat later, in 1944, after the pile under the stadium stands at Chicago had been operating satisfactorily, Enrico Fermi, with his exceptional assistant Herb

<sup>2</sup> While the rest of the mission returned to England after the war, Tuck, after a brief sojourn in the mother country, came back to Los Alamos for the rest of his life. Tuck was a lively, outgoing person, worthy of his namesake, the cleric companion of Robin Hood. To him physics was a joyous adventure or it wasn't good physics. He started the controlled fusion program at Los Alamos, inevitably named the Sherwood Program.

Anderson, joined his European expatriates Bethe and Teller. From the University of Wisconsin, Stan Ulam, a first line Polish mathematician, was recruited by his friend and supporter Johnny von Neuman.

Initially, a gun device was deemed necessary for the assembly of a fission weapon. Therefore, early on, the Los Alamos Laboratory initiated an ordnance program. William S. (Deak) Parsons, a clever, inventive Navy captain, was chosen as head. He was aboard the B-29 aircraft that dropped the bombs over Japan. So was young Harold Agnew, who later became the third director of the Los Alamos Laboratory.

We may set the birthday of the Laboratory at April 1, 1943. Even before that date, construction started at the site. More important, top purchasing priorities for equipment, far more valuable than cash in wartime, were obtained. The university scientists and the Laboratory connived on the transfer of experimental equipment to the site, equipment paid for by the government, more to observe the legal niceties of our procurement system than as a condition of sale. So it was that the Harvard 36-inch cyclotron, disassembled in pieces and parts, arrived at Los Alamos in mid-April with Bob Wilson from the canceled Princeton project as its shepherd. By June, it was operating. The two University of Wisconsin Van de Graff generators were commandeered. The 4-million-volt "long tank" was brought online in May and the 2-million-volt "short tank," in June. Physics measurements could then be made, for example, on the crucial number of neutrons emitted per fast neutron fission. Only specks of uranium-235 and plutonium-239 were available for that measurement, but the experiment was successful. Apparently, nature loves a fission chain reaction, for that number was found to be considerably greater than

1, the theoretical minimum required.

## The Master Scientists

To understand the contributions of theorists to the many-faceted work of the Los Alamos Laboratory, one must understand the relationships of four outstanding scientists: J. Robert Oppenheimer, Enrico Fermi, Hans Bethe, and Edward Teller. Though immensely talented and wise in physics beyond their years, they were still relatively young. Teller was born in 1908; Bethe, two years earlier; Oppenheimer, in 1904. Fermi was the oldest of the group, not only in age—he was born in 1901—but in the respect they rightly accorded him. He had already received his Nobel Prize in 1938. But to the youngsters in their twenties, the main workforce in the Lab, these four men were indeed the wise old souls, though only in their thirties and forties. Together at Los Alamos, almost from the starting date of the Laboratory in April 1943, these four men who had known each other for some years, who instinctively almost innately understood each other, now worked closely, enthusiastically toward the development of nuclear weapons—they were a family of scientific brothers. Much later, the tragic Oppenheimer affair in April 1954 ripped the family apart forever.

The lives of these scientists were sternly contoured by the great changes that occurred in the twentieth century. That was the time of a painful metamorphosis of Western civilization. Starting from the comfortable assurances of the Victorian era as a primarily agricultural society, the transition progressed uncertainly toward the promise and the perils of the new, predominantly industrial society. In the process, the two great world wars were fought, the outflings of a culture trying almost everything to find a satisfactory accommodation. Bethe and Teller, too young to serve in the

military of World War I, escaped the grinding up of a whole generation of men, but they nevertheless experienced real poverty, privations, and hunger in war's aftermath.

Characteristically, Enrico Fermi, temperamentally of earthy Italian peasant stock, passed unaffected through that war.

But while the larger society was convulsing in this cultural transformation, the scientists' own smaller universe of physics was reveling in the scientific revolution caused by the two new central theories of the twentieth century, relativity and quantum mechanics. Whereas the scientists had been victims of the societal changes, they were joyous participants, indeed significant contributors to the new science. At different universities from 1926 to 1933, these young men, instructed by the grand old professors—Bohr, Pauli, Sommerfeld, and Born—used the new scientific disciplines to solve problems in atomic, molecular, and solid-state physics. By day they worked hard and ably at their craft. Compared with classical mechanics, quantum mechanics appeared abstract, nonintuitive. But so simple the premises of that new theory, so universal its application! One equation, the Schrödinger equation, contained almost everything needed. But that equation concealed remarkable subtleties. It took a newly developed nonclassical intuition to penetrate those subtleties. Then easily the new theory served up quantitative results on problems untouchable by classical theory.

These were young men, not yet married, still not settled down. Their whole lives were just opening up before them. They had that zest, that wild joy of living, only partly satisfied by their daytime scientific work. By night they gathered with colleagues and an occasional younger professor in groups where the traditional carousels of youth were overlaid with

something deeper, more . . . the wonderment at the mysteries of their craft. Free, however, from the discipline of their daytime work, they would be expansive, philosophizing about the innate quality of Nature and their own place in her grand scheme. The ever pragmatic Enrico Fermi did not so indulge.

Although the young scholars did not fully recognize it, the dark shadow of Adolph Hitler was lengthening over Europe as the sun was setting on the Weimar Republic. On April 7, 1933, Hitler's captive Reichstag passed the "Law for the Restoration of the Civil Service." These mild appearing words meant simply that Jewish faculty members were to be expelled from the universities. The young scientists' bright dreams of fulfilling careers turned into nightmares. Edward Teller and his wife Mici were Jewish. Hans Bethe was brought up as a Protestant—German law defined him as Jewish. Fermi was a Catholic, but he married Laura Capon, a beautiful, spirited, intelligent young Jew. All three families realized that they had to escape from Europe while they had time. By different paths, at different times, with varied help, anxieties, luck, they all found places in America. When the Manhattan Project started, they were available and well prepared for its demands.

Of the four outstanding scientists, three—Teller, Bethe, and Fermi—were European refugees from dictatorships. They had felt the personal degradation and moral corruption of totalitarian regimes. Oppie was an American intellectual, but he had done postgraduate work under Max Born in Europe. All of these men were not merely great scientists; they were extremely complex individuals. Contrary to the popular stereotype of a scientist as an expert in his field but naïve in other disciplines, and particularly in practical matters, these remarkable men were multiphasic in



their capabilities. Their stern and thorough training in science equipped them to analyze problems in other areas of their interests and, where appropriate, to apply quantitative methods of considerable sophistication in their solutions.

The three Europeans had been together at Columbia University since 1939, all living in Leonia, a pretty New Jersey suburb a few miles across the Hudson River from New York City. Among their friends and an essential part of their intellectual community were Joe and Maria Mayer and Harold Urey. Urey was the director of the Special Atomic Materials (SAM) Laboratory of the Manhattan Project, where the fundamental work was done on the separation of the uranium isotopes. He was already a Nobel laureate for his discovery of heavy water containing deuterium, the rare mass 2 isotope of hydrogen. Joe and Maria Mayer were well known to the three European scientists from prewar days. Later in 1963, Maria, aided to a significant extent by her good friends Teller and Fermi, was to get the Nobel Prize in physics for her work on the magic numbers of nuclear shell structure. The group often carpooled together, driving to Columbia University, where they would need only one parking space. When crossing the George Washington Bridge to New York from Leonia, their car contained four current or future Nobel Prize awardees.

Bethe and Teller arrived at Los Alamos almost at its beginning. Fermi followed in 1944, as soon as he finished his work on the nuclear pile at Chicago.

**Oppenheimer and Groves.** Oppie, of course, was the technical director, but General Leslie Groves was in overall charge of the Manhattan Project. General Groves was a capable no-nonsense engineer, but he was incapable of appreciating the scien-

tists' method of probing the unknown territories leading to the realization of the fission device. He had a timetable driven by desperation, and he was determined to meet it. His skill was in known engineering projects, in programming to meet schedules and budgets. His recent practice was in military command, where orders were obeyed, not questioned, and well-paid contractors, who understood engineering schedules and budgets, were acquiescent to his will or sometimes even to his whim. He would have preferred to do without the scientists, the eggheads who could not seem to get on with the work. Oppie alone could not have moved the general to adopt the exploratory methods of the scientists in place of his proven procedures of getting a job, albeit a well-known one, done. But Oppie knew the power of his scientific associates. He trusted their capabilities. Skillfully, he used the necessity of accommodating to their methods to sway the general away from his accustomed path.

Rather than Groves channeling the scientists to his methods, it was the scientists, through Oppie, who channeled Groves's efforts for their support. Groves became not the project leader directing efforts, but the project enabler who helped the scientists do their job. Capable an engineer as he was, Groves never realized that he had been co-opted to the scientific task. To the end of his life he really believed that he had made the atomic bomb.

**Enrico Fermi and Hans Bethe.**

By 1941, Fermi was already building a nuclear reactor pile in the basement of Pupin, the physics building at Columbia University. That pile could not achieve criticality because of the neutron-absorbing impurities in the graphite blocks then available. Fermi was the most complete physicist of the brothers, perhaps the most complete physicist of the century. He had



J. Robert Oppenheimer



Enrico Fermi



Edward Teller



Hans Bethe

a profound understanding of his subject. In theoretical physics, quantum mechanics was as natural to him as classical mechanics. He said that, after completing his *Reviews of Modern Physics* article on the quan-

tum theory of radiation, he understood it so well that its extension to his theory of beta decay just sprung unbidden into his brain. Both an experimentalist and a theorist, he had the feel of physics in his hands. He



**Fermi socializes with coworkers and friends at a party.**

was the supreme practical problem solver; for example, he developed the age theory of neutron slowdown enabling simple calculations of nuclear reactor performance. At parties, he was the best riddle and puzzle solver of the crowd. The high sport at these events was to stump Fermi. Modestly, he explained his success, not as due to his particular skill or intellect, but simply to his practice at the art. After a while, he had seen most of the usual puzzles and had developed the knack of solving the various types.

Bethe was lecturing from his classic articles published in *Reviews of Modern Physics* (1936–1937), summarizing and systematizing all that was worthwhile knowing about nuclear physics at that time. If one knew what was in that “Bethe Bible,” one need not bother reading any literature previously published. Most theorists, as well as the experimentalists, told Bethe about their work before it was published. His brain absorbed it all, refined the information, organized it, and stored it permanently, but with instant recall. In addition to having formidable manipulative technique, Bethe had a deep understanding of

physics, accompanied by a unique knack of finding the simple way for analyzing a problem. He showed this quality in his 1946 work explaining the Lamb shift in the hydrogen atom 2s and 2p fine structure. By using a nonrelativistic approximation and by staring down some daunting infinities, he captured the essence of this quantum electrodynamic effect before anyone else did. Bethe was a genial father figure to his young associates. They could always find him a sympathetic listener to their troubles, personal or scientific.

**Edward Teller.** Edward Teller was the most convivial of the three Europeans. It seemed as if he did not work on physics—he talked out physics. He almost had to have a collaborator. For Fermi, physics was in his hand; for Teller, it was on his tongue. This story is told by one of Edward’s students who knew Fermi as well. When the student first met both of them soon after they came to the United States, Fermi’s English was so heavily disguised by his Italian accent that he was barely understandable. Teller, on the other hand, spoke well, although he sounded obviously quite Hungarian. Several years later, the student found that he could understand Fermi easily, but Teller’s accent had improved not a bit. The student reasoned that Teller was always talking while Fermi really listened. When the student told this theory to Teller at a dinner party, forty years later, Teller could only laugh a huge Hungarian laugh.

Edward Teller was one of those scientists and human beings who are easy to be with but hard to classify. His outstanding scientific characteristic was his creativity. Before the war, he had already made significant contributions in molecular and solid-state physics. But it was his breadth of comprehension, rather than any specific contribution, that gave him his

effectiveness. Putting diverse aspects together, Teller was able time and time again to hit upon the unusual synthesis; that was his creative mode. Even close colleagues, however, complained that he had too many ideas; they could not work on all of them. With the good ones, there were sure to be some poor ones to be filtered out. But the faint hearted, who never propose anything that turns out to be wrong, rarely propose anything significant. Edward Teller, even by his critics, was never assailed as being faint hearted.

Concealed behind his very clear physical insights was his full mathematical competence; he could, when he wished, work out in detail his inspirational concepts. However, as a teacher in his classroom, Teller



**At Fuller Lodge in Los Alamos, Teller talks with Julian Schwinger (left) while son Paul enjoys his favorite ride.**

would eschew the mathematical derivations and, instead, develop interesting physical reasoning to prove a point. As a trivial example, in his graduate lectures on mechanics, he obtained the resultant of two forces, which leads to the familiar parallelogram result not by the addition of the  $x$ - and  $y$ -components of the forces but by qualitative symmetry arguments that showed the inevitability of that conclusion.

But it was outside the classroom, in one-on-one or one-on-two conversations, that Edward excelled. With his peers, he would stimulate; with the fully developed young scientist, he would inspire; with the fresh student,

he would educate. He helped them all.

Edward Teller was one of a group of Hungarians sometimes called the Martians because of their apparently unearthly intellectual abilities. They all came from a small neighborhood in Budapest: Leo Szilard, Teller, Theodor von Karman, John von Neuman, and Eugene Wigner. They knew each other from their youth. It was Szilard and Teller who approached Einstein to get his backing for the atomic bomb project and through him eventually to get President Roosevelt to approve it. The childhood acquaintance and later mature interactions of these “Martians” contributed greatly to their effectiveness during the Manhattan Project.

#### **J. Robert Oppenheimer.**

Oppenheimer was the director of the Los Alamos Laboratory, chosen and championed by General Groves. With Oppenheimer rested the responsibility for the scientific and technical aspects of the project. For his authority he possessed, through the general, the key to unlimited funds and, what was more important in the wartime economy of scarcity, the highest priorities. But what authority over the minds of those European scientists did he have? Oppie, in his own right, had made significant contributions in physics. He was in part responsible for the Born-Oppenheimer adiabatic approximation of molecular and solid-state physics, which separated the treatment of the degrees of freedom of the rapid motion of the electrons from those describing the more stately motion of the massive nuclei. He also developed and then applied the Oppenheimer-Phillips process to the collisions of high-energy deuterons with complex nuclei. In that process, the proton in the deuteron could not penetrate the Coulomb barrier of the target, and so it sat by as a spectator particle, while its accompanying neutron engaged in

the reaction. Two of the earliest papers on the theory of black holes were his. Had he not been diverted from his career as scientist by his service at Los Alamos, he might have gone very far in that field.

Oppie possessed a very quick and facile mind. He was able rapidly to absorb almost any subject, and since his interests extended far beyond science, he became learned as well in philosophy, literature, and language.



(Left to right) Dorothy McKibben, Oppie, and Victor Weisskopf at a party on Bathtub Row.

His wide-ranging erudition surprised, even delighted, his colleagues but set him apart—he carried himself on a higher plane. His very quickness also enabled him to understand the varied work of the Laboratory staff; sometimes, he would comprehend even more than did the originators. That faculty made him extraordinarily well suited to direct work in the multidisciplinary problems of the Los Alamos program. Furthermore, with his vast command of technique, he could often integrate scattered work into a sophisticated and powerful mathematical formulation. Despite his mental gifts, in the judgment of the European trio, his accomplishments in science were not up to his potential, and they were competent to make such judgments. So, how could Oppie, as director of the Laboratory, make these men work together with him?

Consider the following: Before the war, Oppie was one of the intellectual elite in the company of his students and colleagues at Berkeley. He was

harsh in his criticism of their work, to some extent belittling their efforts. After the war, as director of the Princeton Institute for Advanced Study, he continued the same elitist attitudes. But at Los Alamos, Oppie was different. He was obviously capable intellectually of recognizing that, in the Los Alamos setting, his attitude toward the technical staff must be different from that he showed to his students. Not only must he continue to be a leading scientist, but also he must be an effective administrator and much more. He remained above the staff but not distant from them. He would understand and kindly appreciate rather than criticize them, and they loved him. In the company of Fermi, Bethe, and Teller, he was in no position to denigrate their abilities. Recognizing this, Oppie became their facilitator—he provided the opportunity and the atmosphere for them to do their best work. And he could integrate their work into the overall Lab program. He was the coach and the strategist of his team of star physicists. Organizationally, he realized that it was inappropriate to place them all in one theoretical division with the resulting question of who worked under whom. To Bethe, he gave the name and position of head of T Division. Fermi had his own Division, “F” Division, of course. By temperament, Teller would not permanently be pigeonholed anywhere. He was permitted to float, nominally in T Division, but detachable for special assignment anywhere, either at his own selection or by Oppie’s direction. Edward was a large man with many and large ideas. He would prove to be one of the most creative people at the Lab.

#### **Role of the European Scientists.**

What did these three European scientists bring to the Laboratory during and after the war? Obviously, they

brought their broad scientific knowledge and great scientific talent. They were also problem solvers, comfortable with entering new fields and venturing on untrodden grounds—just what was needed to develop the fission bomb, a device previously imagined only in science fiction. To their tasks, they brought the discipline of hard work and the habit of persistence, not the dogged persistence following a single path toward a destination, but the persistence to try path after path until a broad highway opened up to their goal. There was an enduring legacy these scientists left to Los Alamos. It was the love of science, the enthusiasm of working in science, and the confidence that science was the tool of choice for developing the new industries needed in peace as well as war—to serve the needs of humankind. Moreover, they could inspire others to have the same faith.

As no others, they knew their opponent—the two sides of Germany. They knew the background and the richness of the body of German science, and they knew the genius of Werner Heisenberg. They also knew at first hand the perversion of values that the Nazis had brought to their countries, and the power of an aroused and united nation whose imagination had been unfortunately captured by its persuasive but demonic leader. Therefore, they worked with conviction spurred on by a terrible fear.

#### **Inspiration of the Scientific Staff.**

To the more junior scientists (not necessarily the younger scientists) at the Lab, Oppie acted with regard, care, and understanding. He was their charismatic leader, and they all but deified him. His direction and their combined cooperation made for a great team. Together with the European science masters, they played above their individual capabilities.

They succeeded. They made the fission bombs.

A unique spirit of cooperation and camaraderie among the young staff scientists pervaded the Laboratory. Oppie's leadership was a part of it, but there was more. These young people had left home and family behind. Joined in a great enterprise, they could not afford to dissipate their energies in divergent pursuits, nor did they have much opportunity to do so. They were isolated, confined on The Hill, restrained by security measures. The town was entirely dedicated to their work, no way to escape that fact. Their companions in the few hours off from work were usually coworkers. Such concentrated intimacy they had never known before. Actually, their coworkers were like an extended family. And what coworkers! Some scientists worked directly under the

tion, the work itself or the methods to be developed required good science, and the best scientists were close at hand to advise, to inspire, or just to listen. Here they were not students, working alone on a small, detailed topic in science leading to a doctorate. Here, for most of them for the first time, they were in large groups working together by using big science, big facilities—a brotherhood of effort, companions in accomplishment. But behind all the deadly seriousness of the task were a spur and a satisfaction. For most of the participants, the Los Alamos experience was the highlight of their lives. The work was well done.

#### **War Work at Los Alamos**

When Los Alamos was founded in



**At the Laboratory, the hours were long, and the work hard. The weekly colloquia, one of which is shown in this photo, were stimulating and intense. They are still remembered.**

European masters, who here were not Professor Fermi, Dr. Teller—just Enrico and Edward.

The hours were long, and the work hard. Sometimes it was routine, but often it was science at the edge of their capabilities. They found that necessity had as brother, opportunity. Though the work had direct applica-

April 1943, fission bombs were already known to be feasible, at least in principle. Two avenues to their production were using the separated uranium-235 isotope to be produced by Oak Ridge or using plutonium-239 to be produced by the Hanford nuclear reactor. But all plans were largely on paper. Fermi's pile at Chicago had just

demonstrated the first manmade neutron chain reaction in a critical assembly. Hanford and Oak Ridge were still “to be’s.” How to design a bomb with the constituents available in only small amounts and with some critical material and nuclear properties largely unknown? That was the Los Alamos problem. Getting the required plutonium and the separated uranium-235 isotope—that was the problem of the rest of the Manhattan Project.

Meanwhile, news from the battlefield gave a terrifying urgency to the tasks of the Laboratory. The Battle of Stalingrad, after terrible slaughter, had just ended in Soviet victory on February 2, 1943. It was not yet recognized as the real turning point of the European war. In the Pacific, the island-hopping campaign had barely begun. On the ground in North Africa, the U.S. troops were about to experience their first combat defeat at Kasserine Pass. How to proceed at the Laboratory? There was no time for the conventional wisdom. What was needed was a new wisdom chased by haste, built on scientific insight, ingenuity, and luck, helped by nature’s guiding hands!

The goal of the Laboratory, the development of the fission bomb, was clear, and the basic scientific concepts were known. But the detailed implementing pathway was vague. Urgency dictated that almost everything be done at once. To General Groves, this process could only bring chaos—no firm priorities, no observed schedule, no PERT charts. He felt beset by more and more requests for strange pieces of apparatus and for usually unavailable materials. He did obtain them all; that was his genius.

But the scientists were on familiar ground. In their research, as usual, nature was in charge, but not always



**The building in which criticality experiments were carried out was called the kiva, a term borrowed from the Pueblo Indians. Here, the kiva is photographed from an Indian cave in the nearby wall of the Pajarito Canyon.**

clear and apparent in her direction. In keeping with their background, the scientists organized in traditional academic manner, by topical disciplines. They created divisions in physics, chemistry, explosives, mathematics and computation, theory, and others as well. Senior professors headed each division, with younger persons, much like students, guided and taught by them. Not only was the staff expected to do specific jobs, although that was their primary responsibility, but they were encouraged to learn and also to innovate. Because at Oppie’s insistence there was no security compartmentalization, the senior scientists knew, in some sense, all the work at the Laboratory. The junior scientists were also informed but to a lesser extent. Therefore, everyone could contribute ideas; everyone could join

in their evaluation. Not altogether surprising therefore, creative contributions, out of the mainstream of their work, were made by Jim Tuck—a central idea in the high explosives of the plutonium weapon; by Seth Neddermeyer—in the assembly of the plutonium weapon; and by Bob Christy—for a crucial idea that rescued the plutonium weapon from potential disaster. Moreover, pure science too was cultivated if it had the possibility of mission application—for example, Walter Koski experimented with the collision of two high explosively driven jets to see what high temperatures could be achieved. Teller improved the theory of radiation transport at temperatures attained in nuclear weapon explosions. He developed a practical statistical treatment of the very many spectral lines in the transport medium under those conditions. This was an adaptation of Wigner’s work on neutron transport in nuclear

reactors, where many neutron absorption lines exist and must be considered.

New facilities were built, almost overnight, erupting from the chaos: a critical assembly building, plutonium-handling laboratory, high-explosive range, ordnance firing site, sheltered canyon location for a nuclear reactor, and another one for experiments with very high intensity radioactive sources. Finally, when necessary, the scientists seconded as engineers, and very capably too—there was no holdup or misunderstanding in transition as would occur in ordinary industrial practice.

The entire informal, almost slipshod-appearing organization fostered the nascent good feeling and cooperation of the staff. General Groves did not appreciate the character of this

organization, but strangely, he had confidence in Robert Oppenheimer. Although to Groves Oppie appeared disorganized, actually Oppie could keep up with everything, understand everything—he had all the inputs in the data bank of his brain, and there he could organize them and, as necessary, rapidly reorganize them to take into account new facts of nature. Oppie worked magic with his people, and they worked their hearts out for him, performing way over their heads. That dedication was what made the Los Alamos Laboratory.

The scientists worked in a special type of mental denial. They knew the terrible destructive power of the nuclear explosive they were devising. They subconsciously could not call it a bomb. Instead, they called it, in general, the device, or the gun gadget (the uranium-235 bomb) and the Christy gadget (the plutonium bomb). The plutonium weapon was named after an Oppenheimer protégé, Bob Christy, who very late in the program, proposed an idea that rescued that device from apparent failure. The characteristic time scale of the explosion was never referred to in its scientific nomenclature, but was called a shake, obviously short for the flick of the lamb's tail. The characteristic cross section for the nuclear processes in the explosion was called a "barn," signifying that it was an easy target. Despite this levity, the scientists worked hard and happily at their mission, although at heart they were appalled at its potential for destruction. Also, they worked with a terrible urgency, for the famous European expatriates who worked at the Laboratory knew the capability of the German scientists they had left behind. They feared living in a world where Hitler would be the first to have the bomb.

### **The Equation of State: An**

**Example.** To illustrate the character

of the work at the Laboratory, let's choose one problem of the many. The fissile material, initially of course in a subcritical configuration, is to be assembled by rather violent, explosively driven motions into a supercritical state. To calculate this motion, the equation of state of the various materials in the weapon was needed at pressures and densities starting from their normal state and increasing during the assembly of the weapons to conditions surpassing those at the center of the earth. Then, at the nuclear explosion, conditions would approach those at the center of the sun. Nicholas Metropolis was assigned to the problem.

The "sun" part was the easy one. Astrophysicists had already understood in principle how to calculate that equation of state. Under those conditions, all materials form highly ionized plasma, and the pressure is almost entirely due to the free electrons acting as a perfect gas. Guess the number of free electrons, guided by the astronomers' formula, and you have the answer to some reasonable approximation. Los Alamos is still improving this approximation almost 60 years after the work of Metropolis.

The highly compressed state just before the nuclear explosion was treated by Metropolis, with the help of Julius Askin, using the Thomas-Fermi model of the atom. Of course, molecular or crystal structure had been squeezed out of the material by this time. The model, designed for isolated atoms, had to be altered to account for the pressing presence of neighbors. So, the boundary condition on the electron density had to be changed from one vanishing at infinity to one that is continuous across the boundary to the next atom. Further tinkering was necessary to adapt the model designed for zero temperature to the very considerable temperature of interest.

At the low pressure end of the

range of interest, one had experimental measurements by Percy Bridgeman at Harvard and some further data from Hugoniot measurements in material shocked by high explosives. The gap between these low-pressure points and the Thomas-Fermi results was forbidding. Someone, possibly Teller, suggested using data obtained by the behavior of seismic waves traveling through the earth's core. That expedient would give one intermediary point, valid of course only for iron-nickel alloys, not for uranium or plutonium, under pressures and densities like those at the center of the earth.

For plutonium, however, there was the additional difficulty that only microscopic amounts of that substance were available until too late in the project to make the Bridgeman-type measurements. Still, something had to be done. In wartime Los Alamos, pessimism was recognized as a word but not accepted as an attitude or even as an emotion. With the minute quantity of plutonium they had, the solid-state physicists and chemists did the best they could, and they did very well. They determined the crystal structure, as well as the density and even some alloying properties. All that did not help much in determining the desired equation of state.

The history of what actually happened thereafter is probably not available. At that time, records were not kept, and the people involved are no longer around. But we can imagine what might have occurred. Perhaps Oppie called a group together: Fermi, Teller, Segré, Metropolis. He asked them to make their best estimate of the values required. After some brilliant but disorganized discussion, Fermi said, "This is what we shall use; it will be good enough for our application. Let's get on with the job." Indeed that "Fermi feeling" was often the best, if not the only, method for many problems, one of the special resources of the Laboratory.

For this small problem, one of many in the design of the first nuclear weapon, one had to dip into the core of the earth and the center of the sun, assemble results from quantum theory applications and laboratory and high-explosive experiments, and patch all together in a hurry, with a feel for physics and a sprinkling of luck. That was typical work done in wartime Los Alamos.

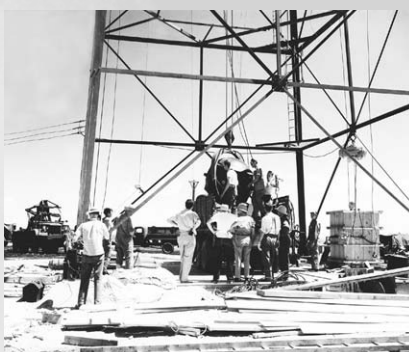
### Trinity: A Culmination and a Beginning

As the work of the Laboratory went on, unexpected successes alternated with heartbreaks. Finally, in the spring of 1945, the gun gadget was ready, its obvious simplicity apparently guaranteeing its performance. The plutonium device was much more complex, but no one really knew whether it would work. Therefore, a crucial nuclear test shot was planned. The test site near Alamogordo, in the southern New Mexico desert, was called Trinity by Oppenheimer. He never satisfactorily explained his choice of name.

In the predawn, dark silence of July 16, 1945, an explosive release of nuclear energy from the fission of plutonium-239 in the Christy gadget caused a supernal flash of light that illuminated the desert at the Trinity test site. The fission bomb was now a reality. Brighter than daylight this flash, with rich promises for the future; darker than midnight, with portents of fear. Observers wearing special dark glasses were able to see the initial flash and the early fireball, confirmation of the value of the time of striving, hope, and heartache in their work. Unprotected eyes looking at the test tower were blinded for minutes, a foretelling metaphor: In the coming age, to see without foresight was to lose sight. The first sensation felt by observers facing the test tower

was the heat of the light pulse, a palpable force. It felt like simultaneous hard slaps on both cheeks. Still not a sound, only visible evidences of the shot. Then, the shock wave raced across the desert floor. When it hit, one could hear the first sharp crack, afterwards continued echoes and reverberations. Eventually, a cloud of bomb debris, desert dust, and atmospheric water droplets lifted off the desert floor and, rising slowly, formed the characteristic ominous mushroom shape of a nuclear burst.

The Trinity shot was the beginning of the nuclear testing procedure as a central feature of nuclear weapon development. Contrary to later test operations, however, Trinity was



The implosion "gadget" is hoisted to the top of the shot tower at Trinity site.

a nuclear explosion test almost entirely devoted to the needs of the scientists who designed the weapon. The military did not hand down requirements for the weapon yield, the nuclear materials, the safety, survivability, and performance specifications that have now become standard operating procedure. There was only one, but an overriding, requirement—the weapon must fit through the bomb bay doors of the B-29 Super Fortress. There was, however, a political desideratum—the test should be made in time to influence the upcoming Potsdam conference. But unknown to the United States, Stalin had considerable knowledge of our nuclear weapon progress through the espionage of his

secret agents, most important among them, Klaus Fuchs. Stalin showed no surprise, although he must have laughed silently, inwardly, when Truman dropped a hint that we had a secret superweapon in the making. The July 16 date set for Trinity, which was the date of the commencement of the Potsdam conference, therefore had an urgency in the mind of General Groves that had nothing to do with the scientific purpose of the test.

Further than the two requirements just discussed, the scientific aspects of the Trinity test were initiatives of the Los Alamos Laboratory, aided of course by the logistics support of the military. The site had been selected by Kenneth Bainbridge, the Harvard cyclotron expert chosen by Oppie as test director. Naturally, General Groves approved the choice of site. The diagnostic experiments were conceived, designed, and executed by the Laboratory. The bomb-firing electronics and even the operating protocols were Lab responsibilities. Academics, only a short time earlier familiar primarily with university laboratories, learned about field operations—operations in a somewhat uncontrolled environment with essentially no second-chance opportunities. The T-Division theorists too learned how to cooperate with the experimentalists, some even going to the field.

After the success of the Trinity shot, the realization that the fission weapon was a reality burst upon the consciousness of the staff. Before the Trinity event, the staff members were too busy with the everyday tasks necessary for the development of the weapon to dwell upon its consequences. They could strive to make it work, while fearing that it would. But after Trinity, the stark reality of success stared deep into their psyches. Each reacted in his own particular way. Learned Oppie, at Trinity, quoted from the Bhagavad-Gita, "Now I am become death, the destroyer of worlds."

Bainbridge, characteristically a realist, said earthily, succinctly, "Now we are all sons of bitches." The rest of the scientists, back at the Lab, were jubilant but worried. They wondered whether the bomb should actually be dropped on Japan.

### War's End: Devolution and Revival

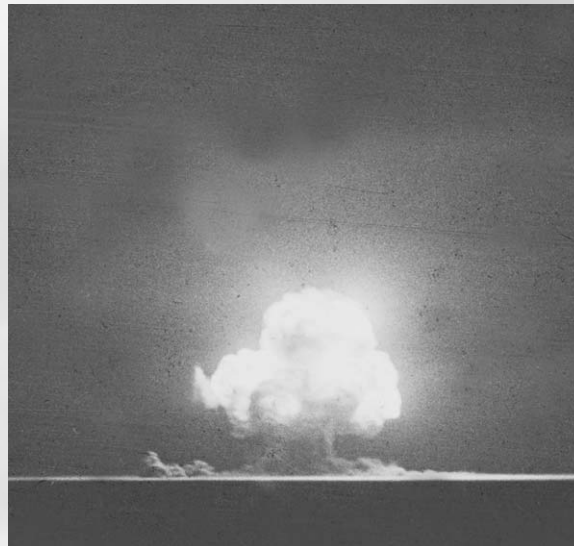
**Achievements.** Two entirely different types of fission weapons were made, the uranium-235 gun type and the plutonium-239 implosion type. One success alone would have been an outstanding achievement. At Los Alamos, the scientists had developed an essentially new mode of operation, the intimate meshing of science with technology. From an idea, the nuclear bomb, and a single new fact of nature, the fission process, to the end use of the finished product—all the steps in between were coordinated by the same tightly knitted community of effort. True, this result was accomplished under the unusual urgency of an overriding national need. In the war effort, no obstacles were permitted. They were only problems to be understood—and once understood, to be solved.

What was understood?—The fission process itself, the processes of the chain reaction, the nature of the supercritical assembly, and the dynamics of the ensuing explosion. Along the path, one needed to understand the chemistry and metallurgy of the various weapon components and the engineering of the assembly of the parts. A critical element was the high explosive in a use for which precision never before achieved was absolutely necessary. Sometimes, a scientist would do engineering, and he was often very good at it—he might even like it. The engineers, for their part, had learned how to cooperate with the

scientists, even though they did not always understand the science involved.

**Dropping the Bombs.** The United States dropped the bombs. Two Japanese cities were destroyed, their inhabitants killed. President Truman's hand had signed the executive order, but his hand was moved by currents he did not control. The debate going on among the Los Alamos scientists about the use of the nuclear weapons had been soul satisfying to them, but that long, hard, devastating war was over.

What would history be like if the bombs were not ready or if the United



The fireball at Trinity site.

States refrained from using them? The bloody battle of Okinawa was a foretelling. The War Department estimated that, in the planned October invasion of the mainland, there would be a quarter of a million U.S. casualties and perhaps one to two million Japanese casualties. Could the wise, constructive peace that actually occurred in Japan be made under such brutal circumstances?

But the bombs were dropped. How much of the current dichotomy in viewpoint about nuclear energy—

uneasy acceptance in the U.S. of its benefits and unreasoning fear of its dangers—is due to its first use in slaughter, however inevitable?

The people of Los Alamos had participated in a unique experience, one that would alter the prospects of the world both in warfare and in the pursuits of peace. When they heard about the two nuclear explosions that had obliterated Hiroshima and Nagasaki and their immediate aftermath in the surrender of Japan on the deck of the battleship Missouri, they realized the import of their work. They realized also that they themselves had been changed. All future experience and accomplishments would be measured

in their own eyes against their Los Alamos achievements. The isolation from their families, the living on The Hill—the beautiful Jemez range in their back yard, the pinked peaks in the sunset of the Sangre de Cristo in their distant view, in contrast with the dust or mud of their unpaved streets and the insubstantial houses in which they sheltered their children—the camaraderie of friends who shared the same privations and delights, the heavy realization that two hundred thousand real human beings had been killed as a result of the microsecond duration of the explosions of

the devices they had created: All this was past and prologue. Their future seemed as rosy as the sunset on the Sangres, but was it not incarnadined as well with the extinguished hopes of the Japanese victims? To the people of that strange city of the hill, their wartime experience had been a singular, slowly evolved epiphany. New sighted now, they were to see a new world.

**Aftermath.** In the autumn of 1945, after successful completion of its



wartime mission, the Los Alamos Laboratory started to disband. Disbanding, however, conveys the wrong concept. It was a diaspora in which the departees formed a web to maintain the intimate relationships that had led to their wartime achievements. As they spread across the nation, they were the missionaries of the new mode of doing science and technology—big science, but soundly based upon fundamental science, linked in a group effort with engineering development, all in support of pragmatic goals in the service of the common good. It was science for humanity's sake, but in an immediate sense, not as an eventual effect. The success of the atomic bomb project conveyed upon these scientists an aura of expertise, which however, outshone their real experience. The federal government placed them on numerous advisory boards and committees not solely restricted to the future development of nuclear weapons and nuclear energy. J. Robert Oppenheimer, for his leadership of Los Alamos, his amazing erudition, and quick understanding in diverse areas of science and human endeavors, was favored as the central member in many of these groups, even those with policy-making responsibilities. Because of this diaspora, the Manhattan Project, particularly the Los Alamos experience, was to illuminate the larger economy. Furthermore, the federal government, which had sponsored the wartime work, now continued to support big science—in government, in industry, and in universities as well. The former Los Alamos staff members were welcomed everywhere. Although the Los Alamos Scientific Laboratory (as it was named later) lost their service, the



**In 1946, two implosion bombs, identical to the one tested at Trinity site and the one that had destroyed Nagasaki, were tested as part of Operation Crossroads in the Pacific.**

commonweal as a whole was well served by the dispersion of Laboratory personnel. With this acceptance in the larger world, the Laboratory itself was not to be abandoned. Some of the departees periodically returned to the Lab. It was possible for a staff member at home in Los Alamos in one afternoon to consult with all three visiting luminaries, Bethe, Teller, and Fermi.

After the war, the great effort that had developed and fielded the two nuclear weapons wound down, mission accomplished. Well done. The future roles of the Laboratory were in doubt. In the euphoria of victory and the passion to return home to resume normal life, most of the scientists were leaving Los Alamos. They believed that their work was done. Not realizing that the way science would be done in the future had forever changed, they were going home, but they were not going back. Though they left the Lab behind, the shell remaining was not destined to fade away. The Lab had demonstrated that a large organization with a mission of great national importance could do what no single university or combination of universities could accomplish. It was historically necessary that the Lab should live. The nation subcon-

sciously recognized this inevitability, although individual scientists by and large did not.

As the Laboratory scientists appeared to evaporate, a stubborn residue remained. The seasoned academics were returning to their universities; other young people were going home to participate in the vibrant postwar economy. Hans Bethe had returned to Cornell, taking Dick Feynman with him. Teller joined Fermi at the University of Chicago. Of

the whole group of talented young physicists only Rolf Landshof, Frederick Reines, and Bob Richtmyer remained. The Canadians, Carson Mark and Bengt Carlson, stayed on as a nucleus of mathematical talent, specializing in neutron transport methods.

Some people stayed because they loved the countryside; others stayed because the intense pressure of the wartime effort was relaxed and they could now pursue science leisurely but skillfully, using the fine facilities of the Lab still largely in place; some people simply had not yet found another place to go. It was slowly to dawn upon these few who remained—and also on the European masters and some others who had left—that the country had given the Laboratory an opportunity, no longer a designated mission, but an opportunity to create an institution using the new big science in the national interest. Los Alamos has been defining and refining this opportunity ever since.

Robert Oppenheimer was replaced by Norris Bradbury, an experimental physicist who had also served as a naval officer. Bradbury considered himself an interim director, a caretaker in the transition from wartime to peace. He planned to stay on but for a short period, less than a year, a matter of

Number 28 2003 *Los Alamos Science*

duty rather than desire. He deliberately kept a low profile. Events, not his hand, were now at the helm of the Laboratory ship. Contrary to his initial plan, he stayed at his post for 25 years.

**Operation Crossroads.** Faced with the diaspora, the remaining Lab core required some defining activity to ensure its success. The 1946 nuclear test operation in the Pacific provided one such focus. Counteracting the fragmenting forces tearing the Laboratory apart was the unifying effect of Operation Crossroads, the test explosion of two atomic bombs over Bikini Atoll in the western Pacific. This operation was much more a military show than a scientific test. The bombs were duplicates of Fat Man, the plutonium Christy gadget tested at Trinity site, the one that had destroyed Nagasaki. Tested, sure to work, their air burst explosion would give little new information about the bombs' internal operation. But some military equipment was exposed to the blasts to measure the effects of nuclear explosions on potential military targets. Thus, the area of nuclear weapon effects was born, albeit in a haphazard and nonquantitative fashion. That was to be remedied in future tests. From the Laboratory, Darol Froman was chosen as scientific test director. A combined Army-Navy force provided the logistic support for the operation—the progenitor of more sophisticated test programs under Joint Task Force Eight, a permanent military organization established for that purpose. Nuclear explosion testing was now established as a central feature of the Laboratory's mission.

Of course, the components of the bombs had to be manufactured, and the parts assembled. That was a job for the Engineering Division of the Laboratory. Here was a unifying task, although not a scientific one, focusing the efforts of a fragmented work

force. In its own way, it was done with the same can-do spirit of the heady wartime developments. Moreover, it was symbolic of a continuing mission for Los Alamos. In fact, nuclear explosion testing was an essential for the rebuilding of the Laboratory. The scattered holdovers now had new hope for fulfilling careers. Darol Froman and Al Graves were role models for others who stayed on—Jerry Kellogg who headed the Physics Division, Eric Jette for chemistry, Max Roy explosives, and Bob Richtmyer for T Division. The renewed Lab could coalesce around these strong men, with Norris Bradbury as the unifying new Laboratory director.

From a weapons viewpoint, the successful explosion of four bombs of the same construction, all giving, so far as was known at the time, comparable yields, meant that the new bombs were a reliable basis for a future stockpile of nuclear weapons. The Laboratory then set about carefully, conservatively modifying and improving the designs of the weapons and systematizing their manufacture.

**The Atomic Energy Commission.** Meanwhile, on the national scene, Congress passed legislation establishing the Atomic Energy Commission (AEC), a decision that placed all nuclear energy programs, including weapons, firmly under civilian control. Furthermore, after some political posturing, Congress confirmed the enlightened and capable administrator, David Lilienthal, former chairman of the Tennessee Valley Authority, as chairman of the AEC.

In Los Alamos, the Army left the region, turning governance over to new civilian agencies. At the Laboratory, the military guards were replaced by AEC-employed civilians in new uniforms. They still rode horseback to patrol the extensive surrounding countryside. In place of the

Army Corps of Engineers, the maintenance and support services for both Lab and town were taken over by the newly formed Zia Company, an offshoot of the McKee construction firm. The doctors in the hospital discarded their rarely donned military uniforms and provided the same skilled and caring medical services with the same stethoscopes and white coats they had habitually used. The same low, subsidized fees applied, but now the physicians were employees of the AEC. Laboratory personnel remained employees of the University of California as before, but somewhat more integrated into the University system, particularly for retirement benefits. Little regarded at the time by the youthful staff, retirement benefits eventually proved to be one of the attractions of the University relationship. New security badges were issued to Lab employees with numbers starting with the letter Z because the Zia Company had the only complete list of residents. The practice of designation by Z number prevails to this day. The town was still closed and resident passes were still needed to enter or leave. The Laboratory itself was still a fenced-in operation with the main technical area near Ashley Pond and varied outlying sites on adjoining mesas. As yet, there was no town government and apparently no need for one.

Budgets now came from Congress through the AEC. Although during the war the military provided essentially unlimited funds to the town and the Lab, the expenditures, while generous, were sometimes made at the wish or the whim of military commanders. But now the AEC effectively gave director Norris Bradbury one overall check to fund the Laboratory. It was always an amount larger than the Lab could sensibly use. The Lab could almost decide for itself what the funds would support. The Laboratory was now ready for a new mission—to use

## Part II: The Dawn of the Thermonuclear Era (1946–1952)

nuclear weapons to secure the peace as they had terminated the war.

### Getting Ready

The development period from the formation of the Lab in April 1943 to the explosion of the plutonium fission bomb at the Trinity site on July 16, 1945, was the fission era. From Trinity to the Mike shot, detonated on November 1, 1952, on Enewetak Atoll in the Pacific, was the thermonuclear development era. More dramatically, these were the atomic bomb and the hydrogen bomb miracles. But when two apparent miracles occur together, there has been no miracle. A causal mechanism must be involved. Some essential culture at Los Alamos must be at work to make both developments possible. Of course, nature was also kind. In what follows, why this all came about will in some part be illuminated.

The fission bomb was made under wartime urgency, when a great nation girded for victory. The greatest nuclear physicists of the time were at Los Alamos, organized under their leader, Robert Oppenheimer, in his finest hour. T Division was led by Hans Bethe—immensely capable, precisely organized—with a brilliant supporting group. In contrast, the first hydrogen bomb was made in a peacetime America, relaxed, reaping the harvest of victory, albeit with the fear of the growing cold war. Norris Bradbury was the unassuming director of the Los Alamos Scientific Laboratory at that time.

Although Los Alamos under

Oppie, with a star supporting cast, might well be expected to perform miracles, under Bradbury, with presumably the second team, it took a miracle to perform a miracle. Oppie opposed the development of the hydrogen bomb; Bethe, after the war, refused for some time to work on it. The principle of the fission bomb was “well understood” early on. Success was more an industrial than a scientific miracle, the gathering of scarce resources in a strained wartime economy to produce the fissile material for the bombs. But the true workings of the hydrogen bomb were involved and obscure. Initially, work was not in the most fruitful direction. How the hydrogen bomb was made was crucially dependent on how the Lab was reconstituted after its almost complete disbandment at the end of the war.

### Chronology

There were two distinct phases in the development of the hydrogen bomb: the classical Super from 1942 to 1950 and the new and successful hydrogen bomb from 1950 to 1952. Thereafter, the hydrogen bomb was refined and exploited, and today it is the mainstay of the U.S. nuclear arsenal.

The Super, proposed by Edward Teller and Emil Konopinski at the 1942 Berkeley conference, was sidelined during the war, but not abandoned. From 1946 to 1949 with the tolerance but not the official support of the national authorities, the work continued at a low level because of its scientific interest. But the first Soviet explosion of a fission weapon on August 29, 1949, changed the relaxed attitude of the Laboratory. On January 31, 1950, President Truman directed the AEC to continue work on a ther-

monuclear weapon. Then the only candidate was the Super. Increased activity but little progress resulted because the basic problems of the Super were just too daunting. However, the Laboratory added personnel and accelerated use of computers. Nuclear tests at the Pacific range, particularly the George shot of Operation Greenhouse in April 1951, explored some of the principles of thermonuclear reactions.

The second phase was initiated with the new concept of a hydrogen bomb discovered in late 1950. It was so obviously sure to work that the total resources of the Laboratory could be focused on it. The concept was brilliantly verified by the Mike shot in the Pacific on November 1, 1952, and the hydrogen bomb was born. During this entire period, very important improvements were made in the performance of fission weapons, significant since they were essential to the operation of the hydrogen bomb.

**Permanent Housing.** By 1948, the rebuilding of the Los Alamos Scientific Laboratory was substantial. Although far from reaching its peak wartime status, the Lab in personnel was well staffed, in equipment even robustly furnished. More important than the actual level of competence of the Lab was the feeling of permanence and the promise of future accomplishments. The Army was gone, an inheritance of wisdom and folly left behind, part of the physical and intellectual capital of the town and the Laboratory. The energetic, but sometimes arbitrary, administration of the military was superseded by the distant and paternalistic oversight of the AEC, responsive to the needs of the Laboratory. But for the present, the Lab was on a solid foundation

with an assured future. The town adjoining was also transformed into a permanent city, emphasizing the permanence of the Lab.

Beginning in July 1947, for the first time in Los Alamos, really permanent houses were built, initially in the western area. Those houses are still in use today, although some have been improved with extra rooms and upper levels: Small but comfortable homes, one story, 1000-square-foot well-designed floor plans, two or three bedrooms, one full bath with real bathtub, hardwood floors, beamed open wood ceiling and fireplace in the living room, full row of extra closets down the central hallway, natural gas cooking, heating, and hot water, one-car open carport with additional outdoor storage. Styling was definitely New Mexico, some fake adobe, single units or duplexes—all with the fresh-paint smell of newness.

Except for the upsloping hill to the far west impinging on the national forest, the western area had been meadow; so the land lacked the tall pines and small shrubs of the fringing woods. Small willows and olive trees were therefore planted. Today, some 50 years after, the area looks richly landscaped. The streets were set out on an interesting quasi-Cartesian grid with some cul-de-sacs for variety. With the open spaces included, each home site averaged half an acre, but there were no defined lot lines. In Los Alamos, as counterpoint to Robert Frost's memorable New England dictum, "null fences made good neighbors."

The physical isolation of The Hill and the fenced-in town site exaggerated further the feeling of isolation from one's family. Obviously, almost no grandparents and no sisters or cousins or aunts either, only coworkers—they were your family. But the comradeship of shared work and shared neighborhood formed bonds closer than kin. To the young scien-

tists awakening in the morning, feeling the crisp clean air, viewing with 100-mile visibility mountain vistas and endless skies, it was like nirvana. When they arrived at the Lab, to the working scientists, it was indeed a nirvana, but with boundless opportunities. And there was that great feeling of comfort and cooperation with friends at work.

**The First Soviet Nuclear Explosion.** Joe 1, the first Soviet nuclear shot detonated on August 29, 1949, was a historical marker for the scientists at Los Alamos. It confirmed their conviction that there was no secret of the atomic bomb—that nature's book was impartially open to all and the Soviet scientists could read it. Although it was no surprise, it brought a shock of realism to their work and changed leisurely investigations into matters of great urgency. Now, additional people joined the Laboratory staff.

**Accelerated Staffing.** Among the senior scientists, John Wheeler and George Gamov were newcomers. The old masters of the wartime effort—Bethe, Teller, and Fermi—took leave from their universities and came back, generously giving part time. These mature physicists brought with them a new contingent of their students. From Princeton, as the Matterhorn Project for civilian applications of controlled thermonuclear burn phased down, Wheeler brought Ken Ford and John Toll. Burt Freeman and Joe Devaney added to his group—four young bachelors injected into a community mainly of young marrieds with children. This group was soon engaged in calculating the radiation transport for some of the nuclear test shots using new methods devised by Wheeler. Hans Bethe sent his students George Bell, Walter Goad, Carl Walske, and Albert Petschek, and then came himself. These men joined Conrad Longmire in the neutronics

group, but they participated more widely in weapon design. Fermi reappeared with Dick Garwin—the equivalent of a whole laboratory capability in that couple, not a metaphor but a reality.

Guided by the old masters, these young men, along with the wartime holdovers, provided the muscle for the detailed calculations necessary for the design of improved fission weapons and, more important, a thermonuclear weapon. New computing machines were ready at their service. Actually, a new method of theoretical scientific work was in the making. No longer was progress made by advanced mathematical analysis, giving numerical results by slide-rule manipulation. Now, scientists programmed the computers, and instead of staying up all night baby-sitting experimental setups, they cradle-watched their computers at their allotted tasks.

The new method gave birth to a new breed. Two other newcomers typified them: Robert Thorne and Art Carson. These men would write their own codes, and nobody else knew precisely what was in them; nobody else could successfully run them. They were opaque to most, but like the mysterious prophesies of the Delphic oracles, the output stream of computer paper was believed to be utterances from the gods.

Finally, the Lab was up to strength to repeat for the hydrogen bomb the miracle that made the fission bomb during World War II.

## Nuclear Testing

Primary among the tools of the nuclear weapon trade is the nuclear test. First used at Trinity site, the nuclear test is a vast expansion over traditional scientific experimentation. These tests are expensive. For the Pacific tests, the military of Joint Task Force Eight deployed ten thousand

men and almost a thousand ships to Enewetak Atoll. The small cadre of scientists was almost lost in the human melee of army and construction personnel. Two entire Pacific atolls, Bikini and Enewetak, were commandeered, their population transplanted to other islands. A military tent-city was set up on Enewetak, the southern island of the atoll, which gave the atoll its name. A little to the north,

on Parry Island, the quarters and laboratories of the scientists were constructed, an invisible security curtain separating that island from the more populous island to the south. Before the war, the islands were covered with coconut palms in cultivated rows. During the fighting in the Pacific, the plantations were destroyed. Twisted military equipment was the new flora, decaying as mementos of the campaigns. Now once again, the palms and the coral rock of the island were to be sacrificed, this time to the aims of the test program. The Mike shot, at 10 megatons, for example, consumed an entire island, Elugelab.

The magnitude of the effort required for nuclear test programs in the Pacific put a somewhat unwise discipline on the Laboratory. The dates for an operation were set long in advance, and the Lab research and development had to be focused on preparing shots in time for the operation. This method precluded some avenues of research considered too long term; in other cases, it resulted in a too-hurried preparation for a test series. To remedy this failing, the AEC opened up the Nevada test site. Whereas in the Pacific the program was “get ready for a test,” eventually at Nevada the Lab’s watchword became “test when ready.”

Nuclear testing became a political, as well as a scientific, enterprise.



View of the Enyu Island in the Bikini Atoll.

International treaties regulated testing. International motivations resulted in test moratoriums, mutual or unilateral, and in the ending of such moratoriums. Nor was the number of tests or the nature of the tests free from political considerations. When the Soviet Union broke the moratorium in 1961, the United States responded by resuming its testing. At the cabinet meeting to decide on the test program, Harold Brown, the Defense Director for Research and Engineering, gave a detailed technical briefing on the scale of proposed tests. At the end of Brown’s talk, President Kennedy turned to his brother Bobby, the Attorney General. He asked how many shots the Russians were planning. When Bobby answered, the President, disregarding all Brown’s technical input, simply ordered that the United States should plan for the same number of shots in its test resumption series. The United States and the Russians have negotiated a comprehensive test ban treaty, which both countries now observe, but since the U.S. Senate so far has refused to ratify the treaty, it is not the law of the land. Since March 1992, however, the United States has not conducted any nuclear test.

#### Transport to the Pacific Atolls.

To the Los Alamos scientists, working at the Pacific nuclear test range was

a whole new cultural experience. It started at Hickam Field in Hawaii, the departure point for the military aircraft transport to the test area. Before takeoff, each civilian was given an equivalent military rank. There was company grade, corresponding to lieutenant and captain; field grade, corresponding to major and colonel; and general officer grade. Your quarters at

Hickam were based on your rank. Quarters were important because the schedule of takeoff of the Military Aircraft Transport Service planes was rarely adhered to. The procedure was to assemble all travelers at the site ready to board the plane whenever it became available—that way, no time would be lost in rounding up the passengers from the presumed pleasure spots of Oahu. This was very efficient for the airplane, but it often meant sitting around for many hours waiting for your aircraft. And it was your assigned aircraft. If on final checkout for takeoff, some slight problem was found, you were not given another plane; you just waited for yours to be fixed. Often, the delay was just a few hours; sometimes, it was days. Then the level of comfort of your quarters would be important. Carson Mark, although he was of general officer rank, stayed with his T-Division scientists in their field grade quarters.

The C-54, the military version of the commercial Douglas DC-4, took off. This was a cargo carrier, with passengers only as a courtesy. Along each side of the aircraft was a long bench of canvas supported by aluminum tubing. These were the passenger seats. The center of the fuselage was filled with bulk cargo strapped down to lugs in the floor. The aircraft was not pressurized, so the top altitude in flight was limited to about 8000 feet. No problem flying over the Pacific. Once

you passed the Hawaiian Islands, the elevation of the coral atolls of the Pacific was only a few feet above sea level. However, at those low altitudes, there was considerable weather that made for a bumpy ride. The C-54 is a rugged aircraft that could take more jolting and wind shear pummeling than the passengers' composure could accommodate.

Slow is this airplane, and large is the Ocean Pacific. Hours, slow hours in flight, fitful sleep sitting up, too hot, too cold; these aircraft have no sophisticated climate control. Bulky Mae West lifejackets on at all times bring comfort in being uncomfortable—one can last long if ditched into the warm Pacific. Then landfall, in the wide ocean that small oval lagoon, pearled with foam on the seaward rim, with deep blue water at its heart, so welcome, as our bird swoops in for a landing. Enewetak Island.

All is protocol as you leave the plane by grade when your name is called. First, the security check; then you are assigned sleeping quarters. An orderly, a lieutenant colonel, takes care of you, bringing your ration of duty-free spirits. Then, a quick wave-rocked ride in an LCM (short for landing craft man), and you arrive at Parry Island. Was this one of the boats used in the Enewetak campaign? Now you are on station.

**Life on the Atoll.** The atoll is the top of a gigantic sea mound peeking out above the ocean's waves. Long, long ago a hot spot in the earth's upper mantle forced a flume through the thin Pacific tectonic plate. Over millions of years, molten rock under great pressure pushed through the flume, spread out upon the ocean floor, and piled up to form an underwater volcanic mountain. Eventually, from the benthic depths 5 kilometers below the surface, the mountain grew to pierce the waves, and thence 2, 3, 4 kilometers into the sky. In one or

a series of cataclysmic explosions, the volcano blew its top, leaving a ringed caldera remnant. Slowly, moving only 100 kilometers in a million years, the tectonic ocean plate drifted northwesterly, leaving behind the volcano's source of fire. Now dying into new life phases, the rim of the caldera weathered down. Combined with the subsidence of the ocean floor depressed beneath its mass, the caldera disappeared below the surface of the ocean.

But then a billion billion little coral animalcules went to work. These small creatures can survive only in the narrow subsurface depth of about 20 meters. They built a coral crown atop the sinking volcanic rock. They could not invade the deep center of the caldera, which then formed the beautiful interior lagoon of the atoll. As the remnant mountain continued to subside, the lower parts of the coral reef died, crushing into limestone, while new live coral was added just below the surface. Wave action continually broke off small pieces of this coral and threw them up to form a ridge above the surface, the multi-islands of the atoll. The weathered coral formed the fine white sand of the islands and their beaches on the lagoon. These isles are evanescent. What the waves disgorge in great storms, they can devour again.

At high tide, the great ocean rollers come in and break at the atoll's outer rim. They spill over onto the tidal plain, perhaps 100 yards out. The water there is only 4 feet deep, more or less, depending on the tide. Rushing toward the shore, the breaker's water mass reforms into small waves, and they in turn make puny crests, which break and spill upon the island sand. As the tide recedes, it bares the black dead coral ridges of the plain. In thousands of little tidal pools, it traps the itinerant dogfish among the resident flora and the sea cucumbers that live in these puddles.

The dogfish have sought shelter from the deep sea beyond the atoll rim, where the large predators, sharks and barracuda, play. But many times, the pools become isolated, the sun warms the waters, the dogfish have no room to maneuver, and they become prey to the shore birds. Humans leave them alone; they are not good to eat.

When the tide is out, the coral rock of the tidal plain is exposed. It has sharp ridges; the general surface is slippery with sea slime. A cut from the coral quickly becomes infected.



**This chapel was built for the people involved in the tests on Parry Island (now known as Medren Island) in the Enewetak Atoll.**

Bacteria too enjoy the plenty of the tidal pools.

From the atoll rim, the water bottom, which is the edge of the sea mound, falls rapidly away to the sea floor, 5 kilometers below. The slope is about 20 degrees. As a result, the low-tide boundary is sharp, indented by old dead coral reefs at the surface, live coral heads a little way out below the surface. No one is crazy enough to try to swim off the atoll rim. The waves break sharp, the rocks are sharp, and the sharks patrol the border.

For the experimentalists, life on the atoll was a race against time, 12 hours a day, 7 days a week. Construction, installation of equipment, servicing,

checkout, calibration—all done for a microsecond of data, with no second chance. In contrast, the scientists from T Division were on hand as experts in the details of weapon performance or, in some cases, for additional information on the theory of the experiments as required. For them, the hours were sometimes heavy with boredom, further weighted by the monotonous weather—air temperature 80 degrees Fahrenheit, water temperature 80 degrees, humidity 80 percent, the trade winds blowing 22 knots, all unchanging day or night.

But sometimes there was, for these island-bound theorists, the excitement of the scientific and technological environment itself. You climbed the test towers 100, 150, 200 feet into the sky to check the layout of experimental equipment because specifications and blueprints did not always contain the full information needed for a successful test measurement.

Modifications to equipment were sometimes made on the spot under a theorist's instruction. You would lift your eyes from the apparatus, and there you would stand atop the tower, with the constant trades cooling the lingering sweat from your climb, and the Pacific stretching to a horizon lost in the faint sea haze, which to your view might not exist at all, except as another theory.

On the atoll, the sea is ever present. The sound of the waves is a steady background, pleasant when you wish to listen, ignorable when you are involved. No matter how busy, you would take some time off to swim in the warm waters. No shock as you plunged in, just the warm wetness refreshing. Most of the men would swim out to the coral heads in the lagoon, facemask on, snorkel set. There the sea creatures that belong to the atoll play. No prior experience could prepare the invading mesa dwellers for the richness, the variety, the colors, the beauty, and the strange-

ness of these underwater gardens of the sea.

Humidity was an enemy. In the damp and the heat, electrical equipment deteriorated, metals corroded, catalyzed by the tiny salt crystals always present in the air. Experimentalists were readjusting, repairing, replacing. The theorists' Marchant calculators were almost unusable after an overnight exposure to the damp. To prevent this deterioration, you kept them, when not in use, in a locker heated by a 100-watt bulb, lit all the time. The same lockers also protected your clothes and particularly leather shoes, all of which otherwise became moldy almost overnight.

Food was one of the every day recreations on the islands. Holmes and Narver, the support contractor providing the general services, knew how to keep their construction workers happy: Give them lots of overtime and lots of food. The budget for food alone, in the 1950 era, was \$5 per person per day. It was a bulk no-menu mess hall. You ate what was served. And what food! Choice meat sirloins or filets mignons were piled on great platters passed down the long tables, 20 workers sitting on benches on both sides filling up their plates. It was more than all you could eat. No worry about excess quantities prepared—the leftovers were the next day's stew, stew to shame French cuisine. On arrival, the new visitors watched with amazement as the old timers, competing with the construction contingent, heaped two or three steaks on their plates and sometimes asked for more. After a week, one joined the food orgy. But milk was that terrible tasting mixture made from powder, except when a freighter had just come in. Then there was fresh milk, the right stuff. Otherwise, no decent milk, but there was always plenty of ice cream. The burly construction workers were able to use up all that food, but some of the sedentary scientists stored the

excess on their middles.

**Transportation at the Atoll.** Were it not for the support services, the scientists could not get their jobs done. Since the shot island itself and other northern islands with experimental sites were about 50 kilometers away from Parry Island, communication and transportation links were needed among them. The joint task force military personnel supplied both.

Water taxis and LCMs were the standard interisland carriers. The speedy water taxis were a traveler's adventure, if he were up to it. In any but the calmest seas, these taxis set up a spray that, but for the canvas-top protection of the rear two-thirds of the boat, would soak any passenger. You had the choice to sit under the canopy and inhale the exhaust fumes or to sit up front and feel the salt spray dashing on your face. Most of the young men enjoyed the ride, but Roy Reider, the safety engineer, whose job required him to visit the experimental stations often, was seasick susceptible and hated the transit. The LCMs were noisy and slow, an experience in the practical results of shock waves as the flat forward ramp in its up position smashed into the choppy sea.

For rapid transit, there were the Army's L-5s and L-13s, the small light observation planes. They could make 100 knots, and Parry to Eleleron took only 20 minutes. Operating in the trade winds, these aircraft were remarkable. Landing speed was as low as 25 knots. In the 25-knot trade winds, they could land on a dime. Actually, a passenger could jump off a landing plane before touchdown with but the care needed to step off the end of a moving passenger walkway at a modern airport. Not much of a runway was needed for takeoff either. These planes could take only one or two passengers, so it took a high priority to get a ride. Overcoming these superficial hard-

ships was part of the culture of the test site.

**The Pogo Planning Staff.** The Pogo staff, the name borrowed from the famous comic strip, was the overall technical planning staff for the scientific operations. That group had done most of the stateside definition of the experiments. Now, they were on-site to help in the execution. Fred Reines, head of the Pogo staff, was on loan from T Division. Fred was a broad-range idea man; some of the experiments were his personal suggestions. He was dynamic—no boredom in his presence. Pogo staff meetings were almost continuous, monitoring every aspect of the tests. If there were no apparent problems, Reines might suggest something overlooked, or he might even improvise an additional small experiment. Fred always held an extra Pogo staff meeting at 8 p.m.—too early to go to sleep, so why not staff up a little more, nothing else to do on the islands. But there was something else in which Fred did not indulge: the 9 p.m. movie. Darol Froman, the associate director of the Lab and a former scientific test director himself, now out on the islands and a much-valued presence at the Pogo meetings, did like movies. So did Harris Mayer, theorist and very good personal friend of Fred's. Promptly at 8:55 p.m., Darol and Harris would stand up, deliberately disturbing the meeting, to show where the proper priorities were, and leave to go to the movies.

Fresh movies came in on the C-54 transports. When they were not available, old ones were reshowed. The theater was open air, with rows of hard benches set out before a big screen, a small bright cutout on the dark sea supporting the lighter sky. Darol and Harris sat down close together because they knew what was coming: not the suspense on the screen, but the weather. Predictable, as were the

trades, the rain would come in at 9:45 p.m. The two scientists were prepared; one poncho covered both of them. They sat together until the rain, as it sometimes would, came down in sheets so dense that the scattered light from the droplets overwhelmed the image on the screen. Except for these very dense showers, the rain was pleasant. It mattered little that you were wet on the outside of your clothing, when because of the humidity, moisture was always condensing on the inside.

Characteristically for those years, the tests were an all-male operation. Indeed, there were but few women scientists at the Lab—Jane Hall, Diz Graves, and Cerda Evans among them. It was understood that it was no denigration of their competence that they were excluded from the test operations. That's just the way things were. Times have changed, but then the camaraderie was a man thing, an enterprise of brothers.

But one got to know one's coworkers in a total living experience otherwise impossible, even in the close-knit Los Alamos town community. Overall, life on the islands was a comradeship in purpose, expressed in activity in test preparation, in sharing experimental results, in recreation, and in boredom.

### **The Path to the Hydrogen Bomb**

The fission bomb was a reality. Nature had indeed been generous in her choice of nuclear cross sections and the number of neutrons released per fission, so that the task was daunting but doable. But nature had also generously provided for another much greater and more pervasive energy source—the thermonuclear furnace of the stars. There, in contrast to fission, where a heavy nucleus is split to release its energy, four light hydrogen

nuclei are ultimately fused together to form a helium nucleus. In the process, the very high binding energy of helium is released. Scientists had been captivated by the fascinating complexity and excited by the potential of this thermonuclear fusion reaction even before the Los Alamos Laboratory was opened in April 1943. Here was the possibility to harness on earth this energy of the heavens. But instead of using hydrogen, the stellar reaction chain was to be short-circuited by starting with deuterium—deuterium nuclei in thermal agitation colliding with deuterium, a much more rapid process. Although nature had been generous with her margins in the physical parameters, as she had been in the case of the fission bomb, she was not nearly as transparent in revealing the proper physics to follow. In fact, the path to the hydrogen bomb was a tortuous one, with many interesting side branches to cause delays.

Clearly, the tremendous concentrated energy release in a fission device was the key to initiating the much greater energy release in a thermonuclear fusion assembly. It was time to start serious work on that possibility. Now, in a star the thermonuclear furnace is contained by the pressure generated by the gravitational effect of its huge mass. The reactions proceed slowly, majestically, on a grand scale, the overall cycle taking thousands of years. On earth, the problem would be the confinement of the exploding thermonuclear bomb. This necessarily precarious balance between explosive force and containment restraints depends upon the complex relations of the many physical processes involved.

The idea of a thermonuclear bomb powered by the deuterium-deuterium reaction was brought up in the pioneering Manhattan Project 1942 summer study at Berkeley by Edward Teller and Emil Konopinski. Oppenheimer, the leader of the study,



and Hans Bethe, one of the “luminaries” involved, enthusiastically adopted the idea and focused the best efforts of the study on it. Oppie called the bomb the Super. During the war, however, the Laboratory could not fully indulge this intense interest of the scientists. But after the war, there was time to explore the concept of the Super.

Although the Super was not on the main line of the Laboratory’s mission, the curiosity of the scientists could not be constrained. The everyday work of increasing the yield of fission bombs by small factors paled before the promise of the fusion bomb to increase yields by orders of magnitude. Understandably, volunteers were eager to work on it. In an informal arrangement characteristic of the organization of the Laboratory of that day, they started work on the Super—the measurement of the cross sections of thermonuclear fuel by Jim Tuck, the calculations of the equation of state and radiative transfer opacities by Harris Mayer, the development of computer codes by Marshall Rosenbluth, Art Carson, and Foster and Cerda Evans, while overall, Teller was dreaming, thinking, analyzing, inspiring the other scientists, and yes, educating, persuading, the administration of the Lab and the powers in Washington to advance the cause of the Super. On the action level, he drafted Enrico Fermi and Johnny von Neuman to engage in the concepts and the calculations, and he had Stan Ulam with his dedicated helper and tireless worker Cornelius Everett to carry them out. However, contrary to her transparency in the case of fission weapons, nature was more complex and subtle with the Super concept. She was not prepared to give up its secrets readily. Our understanding of the processes involved, our techniques and tools for diagnosing the device were not adequate for the job. Will it, won’t it, will it work? Our results were discouraging.

Progress by grand concepts and giant steps was stalled. It was time to make haste slowly. Consequently, it was decided to place a small mass of thermonuclear fuel adjacent to a very large yield fission bomb. Such a small mass would not give enough energy to ignite the Super, but it would demonstrate at least that some external thermonuclear burn could actually be accomplished. So was born the idea of the George shot. Tested at Enewetak in Operation Greenhouse in April 1951, George was a complete success. The yield was about 225 kilotons, only a small part of which was the precious thermonuclear yield.

But George was much more important in the process of its conception than it was in the success of its testing. Because of its influence, George deserved its ranking as the nonpareil shot in nuclear testing. Although the scientists were involved full time in the details of the test, in the shadowy recesses of their brains, waiting to be brought to full consciousness, were all the physics and components of the hydrogen bomb. The daunting difficulties of the Super concept could be avoided by a new approach. Not obviously suggested by the planning for the George test, that approach required a new synthesis of all its elements. Consciously, the scientists worked in the usual combination of inspired conception followed by critical analysis—sometimes with both going on almost simultaneously within one mind. When both aspects clicked together, then an idea was born. Consciously, the scientists did not proceed by considering a logical extension of the elements of George or the principles of its action. They thought they were considering the problem anew. But in their subconscious, all the elements were there for the viewing; just the new synthesis had to be made. How very clever of the human mind to uncover the obvious when it was not obvious!

This new idea transformed the concept of the Super into the beautifully workable hydrogen bomb. Remarkably complex, and devilishly interesting was this new concept—and capable of great flexibility in applications. It led to the felicitous design of a considerable variety of thermonuclear weapons. A transparently workable design with many important details was worked out by the early days of 1951, even before the firing of the George shot. The shot itself was then an irrelevancy.

There has been a continuing discussion among scientists, historians, and the curious public about the contributions of Teller and Ulam to the concept of the thermonuclear bomb. The overriding fact is that the bomb is an actuality. The Soviet Union and the United States made thousands of them. China has some in its nuclear arsenal. So has the United Kingdom. Nature had provided generous margins in the properties of radiation flow and nuclear reactions that made this complex concoction challenging and intriguing but, ultimately, not excessively difficult to master. Teller, Ulam, Sakarov, or some unknown researchers in other countries—that is no longer important. All these men were talented, creative, worthy of respect, even if afflicted by some modicum of “Fame . . . that last infirmity of noble mind.” At this late date, the distribution of fame or blame is a diversion from the stark reality. We now have the knowledge of the hydrogen bomb. Long ago, humanity took upon itself the knowledge of good and evil. What good and evil will we make of this?

### The George Shot

The nonpareil thermonuclear shot, wittingly and unwittingly testing many principles of thermonuclear weapon design for the first time, was

code-named George. It was one of four shots in the Greenhouse Operation of April 1951 on the small island Eleleron of Enewetak Atoll. George was a Teller initiative, with ideas and sweat contributed from all over the Laboratory. It was a test of the principle of a thermonuclear reaction, but it was not, by any means, designed as a complete thermonuclear bomb. A fission bomb provided the energy to start the burn reaction in a small mass of thermonuclear fuel. But to examine the reaction in detail, it was necessary to separate the fuel from the fission bomb. Therefore, the design tested had a separate implosion fission bomb with a hefty yield of about 225 kilotons. Energy from the bomb would ignite the fuel. The test device was placed atop a sturdy 200-foot-high tower. Many different instruments were arranged with a clear view of the external mass in order to diagnose its performance.

Two experimental stations were at the base of the tower. One, belonging to the University of California group under Herb York, was to measure the thermal x-rays from the hot fuel mass. The other station housed the electronics for the Naval Research Laboratory (NRL) experiments measuring the time dependence of the neutrons produced in the thermonuclear fuel. That group was headed by Ernst Krause. Their aims were the same, diagnosing the thermonuclear reaction, but the ethos of the two groups could not have been more different.

The NRL group under Ernie Krause was a well-practiced machine; the men had worked together for many years. Krause was careful, meticulous, well organized, a hard worker himself, and a hard driver of his men. Here was a team that knew by prior experience how to get the job done. On shot day, they were ready, their station buttoned up.

The University of California group under Herb York was a newly gathered

assortment of smart, eager young men in their first field experience, many later to become stars in their own right. Besides York, who became head of the Advanced Research Projects Office in the Department of Defense and later Chancellor of the University of California at San Diego, there was Harold Brown, a future Secretary of Defense; Mike May, future director of Lawrence Radiation Laboratory (known as Lawrence Livermore National Laboratory since 1979); Robert Jastrow, Chief Scientist at NASA, Goddard; Hugh Bradner, undersea photographer and underwater equipment designer par excellence; and Bill Grassberger, who had a long, fruitful career at Livermore as a radiative transport expert. Under York's leadership, the team members worked with the enthusiasm of youth and the luck of the blessed. On shot night, they had an unexplained slow leak in their vacuum pipe—no way to treat it. If it continued, the experiment was lost. In the dead of night, miraculously, the leak stopped. At shot final countdown, the experiment was a "Go!"

Herb York's group later became the nucleus of the new Lawrence Livermore Laboratory, the second weapons laboratory of the United States.

**The 14-MeV Neutrons.** The experiments to measure the total number of thermonuclear neutrons from the external mass of the Greenhouse George shot were straightforward in concept, massive in practice. Louis Rosen's especially designed detectors consisted of nuclear-emulsion plates mounted in a massive concrete collimator aimed at the fuel mass. Emitted 14-million-electron-volt (MeV) neutrons passing through the collimator would cause recoil protons upon elastic scattering from the hydrogen atoms in the emulsion of the plate. The ionized track of the protons would be revealed when

the plate was developed. The tracks could be measured and counted under a microscope. The collimator would not see the much more numerous fission neutrons from the multikiloton energy yield of the fission bomb itself. A strong 14-MeV neutron source placed on the tower at the fuel location was used for calibrating the entire setup before the shot.

In this experiment the really important problem was protecting the plate from the blast and shielding it from the gamma rays of the fallout. Heavy concrete shutters, released by explosive bolts fired synchronously with the detonation of the bomb, fell by gravity to tightly seal the detector. The concrete walls were sufficient shielding to attenuate the late-time gamma rays from fallout.

The George shot on Eleleron Isle, viewed from Parry Island base 15 miles away, was a terrifying sight. The fireball flashed, the cloud formed and rose, the characteristic mushroom shape developed—and rose and rose. One's head tilted up and up to follow it, the angle of view increasing. The radial expansion of the cloud and the distortion of perspective made it appear that the top of the cloud was marching with increasing velocity toward us at Parry Island, menacing the puny viewers. Shivers of latent feral fear crept along their spines. The great yield of the fission bomb had surely been achieved, but what of the thermonuclear fuel?

A quick water-taxi ride to the shot island. The recovery crew disembarked on a devastated wasteland: just coral sand left, the shot tower gone, and a crater in the coral rock left instead. The recovery crew had to wait for clearance from the radiation-safety monitors before going in to recover the precious plates. They had made a quick aerial survey of the radiation levels and plotted a reasonable approach path. The levels were variable, hot spots here and there,

most not really dangerous but not trivial. There was caution in choosing a path among the hot spots, where dose rates reached 100 rad per hour. The crew was in no danger of getting a lethal dose; 500 rad was the so-called LD<sub>50</sub>, meaning a 50 percent chance of dying as a result of exposure. The real fear was that one would accumulate the maximum permissible dose of 3 rad for the whole operation. That meant a quick ticket home.

To keep within safe limits, one needed a good entry and return path and a quick recovery at the detectors. When Louis Rosen and his recovery crew got to the massive concrete neutron cameras, they saw that the covering, protective layer of sand had been blown away by the blast. The great hunks of concrete had been tilted—of course, alignment now was irrelevant, but were the camera films all right? The opening of the concrete block cameras to extract the film was not going as it had been in the dry runs. Louis was working hard, but he was unhurried even in that radioactive environment. The plates came out intact.

With the precious nuclear-emulsion plates safely stowed in shielded recovery packets, the group took a quick trip by water taxi back to base camp on Parry Island. Strangely, the time on the return trip passed much more quickly than on the approach. Then, Louis Rosen went into the darkroom to develop the plates. After a wait that seemed longer than it really was, he came out and placed a nuclear-emulsion plate under his microscope. The T-Division contingent that had gathered to hear the results was waiting anxiously for his reaction. If the burn was a success, they thought it would be obvious, and Louis would say so in a minute or so. But Louis, face expressionless, said nothing at all. Not to disturb him, they started up a game of poker, Carson Mark, Frank

Hoyt, Marshall Rosenbluth, and Harris Mayer, with Rod Spence and George Cowan from the radiochemistry group and Bill Ogle, deputy scientific test director. As the minutes passed with no signal from Louis, they thought in despair that he was anxiously seeking for a few, even one, true neutron track. The game got wilder, deuces and treys wild, then eights and jacks added—no one of the group knew the odds of the permissible fantastic combinations—and



**Louis Rosen**

do five aces top a royal flush anyway? Yes. An hour and a quarter later, Rosen got up from his microscope, stretched, and said simply, “We got them.” The careful scientist and the unforgivable miscreant had been counting and measuring proton recoil tracks from the 14-MeV neutrons, hundreds of them, for the whole time. He had folded in the proper calibration, done a statistical analysis, and gathered the data in good shape. Yes, that small external mass was one hot thermonuclear source. No question, thermonuclear fusion was incontrovertibly demonstrated.

### **Mike—The New Hydrogen Bomb**

The scientists returned from the

Pacific proving grounds bearing rich treasures of experimental results. The Booster device, also tested in the Greenhouse Operation, worked well. The ignition of the small fuel mass in the George shot demonstrated a thermonuclear burn and supplemented the understanding achieved in the Booster. However, the success of George gave only vague moral support to the new concept of the thermonuclear weapon. The immediate importance of George was the proven performance of the sophisticated diagnostic tools used. Those instruments would be available for the diagnoses of future, much more complex, weapon designs.

The sun-tanned crew from the Pacific test range returned to a Laboratory that was about to be dramatically changed by the new thermonuclear weapon concept. Los Alamos was now committed to developing a concrete realization of the concept and testing it at the Pacific proving grounds.

Consider this recipe for disaster facing the Laboratory and discern how such a startling accomplishment could have come from it: A new technical idea proposed in March 1951 to be tested in only 19 months; the two leaders of the new program, Edward Teller and Marshall Hollaway, who cannot get along with each other; a Laboratory director, Norris Bradbury, who cannot make them cooperate and regretfully but decisively chooses Hollaway as project leader, causing Teller to resign; engineers under Hollaway who need to freeze the design in order to meet the test date; theorists, no longer the wartime stars, trying first to understand the applicable principles of the design to find the appropriate one and unable to come up with a final design. What organization could complete the task on time? An organization conceived in the value of each individual scientist, banded together as coworkers, friends,

neighbors, in an enterprise of national significance, with a tradition and a will to succeed—that was the Los Alamos Scientific Laboratory of the 1950s.

Out on the Pacific test range on the island of Elugelab at the north end of Enewetak Atoll, the Mike device was set up. No tower for this 60-ton monster—it stood upright on the coral sand. A bewildering array of experimental equipment was placed on Elugelab and other islands up to 25 kilometers away to diagnose the shot. Ernie Krause's crew, veterans of the Greenhouse George shot, were there but with much more elaborate experimental equipment this time.

After the preparations on the islands had been completed and the northern end of the atoll evacuated, the center of activity shifted to the command ship, the *Estes*, stationed 50 kilometers south of the detonation point, presumably a safe distance. From the ship, the firing signal was sent by microwave transmission to Mike, the deserted monument to destruction.

At 7:00 a.m. on November 1, 1952, Mike exploded with a 10-megaton yield. The island of Elugelab vanished. A crater 1 kilometer in radius and 50 meters deep had been blown in the coral rock; a fraction of the missing material had been vaporized or shattered into dust, which was carried up in the mushroom cloud to pierce the tropopause, high though it is in the tropics. At about 15 kilometers altitude, the cloud entering the stable stratosphere was forced to spread out, forming a flat pancake, instead of the rounded top of the mushrooms from the smaller explosions of fission bombs. The stratospheric circulation would carry the radioactive cloud around the entire Northern Hemisphere. Once again, as it had done at Trinity with the fission bomb, the Lab succeeded spectacularly with the hydrogen bomb. The Mike shot of Operation Ivy went off only one year



**The Mike device is pictured just before it demonstrates the new thermonuclear concept.**

and seven months after the George explosion.

#### **Reaction to the Hydrogen Bomb.**

Why was there interest among the military and the Los Alamos scientists in a weapon that would produce a yield of 1, 10, or even 100 megatons? The Trinity bomb was 20 kilotons in yield; it obliterated a city and terminated a war. Immediately after the war, the Air Force generals—familiar with bombing missions of hundreds of airplanes, each carrying about 10 tons of high-explosive bombs, a total of only a few kilotons—regarded the fission bomb as the proper successor to the mass bomber raid. They had no experience with larger yields, and in that mindset, they had no requirements for them. Furthermore, because of secrecy, they had no knowledge base to understand that much higher yields were possible even with the infant fission weapon technology.

During the 1946–1950 period, the military were not involved in the workings of the Laboratory; they placed no requirements on the bomb yield because they had no vision of its utility. They wanted as many bombs as needed to load their aircraft. Their

interest was in packaging and saving scarce fissile material. When approached with questions about bomb yields, they said 20 kilotons is fine, no need for more. But once the military were given the confidence that megaton yields would be available easily, they quickly changed their views about requirements. Now it was “no yield too high.”

Why did Los Alamos leap so quickly into the development of the new hydrogen bomb? First, there was the scientific knowledge of its inevitability. The concept in Oppenheimer's word was “sweet.” That concept indeed was “Nature's sweet and cunning hand laid on” (Shakespeare, *Twelfth Night*). But Oppie, in his years of government committee work since Trinity, had learned the wisdom of the laconic, short, and sweet. “Sweet” was his appropriate appraisal. He knew this idea for a thermonuclear weapon would work. No longer was this concept of the bomb a distant hope like the classical Super.

This was pregnant reality. For better or for worse, we had to fully understand this unborn thing. Surely too, the Soviets would know about it, and we had to understand what it would mean to them. The wisdom of the wide deployment as a weapons system could not be determined without an understanding of the characteristics of the hydrogen bomb. It had the potential of increasing the yield of nuclear weapons a hundredfold. At that time, the Lab was spending much effort in improving the fission weapons by what was considered great steps: up to a factor of 2. The near perspective was shadowed over by the great looming shape of the future.

When only the classical Super was the thermonuclear weapon candidate, Oppenheimer and Bethe particularly, but most of the other scientists as well, had an easy decision. There was in their view an abhorrence of the

weapon they had created. Strongly, they held the position that the Super was an immoral weapon that should not be developed. Since the probability that the Super could be made was low because of the great technological problems, they could reinforce the moral position by the pragmatic one. Because the scarce commodity was scientific talent, why waste that on the Super, which had only a slight chance of success and was unnecessary and immoral to boot?

But the new hydrogen bomb changed the conditions of the argument. Now, Oppie and Bethe and Rabi, with the scientific knowledge that they had, knew very well that nature would readily serve up the hydrogen bomb, leaving it to humankind to feast upon and digest. Now, morality must face up to reality, not to a remote possibility, as was the Super.

**Mike's Role in Nuclear Weapons Development.** Mike was designed with all the features of the new concept for the hydrogen bomb. As such, it was an integral demonstration of a practical weapon, even though it was much too heavy to be carried by aircraft. However, it was fielded as an experiment liberally instrumented to test specific features of the design and to learn about the behavior of the weapon if the yield was not as planned.

Fortunately, since the behavior was actually as planned, those measurements confirmed the methods used in several detailed features of the design, validating the methods for future applications in variations of the specific Mike configuration. Therefore, sufficient information was collected to allow the modifications required to reduce mass and even size so that a weaponized version could be made with confidence. The test also verified the procedure used in the design—understanding of the physics, idealization of the physics into models for the computer, and further idealizing the

actual configuration to one that could be calculated with new computer codes on the supercomputers then available at the Laboratory.

Mike was designed as a test of feasibility, not as a fieldable nuclear weapon. But its success, a 10-megaton yield in line with projections, clearly meant that the thermonuclear weapon would take its place in the U.S. nuclear arsenal. By the time of the next test series—Castle, in 1954—several versions were available and were tested. Smaller, lighter, cheaper, more readily maintained in the field, new weapons appeared in the stockpile. However, the strategic requirements began to change as intercontinental missiles came into the force, complementing the great bombers. Since multimegaton yields were no longer optimum, the national laboratories developed submegaton weapons in a surprising variety of designs. The heritage of Mike had to be greatly modified. As a result, a healthy competition between Los Alamos and Livermore arose, both laboratories contributing to the new stockpile.

## Afterword

An unusual circumstance, the wartime years. Los Alamos Laboratory was born for a single mission: Make the fission weapon and make it in time to defeat Nazi Germany. The legal framework, a contract with the University of California to operate the Laboratory, was merely a convenience. The University was a pass-through for funds, nothing more. In fact, the Army through General Groves and his appointed deputies ran the Laboratory in every detail. But there was wisdom beyond intent in this arrangement. An entirely new entity was created, a national laboratory, that in the ensuing years developed into an essential component of the body politic and the national economy. The legal conven-

ience now supported a powerful reality.

We, the staff of the Los Alamos National Laboratory, essentially work for the U.S. government, but we are not part of it. Because of our unique legal structure, we can help our government objectively, we can provide our results to industry impartially, and we relate to academia as partners philosophically and intellectually. The years have proved the worth of such an institution, more efficient than a government arsenal, more creative for the commonweal than profit-oriented industry, and more focused on national needs than academic institutions. That being the present, what of the future?

In the next quarter century as never before, the riches of our planet may be spread out before us. We may waste them, or we may use them wisely. Perhaps, unforeseen events may foreclose our choices. Nevertheless, at Los Alamos National Laboratory, we have the opportunity to use science and technology for the benefit of humankind. In that use, we may hope that wisdom will prevail in our society. ■

*For further information, contact Harris Mayer (505) 667-2604.*

*The Laboratory Today*

# Nuclear Stewardship in the 21st Century

As we close our celebration of the Laboratory's 60th anniversary, the times are changing. A new global environment challenges our presumptions regarding nuclear deterrence, nonproliferation, and national security. For the last 10 years, this Laboratory has been committed to developing the scientific base and technical tools to certify the U.S. nuclear stockpile without nuclear testing. The articles in this section document our significant progress toward that goal. Whether the route we are following will give the nation the deterrent it needs and enhance the nuclear nonproliferation regime is an issue under debate. We hope that the efforts of scientists and engineers at Los Alamos will help the nation to a prudent conclusion on this issue.



# Science-Based Stockpile Stewardship

## *An overview*

*Raymond J. Juzaitis*

Los Alamos was founded as the world's first nuclear weapons laboratory. Brilliant scientists from different nations, all committed to defending freedom, dedicated their time and offered their best understanding of physics, chemistry, engineering, and materials science to design and manufacture the first nuclear bombs. They had no previous experience, only the minutest amounts of the nuclear material for most of the project, and at the end, only material for one test. The scientists' only option was to exploit the full power of the scientific method, whereby concepts are challenged and the iterative cycle of theory, experiment, evaluation, and innovation leads to confidence in prediction.

When the Laboratory opened, the basic concepts for a gun-assembled weapon and an implosion weapon had already been formulated at the 1942 University of California, Berkeley, summer study, but the detailed physics necessary to assemble several critical masses fast enough to produce a successful nuclear explosion had to be acquired and demonstrated. The fundamental properties of the neutron chain reaction—the number of neutrons released per fission, the energy spectrum of fission neutrons, and the cross sections for neutron-induced fission, neutron capture, and neutron scattering—were measured at a feverish pace. Basic material and chemical properties of uranium-235 and plutonium-239 were determined. Diagnostics, such as flash radiography, were devel-

oped to measure the progress of an explosively driven implosion, and numerical methods were developed to calculate the implosion. Analytical methods, combined with judicious approximations, were used to estimate the amount of nuclear material that would be needed and to predict the efficiency of the nuclear explosion. As the experimental numbers became available, they were used to determine the parameters in the theoretical models. The resulting predictions for the explosive power released could be trusted within some margin of error.

The yield predicted for the Trinity test at Alamogordo, New Mexico, was the equivalent of 5 to 13 kilotons of the explosive TNT. The measured yield was even higher—17 kilotons. Considering that the designers were treading on unexplored terrain, these results were an awesome testament to the power of scientific prediction.

Sixty years later, our core mission bears some remarkable similarities to the mission of the early days. Today, the vast nuclear weapons complex of the Cold War, built after Trinity, has been reduced in size. Los Alamos and our sister laboratories, Sandia and Lawrence Livermore National Laboratories, are now responsible for all the science, much of the engineering, and a significant portion of the manufacturing needed to maintain the enduring stockpile. As a steward of the weapons in our stockpile, Los Alamos was challenged by the president to cer-

tify their safety, security, and performance—and to do so in the absence of yield-producing nuclear tests. The nuclear weapons program at Los Alamos relies on the scientific method to acquire the needed knowledge and to formulate predictions based on that knowledge. Stewardship means that we must predict performance as weapons age, identify the parts that need refurbishment, certify performance when weapons contain parts that are made from new materials and that have been manufactured by new techniques, and prepare for possible redesign of present systems to meet the changing needs of an increasingly complex world. However, we must accomplish these tasks through predictive capabilities, without resorting to actual nuclear weapons testing. This approach has never been attempted in the history of engineered devices. Achieving and demonstrating the required level of predictability demand at least as much (if not more) ingenuity and skill today and in the future as they did 60 years ago.

Los Alamos has been a science laboratory throughout its history. It has built and maintained the nuclear deterrent through its broad investment in science and technology and in the talented people who continued to create ideas that change the world. This overview and the articles that follow it show how our continuing investment in frontier science, first-rate scientists, and the rigor of the scientific method are producing sustainable nuclear stewardship in the twenty-first century.

### **Development of the Enduring Stockpile**

Only by reviewing the methodology used to create the existing nuclear weapons stockpile, can we convey the scientific challenges of modern stewardship. During the Cold War, changing military requirements drove the design of new weapons systems. Increasingly,

lighter, smaller, more accurate, and specialized warheads were required to maintain deterrence against the growing sophistication and hardness of the threat. These new weapons were designed to perform reliably during much more rigorous and demanding operating conditions, referred to as stockpile-to-target sequences, and ultimately deliver on target the certified yields, known as military characteristics. Later, requirements for increased safety and security led to the development of insensitive high explosives, fire-resistant weapons components, and other surety features. Weapons were manufactured in large quantities to counter the Soviet buildup. However, for logistic and maintenance simplicity, as well as to ensure a credible deterrence posture, the military required many identical copies of a few, well-honed, and fully characterized designs. That is, all these designs had their pedigrees in nuclear tests and in nonnuclear integral tests (weapons tests in which surrogates replaced the fissile materials). These tests improved our basic understanding of weapons physics and permitted us to develop an expanding body of empirical experience. This experience provided us with a means to improve and fine-tune weapons performance.

At the same time, weapons designers developed a series of computer codes, now designated as “legacy codes,” for weapons design. These were design aids to refine the qualitative understanding of the physical processes involved. Although not capable of directly predicting the results of nuclear tests to the accuracy required for the military, the codes were calibrated empirically to fit test results. Hence, the codes were a valuable, very sophisticated interpolation, and even extrapolation, device for designs in the neighborhood of those tested. The adjustments to the codes made directly from test experience gave designers a “feel” for how their incomplete simula-

tion tools related to materials behavior under the physical conditions achievable only in a nuclear test. The expert judgment gained from full-scale tests remained a key component in the designers’ craft during the Cold War era.

### **Stockpile Maintenance without Nuclear Testing**

Los Alamos designers were very successful at meeting the safety, performance, and reliability criteria of the military: They designed five of the seven weapons systems currently in the enduring stockpile. However, the focus of their activity changed abruptly toward the end of the Cold War. First, the nuclear weapons stockpiles that had accumulated in both our country and the Soviet Union far exceeded the size necessary to maintain stability. Building down the stockpile became more important than building it up. Second, there was a growing national commitment to global nonproliferation goals and to preventing terrorists from acquiring nuclear materials and weapons. Finally, by fiat, the United States and other declared nuclear states announced a moratorium on underground nuclear tests. Our last nuclear test occurred in 1992, just after the end of the Cold War.

In 1992, our nation adopted testing constraints laid down by the Comprehensive Test Ban Treaty. That is, we agreed not to perform a weapons test involving an uncontrolled nuclear chain reaction. The complete ban on nuclear tests, at “zero yield,” was seen by some policymakers as a mechanism to slow the proliferation of nuclear weapons to nonnuclear states. Without the option to test, it was argued, treaty members would be denied the key means of assessing and demonstrating nuclear capability.

For the weapons designers at Los Alamos and Lawrence Livermore, the two weapons design laboratories, the



change to a nontesting environment was intellectually as “seismic” as the nuclear tests had been in actual fact. Testing had been the ultimate guarantor of reliability and performance. Testing was the key means not only for certifying new systems and developing expert judgment but also for verifying the continued safety, security, and performance reliability of the weapons systems for which the designers were still responsible. What could possibly replace the sensation of having the ground heave underfoot after an actual nuclear test?

The answer to that question, arrived at jointly by the Department of Energy and the design laboratories, was a formal program in science-based stockpile stewardship. The idea was to build a strong base of scientific understanding, combine that base with our historical test experience, and from that combination, develop the tools to predict the performance of stockpile weapons without resorting to new nuclear tests. From small-scale physics experiments combined with theoretical analysis, scientists would develop a deeper understanding of detonations, hydrodynamic behavior, and materials behavior and hence be able to develop more-accurate weapons physics models. The new models would be incorporated into a new generation of simulation codes developed under the Advanced Simulation and Computing (ASCI). New facilities would be built to do more accurate nonnuclear integral tests of whole weapons systems. The integral tests would provide a method to validate the computer simulations of the early stages of weapons performance. Archival data from past nuclear tests would be used to validate the codes during later stages of weapons performance. Finally, through vastly expanded computers for carrying out more realistic simulations of weapons performance in three dimensions, weapons scientists

would be able to predict performance of the stockpile weapons with acceptable levels of confidence, maintaining the stockpile with no additional tests.

The need for scientific prediction, handicapped not by a lack of nuclear material but by the injunction against nuclear testing, has required a major cultural change for the weapons program. However, as new simulation capability has come online, as new theories and models have been developed and incorporated into the weapons design codes, and as new experimental tools confirm our predictions, optimism has grown among designers that science-based stockpile stewardship could be sustained for “near”-stockpile configurations.

## Successes of Stewardship

**Enhancing Predictive Capability.** A brief sampling of successes over the last decade illustrates the new understanding and scientific tools that are leading to enhanced predictive capability. Many of these successes are discussed in the articles included in this section on nuclear stewardship. Most remarkable are the increases in simulation capability achieved through ASCI. Both the level of detail in the simulations and the speed and size of the computing platforms have increased by many orders of magnitude. A major milestone for the ASCI multiphysics codes was the first end-to-end three-dimensional simulation of a nuclear weapon explosion—from high-explosive detonation to nuclear yield. This capability provides a strong foundation on which to build predictive simulation.

Although such calculations take several months even on the new machines, they were simply unimaginable just a decade ago. One of the challenges now is to achieve the shorter turnaround times needed for code validation and production use.

To be predictive, our simulations

must incorporate theories and models derived from and validated through a strong experimental science program. Our experimental program covers all the scientific areas related to weapons performance. It also spans the range from small-scale basic physics experiments to so-called integral experiments, which test the behavior of a whole weapon system just short of a nuclear test. In our gas-gun experiments, for example, we shoot a projectile at a small flat plate of plutonium to measure the material ejected from the surface. Those experiments provide basic physics information on dynamic response to shocks. On the other hand, in an integral experiment, we might replace plutonium with a surrogate, say, a heavy metal, in a geometry that closely represents that in a weapon system. Integral experiments known as subcriticals are conducted underground at the Nevada Test Site. In these experiments, high explosives drive the implosion of an assembly in a weaponlike geometry using amounts of plutonium that do not give nuclear yield. Thus, the simple experiments build the basic physics knowledge that is incorporated into the simulation codes, and the integral tests help us validate the predictions of systems performance.

The iterative process of experiment, theory, and simulation has already yielded significant improvements in some of our physics models, including a model for the propagation of detonation waves around corners and the development of more accurate equations of state for plutonium. The materials models have a direct impact on certification. Our new ability to accurately model the detonation of insensitive high explosives in complex geometries has helped us address a major stockpile issue. That new model has also helped us analyze accident scenarios and support the authorization basis at the Pantex manufacturing facility. The work on the equation of state of plutonium is contributing to

the certification of the newly manufactured pit for the W88 warhead. The pit will be certified through a large number of subcritical tests in which the weapon assembly contains a partial plutonium pit.

Simulation tools are also being developed to model manufacturing processes such as plutonium casting and to model materials behavior under weapons conditions. These computer simulation tools allow exploring a whole range of these processes for a fraction of the time and expense involved with real materials and equipment.

Another major success is the development of DARHT, the world-class dual-axis x-ray machine for obtaining high-quality, high-resolution images of hydrotests, which are nonnuclear integral tests of hydrodynamic implosion. Experiments at the DARHT facility are being used to address system performance and to validate weapon system codes. Very recently, radiography of a hydrotest at DARHT enabled us to resolve a major uncertainty in the calculation of implosion and thereby address an important stockpile certification issue.

The invention and application of proton radiography, a powerful new imaging capability, is one example of the enormous creativity of our scientific staff. This new technique is now being implemented at the rate of about 40 experiments per year at a proton “microscope” system installed at the Los Alamos Neutron Science Center (LANSCE). Short proton pulses passing through an electromagnetic lens system produce rapid multiple-time images of dynamic events with a resolution that can be as good as 15 micrometers. The movielike sequences lend insight into basic material behavior under extreme pressures and speeds and under dynamic conditions that would otherwise be difficult to access diagnostically. Protons have the advantage of discriminating among materials of different atomic

numbers, thus enabling the capability to “identify” materials in mixed conditions. (X-rays are not sensitive to atomic number.)

The two intense neutron sources at LANSCE also continue to yield important new nuclear data for weapons design and new characterization of plutonium and other weapons materials. Recent measurements at the Weapons Neutron Research facility at LANSCE, combined with theory, led to a major (30 percent) change in the cross section for the important (n,2n) reaction, in which the isotope plutonium-239 becomes plutonium-238. As a result, relative changes in plutonium isotope abundances became a reliable metric for determining the fission yield of plutonium in past nuclear tests. That development, in turn, resulted in an important reanalysis of the nuclear tests that underpin certification of the current stockpile. At LANSCE’s Lujan Center, inelastic neutron-scattering measurements have produced the first-ever determination of the phonon density of states of plutonium, an important component of our understanding of the equation of state of plutonium. Also at the Lujan Center, a major new detector system will enable us to measure the nuclear properties of very small radioactive samples, some weighing as little as one milligram. That capability will allow us to reanalyze radiochemical information from past underground nuclear tests with confidence that the physical processes determined from the data are correct and predictive.

**Stockpile Maintenance, Manufacturing, and Manufacturability.** Science-based stockpile stewardship involves more than developing the tools to predict performance. As a steward of the stockpile, Los Alamos is also responsible for maintaining the existing stockpile through a program of surveillance and response—taking weapons out of the stockpile, examining them, and solving any observed problems. One

type of response is the life extension program. In the next decade, this program will call for replacements or modifications of specific components in the stockpile, and thus it presents major engineering and resource challenges.

Los Alamos has also taken on some production responsibilities as facilities were shut down across the national weapons complex. Our most visible new task is to manufacture the plutonium pit, the heart of the weapon primary, but we are also responsible for manufacturing detonators, neutron generators, beryllium components, and other parts.

In pit manufacture, we have had to recreate the entire technology of the Colorado Rocky Flats Plant in a changed environment, where many materials and processes used at Rocky Flats are neither available nor permitted. Developing and qualifying the new processes and certifying the performance of the product without full-scale testing have been the first big test of the stewardship regime. We have changed not only our technology but also our traditional ways of doing business. Fortunately, our dedicated staff at the plutonium facility responded with their full measure of skill and intensity. By the start of the calendar year, they had produced a number of system qualification test pits and just recently delivered a completely weapons-qualified (“certifiable”) pit—a major achievement. In a parallel effort, our program leaders have initiated the development of sophisticated process monitoring and control procedures that guarantee quality during the manufacturing process. This investment in yet another aspect of predictive capability should enable us to sustain the pit manufacturing capability in the present environment of changing requirements and small throughput. Both the life extension program and our different production tasks clearly call for a science-based methodology to establish priorities and quantify our

level of confidence in the new or changed components. Responding to an aging component with a plan to replace all identical components in the stockpile and thus “rejuvenate” the stockpile may be a very expensive decision. Without careful assessment of performance versus impact, one can make poor decisions. As described in the next section, we are currently developing a quantitative framework for guiding such decisions and building confidence in stewardship.

### **A Certification Methodology**

Each year, the director of the Laboratory must assess the weapons in the stockpile for safety, performance, and reliability. This assessment must consider whether military characteristics and requirements can be met without a return to nuclear testing. In the current stewardship regime, the key question we face in the annual certification is, “What is the relationship between key weapon-performance metrics and the design margins of the system?” Furthermore, how far can we stray from the ideal design environment (materials, age, and tolerances) before a weapon will fail to meet its military requirements? And how can we quantify our confidence? That is, how much do we trust our predictions?

These are tough questions that have never before been addressed or quantified. Consequently, the policy community has challenged us to provide a rigorous scientific approach to reach closure on scientific issues and to quantify the level of confidence with which we certify the stockpile. In response, both Los Alamos and Lawrence Livermore have developed a certification methodology that revolves around quantifying margins and uncertainties for the various stages of weapons performance. By judging our progress on the problem of decreasing the uncertainties, we

have the means to rank scientific and technical investment. For example, we will be able to decide whether a particular process must be modeled at the molecular or macroscopic level to reduce uncertainty or whether some modest parametric representation would be adequate—all based on assessing the impact of the uncertainty on our confidence in performance. We know that complete predictive capability of weapons performance is not possible, but we will be able to estimate our degree of confidence and specify the requirements for increasing that confidence based on quantitative performance-related measures.

This new methodology has an important corollary. It can help translate the unwritten lore of our best designers into solid guideposts for the emerging generation of new designers. Our best designers, like innovators from every field, did not always write everything down, nor was there ever a prescribed method to document the detailed interplay between simulation and testing. The experienced designers had learned how to compensate for less-than-predictive models by adjusting empirical parameters to ensure enough “predictive ability” in yield and diagnostic measurements and to anticipate the “next” underground test. Now, a new methodology focusing on margins and uncertainties allows for more explicit representation and quantification of essential design decisions and judgment.

Most important in the long term is that certification without testing be sustainable. Sustainable means not only that we continue to increase our science and engineering understanding of the weapon system but that we use that knowledge to make cost-effective decisions about the scope of weapons refurbishment and to better address the issues observed in the aging stockpile.

### **The Current Global**

## **Environment**

Today, the international and national security environments have changed radically and have, to some extent, become entwined. Nations that were once our formidable and determined nuclear enemies have now become our real or emergent allies. Although the Cold War, a struggle that seemed destined to permanence, has ended, the threats to world peace remain real, and arguably, the instability around the globe is greater. Among the new and emergent allies, there is a new determination to stop the growth of this incipient instability—one brought to us by the harbingers of terror.

Against such a backdrop, our nation has been reevaluating its nuclear posture. Of course, nuclear capability remains the ultimate deterrent, but ever more voices raise questions about the nature and effectiveness of that deterrent. Here, effectiveness is not discussed in destructive terms, but it refers to maintaining real deterrence against radically different enemies and targets. It may be argued, and it would certainly be ironic, that the existence of nuclear weapons with lower levels of collateral damage and therefore increased “usability” may be the greatest deterrent and thereby the greatest force against their own actual use. The aim would still be to never have to use the weapons.

## **Policy Changes**

In early 2002, the Bush administration issued the findings from a Nuclear Posture Review that placed nuclear weapons in a new and different context. In the past, we described deterrence in terms of an offensive triad composed of intercontinental ballistic missiles, submarine-launched ballistic missiles, and strategic bombers, each carrying nuclear warheads capable of delivering kilotons, if not megatons, of

explosive power. Having evaluated the changed environment in both threat and technology, the Nuclear Posture Review offers a new triad, in which the three nuclear offensive capabilities above appear on one leg of a triangle, joined and complemented by strategic nonnuclear weapons. This change recognizes that precision delivery systems with conventional warheads, such as those exercised during the Gulf War and, more recently, in Afghanistan and Iraq, can now operationally achieve some of the strategic objectives that only nuclear weapons could have achieved in the past.

The first choice is always to avoid direct use of nuclear weapons and to use them only as a deterrent. However, in the event they were required because the destructive effect needed is achievable only through nuclear processes, our nation would not want them to have unacceptable collateral effects. For example, it would be less “effective” to threaten to use a nuclear weapon to destroy chemical and biological agents in a deeply buried and hardened arsenal if the explosion would produce widespread nuclear contamination. Consequently, there may be fewer nuclear weapons in the new triad, but they will probably have to be more robust and address new strategic problems.

The review also introduces a vital, new component to the new triad, namely, responsive infrastructure. In a world where technology is changing quickly, where emerging threats are difficult to identify in advance, the review challenges the science and technology community to develop flexible and adaptive capabilities. What does that mean for the nuclear weapons community?

### Advanced Concepts

In the past, we were asked to build thousands of identical warheads to be

placed in ballistic missiles, each directed toward specified targets. Today, the technical and policy communities are increasingly seeing a need for new kinds of devices. Depending on how the threat evolves, we may be tasked to build relatively small numbers of weapons of very special and limited capability. If so tasked, we may extrapolate some of those weapons designs perhaps from the designs in the existing stockpile. Those would be moderately easy to certify without testing. A great number of possible “new” weapons might be based on design concepts and weapons systems that were tested in Nevada before 1992 but never implemented in the stockpile. Depending on the testing pedigree, these may or may not be straightforward to certify without testing.

The Nuclear Posture Review has opened the door to serious thinking about advanced concepts. The timing could not be more opportune. Our experienced designers are nearing retirement, and before they stop working, they must mentor the new designers. Study of advanced concepts offers a dynamic environment for training and transfer of expertise to a new generation. Unlike stewardship of the last decade, which focused on narrow aspects of weapons physics at times, advanced concepts require thinking through the performance of the system as a whole and thus keeping the integrated design capability alive.

### The Future and the Need for Talent

I believe that stewardship is at a crossroads. In the last decade, we have achieved a great deal without testing and have been able to continue to certify the stockpile. However, we are starting to address physics and engineering issues that may not be so amenable to our present tools. For many reasons, the weapons laboratories are not yet able, unfortunately, to

develop and validate the new tools fast enough. We have several major stockpile systems to maintain (for example, through life extension), and those efforts are as significant a load as any placed on us during the Cold War.

While the national and international environments compel us to maintain, for the foreseeable future, the science, engineering, and manufacture that underpin the existing nuclear weapons capability, we must also envision how the nuclear community might contribute to a more agile and responsive defense without resorting to testing. In other words, we must create the deterrent of the future.

During the past 10 years, we have prepared for these demanding challenges by embracing a strong scientific approach and developing the tools for sustainable stewardship. Now, we need to continue recruiting and nurturing the best talent to solve the wealth of science and engineering challenges that the program faces. The fact that those problems can now be tackled with some of the most advanced simulation and experimental tools available gives us hope. The determination and continued dedication of our staff sustain that hope. ■

*For further information, contact Raymond Juzaitis (925) 423-3096 (juzaitis2@llnl.gov).*



# How Archival Test Data Contribute to Certification

*Fred N. Mortensen, John M. Scott, and Stirling A. Colgate*

The testing of a nuclear explosive was a complex physics experiment with a far richer content than a simple “yes” or “no” answer to the question, “Did it work?” The numerous physics measurements performed during the experiment (see Figure 1) were designed to ascertain what occurred during the nuclear explosion. Detailed knowledge from a series of similar past tests can lead to a number of accomplishments, including the following: (1) a sufficiently convincing understanding of how the weapon operates to enable the Laboratory to certify that it will work as expected, (2) the calibration and perhaps an increased confidence in the simulation codes that are used to assess and certify the performance of weapons in the stockpile, (3) the design of a higher- or lower-yield explosion with the same or with a greater or lesser amount of special nuclear materials, and finally (4) a basis for evaluating and possibly certifying new and untested devices that are near the configuration of the tested devices. Ultimately, data from past nuclear tests corrected and guided our perceived understanding of device performance.

*The photo at left shows a diagnostic rack suspended from a crane as it was being installed into the adjacent tower (white). The tower, which covered the opening to a deep hole drilled specially for the test, would protect the rack against the weather while the diagnostic equipment was placed at strategic locations along the length of the rack. After the rack and the nuclear device had been placed inside the hole, the rack tower was disassembled, and the hole was back-filled with sealing, or stemming, material designed to prevent the blast from breaching the surface.*

The complex fundamental physical laws and interrelated measurements that must be accurately interwoven to explain the performance of a weapon are awesome. The depth of understanding, gained from over a thousand past nuclear tests, is what ultimately gives conviction to the testimony of the nuclear laboratories' directors before the Congress and the nation that our nuclear stockpile is safe and reliable and that it will perform as designed. For the scientist, essential proof of that understanding is the ability to develop a numerical model that accurately reproduces the results of the diagnostic measurements. The models, which are applied to current weapons undergoing aging or manufacturing changes, can only use the nuclear test data that already exist. The ability to answer current stockpile questions is evolving as experience is gained and calculations improve. In the end, our Laboratory director relies on the peer-reviewed scientific judgment of the weapon designers to certify the stockpile.

In this article, we discuss various diagnostic measurements, how they are made, and the information they provide. These measurements were recorded and then preserved as archival data. Today, they represent a major legacy of research that must be employed in the process of certifying aging and altered devices without nuclear testing.

### What Do Diagnostics Measure?

To understand what can be learned from diagnostics, one needs to know how a device operates. A modern thermonuclear weapon consists of four elements: a primary, a secondary, a separating volume, and an enclosing radiation case. Nuclear device operation begins with the initiation of the detonators for the high explosive

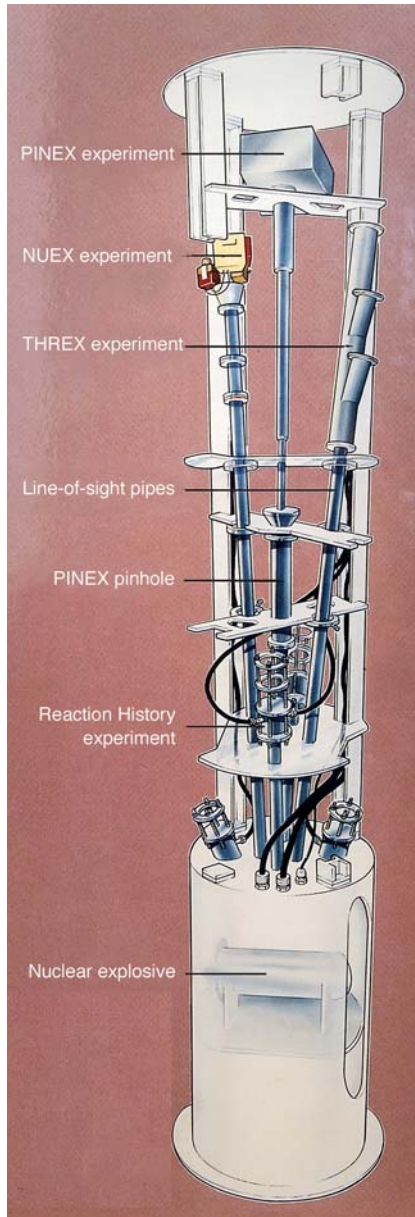


**Figure 1. Line of Sight (LOS) from the Blockhouse to the Bravo Test Site**

The first weaponized version of the hydrogen bomb was tested under the code name Bravo in 1954. Yielding 15 Mt, it was the largest test conducted by the Los Alamos Scientific Laboratory. The design and execution of the diagnostics were performed, however, by Lawrence Livermore Laboratory under the direction of the last author. The view shown is from the block house on Bikini Atoll (housing the detectors and recording oscilloscopes) toward the test site 4 km away. A dozen vacuum pipelines were placed level to provide an LOS between the detectors and the device. At a distance of 4 km, the curvature of the earth is sufficient to occlude the view through the pipe aperture unless the pipes are straight rather than level, a point corrected in some haste. Less obvious was a late worry that a “fireball” of energy might travel along the pipe lines and destroy the block house and recording instrumentation. Such fireballs had been observed many times traveling along the guy wires of the nuclear tests placed on towers (at the Trinity test and later at the NTS). No satisfactory explanation existed. Consequently, additional coral, 100,000 tons, is being piled on top of the block house, a fortunate last-minute correction. Later pictures showed a fireball of 1 kt equivalent energy traveling down the pipe lines to the block house. The block house, equipment, and data survived, but not until 30 years later has a possible explanation emerged: Gamma rays from the bomb, traveling at the speed of light and incident tangentially on the surface of the cable (or pipe lines), absorb and heat the surface of the cable and blow a “hole” in the atmosphere around the cable. Slightly later, a powerful radiation-driven shock wave travels in the air, along the cable and drives a widening wedge of energy into the gap in the atmosphere surrounding the cable. Ever more energy flows into the wedge, and the gap opens in the atmosphere producing a “gap shock” or fireball.

(HE). The HE detonation assembles the nuclear materials of the primary into a supercritical configuration. Once the materials are in this configuration, neutrons introduced into the material will cause fission reactions,

each of which releases 180 million electron volts (MeV) of energy and several more neutrons. In turn, these neutrons will cause more fissions and the release of more energy. As an example, if 1 kilogram of uranium-235



**Figure 2. Diagnostic Rack Layout**  
 This drawing of an underground test rack shows the typical positions of the nuclear explosive, timing and firing instruments, and radiation-measuring instruments. Each custom-designed rack required about 6000 h of effort to build and represented work from all the skilled crafts. Upon completion, the tensile strength of the rack and supporting hardware was tested and certified. Racks weighed up to 300,000 lb when fully loaded. Once completed and certified, the rack was trucked to the NTS on a flatbed trailer.

were to completely fission, it would liberate an amount of energy equivalent to the detonation of 17,600 tons of the explosive TNT. That amount is approximately the energy content in 600,000 gallons of gasoline. Additionally, use of deuterium-tritium (DT) fusion reactions in the primary enhances the fission energy release from the primary, a concept known as boosting.

Most of the energy released in the fission reaction is deposited within micrometers from where the fission event occurred. The release of this energy occurs in nanoseconds, heating the materials in the primary to temperatures of about  $10^7$  kelvins. At these high temperatures, the materials in the primary radiate a large amount of energy (mostly x-rays), similar to an electric stove element glowing red when set on high. This energy can be used for the radiation implosion of the secondary if both the primary and secondary are surrounded by a radiation case that is partially opaque to the radiative energy emitted by the primary. Because the radiative energy leaving the primary cannot quickly escape through the radiation case, it is forced to surround the secondary. As the radiation energy surrounds the secondary, enormous pressures are created, and the secondary implodes, releasing nuclear yield.

Diagnostics play an important role even before a nuclear test occurs. They record the results of hydrodynamic experiments (hydrotests) that aid in the modeling of primary performance. These nonnuclear (or noncritical) experiments examine the implosion of the primary using surrogate nonfissile materials. In other words, hydrotests have the proper geometry of a real device but do not use special nuclear material. In one type of diagnostic, devices called pin domes measure the time of arrival of primary materials at certain locations during the implosion. Because the implosion is spherical, a

pin dome uses a set of wires mounted in the shape of a dome. During the implosion, the electrified wires are short-circuited when the imploding metal contacts the wire. The recording of this signal indicates when material has arrived at the location of the wire and results in a series of measurements that give position versus time. In another diagnostic, pulses of high-energy photons, timed to pass through the primary near maximum implosion, record x-ray-like images of the configuration. Together, the measurements of the HE detonation velocity, the timing of material motions, and the surrogate material positions are a confirmation that the actual primary design produces the calculated supercritical geometric configuration. Those types of data also provide a means to validate the models used for simulating the primary implosion. Because those data are so useful, a significant effort is being put forth to determine the potential of proton radiography for even more precise imaging of hydrodynamic experiments.

Hundreds to thousands of HE experiments and hydrotests have been done and are continuing to be done. The results of those nonnuclear tests are extremely important to certification. They are the cornerstones of primary design because they provide evidence that the assembly of the primary materials into a supercritical configuration proceeds as planned, albeit, using surrogate materials. Of course, age and environmental factors such as temperature can degrade the HE. Given that degradation occurs, the hydrotest becomes a measurement of the robustness of the bomb design in the face of the degraded HE.

In the past, when results of hydrodynamic experiments gave enough confidence in a particular primary design, a nuclear test was used to confirm that the primary worked as models indicated. The high-energy, high-intensity emissions from a

### Getting Out the Signal in a High-Radiation Environment

How far must a detector be from a nuclear device to deliver a clean signal to the recording instruments? Many detectors are typically made of scintillator material. The incident flux of neutrons or gamma rays causes ionization in the scintillator, which converts part of that ionization energy to light. A photomultiplier, or photodiode, converts the light into current, and the current pulse is transmitted through coaxial (coax) cables, like a television signal, to recording electronics, oscilloscopes, or digital recorders protected in a trailer aboveground or a “block house” for atmospheric testing.

Surprisingly, the coax cable itself is the cause of the most stringent restrictions on the distance between detector and device. The reason is that the incident flux of gamma rays can Compton scatter from electrons of the central conductor and produce a spurious signal called the Compton recoil current. That recoil current per centimeter of cable length, must not give rise to a voltage pulse in the cable that is even a small fraction of the signal to be recorded—typically 50 volts, or 1 ampere in 50 ohms of cable.

Let’s first estimate the distance  $D$  at which the radiation from a typical aboveground fission explosion with a 15-kiloton yield would induce a spurious signal level of 1 ampere in a coax cable 1 centimeter in length whose radius is also 1 centimeter. To estimate the flux, or number of particles per second, emitted from that canonical source, let’s assume that one gamma survives from each fission and that the fission rate is one mole per shake ( $10^{-8}$  second), or a 4-kiloton equivalent yield of gammas every  $10^{-8}$  second. That gamma flux is Avogadro’s number ( $6 \times 10^{23}$ ) in  $10^{-8}$  second, or  $6 \times 10^{31}$  gammas per second, or about  $10^{13}$  amperes equivalent flux of charged particles (1 ampere =  $6 \times 10^{18}$  electrons per second). Distance, attenuation, and efficiency for converting gamma rays to a Compton current must all contribute to reducing this flux by a factor of  $10^{13}$ . When these factors are used judiciously, the distances required become kilometers for aboveground testing and meters for underground testing, in which high-density stemming materials are used. However, for safety and signal-to-noise margin, underground dimensions are up to tens of meters (see Figures 1 and 4 in the text).



View of the coax cables looking down from the rack tower.

device during a nuclear test, including gamma rays, neutrons, and x-rays, present a different measurement problem than the signals in a hydrotest. The radiation flux from a nuclear explosion is so large that, even before reaching its peak, the flux would destroy any detector placed close to the explosion. That destructive potential has led to the complicated geometry of the diagnostic racks (Figure 2) of test equipment. These racks are lowered to the bottom of a hole, typi-

cally a few thousand feet deep. Detectors for recording peak signals are placed at the top of the rack, each with a view of the device through a long line-of-sight (LOS) pipe. Many neutrons and gamma rays from the nuclear explosion scatter within the rack, thereby producing additional particles that can interfere with the collection of the desired data. Shielding materials placed in the rack to protect the diagnostic experiments are designed to attenuate these extra-

neous fluxes of gamma rays, the slower neutrons, and delayed x-rays, allowing the desired signals from both the primary and secondary to get to the detectors without contamination. Atmospheric testing from the 1940s to the 1960s required longer LOS. In the Bravo-Shrimp test, the nation’s and Los Alamos’ largest thermonuclear test (15-megaton yield), vacuum pipe lines (long pipe lines from which the air had been removed) 4 kilometers long were used to give a highly colli-



**Figure 3. PINEX Measurements**

(a) The PINEX camera includes a pinhole assembly (b) that focuses neutrons from a nuclear explosion onto a piece of fluorescent plastic. The plastic produces fluorescent light in proportion to the neutron fluence striking it. Modified TV cameras view the pattern of light through reflecting mirrors and record the image. Before the TV cameras are destroyed by the shock of the explosion, the PINEX image, which is usually only one frame, is relayed to recording instruments above-ground. (b) This PINEX “lens,” or pinhole assembly, is made of tungsten, a metal that shields unwanted neutrons. The size of the hole regulates the number of neutrons passing through it. Changing the position of the pinhole assembly varies the image size. (c) This calculation of PINEX data shows intensity levels (by color) of the neutron fluence as measured by the light from the scintillator. The color levels show intensity levels differing by 10%.



mated view of the nuclear reactions. At that distance, the signal-recording detectors escaped most of the damaging radiation (Figure 1). In addition to measuring gamma rays, neutrons, and x-rays emitted by the device, diagnostics can measure the effect of a device. For example, measuring the ground shock of an underground test allows one to infer the device yield.

During a nuclear test, the start of criticality is observed as the exponential growth of either neutrons or gamma rays from the nuclear core. The neutrons result from fission, and

the gamma rays result from fission or the interaction of fission neutrons with other elements. A diagnostic known as a reaction history measures the gamma-ray flux with good time resolution. Because the flux varies over many orders of magnitude, measuring its time history is quite a feat. Those data provide a time history of the criticality of the device, a quantity known as alpha. The prediction of alpha is one of the most exotic calculations in all of physics—it requires simultaneously modeling the hydrodynamics and the transport, absorption,

and multiplication of the neutrons by fission and fusion burn. Thus, the measurement of alpha at various points in time during the exponential growth of neutrons from fission and fusion becomes a critical diagnostic of the implosion and explosion. The measurement indicates how the fissile material becomes supercritical and explodes. Usually, separate LOS on the diagnostic rack are used to measure the reaction histories of the primary and secondary. This measurement is considered so important that it has been taken on every nuclear test event since Trinity. The interval time, roughly the time between primary and secondary operation, can be assessed from reaction history measurements of the primary and secondary.

A NUEX (for neutron experiment) measures neutron output versus time. That measurement has lower time resolution than a reaction history measurement because the time of flight of neutrons from their point of emission to the detector is longer than the time during which they are produced. Because a neutron’s velocity is proportional to the square root of its energy, NUEX is a measure of the time-integrated neutron energy spectrum from the device.

PINEX, for pinhole camera experiment, uses a pinhole camera to image neutrons (or sometimes gamma rays) from a device (Figure 3). The experiment can image all neutrons over time or may be gated in time to measure only the 14-MeV fusion component of the neutron spectrum. (Time gating is possible because, again, the velocity of a neutron scales with the square root of its energy.) PINEX gives a time-integrated but spatially resolved image, indicating where neutrons are being emitted from a device. Essentially, it can give the shape of the regions in a device where neutrons are being produced. If PINEX is gated to measure only the 14-MeV neutrons,

the result of the measurement will indicate where DT fusion reactions are occurring.

A THREX (for threshold experiment) measures neutron output versus time from DT reactions. As a material containing both deuterium and tritium becomes very hot (about  $10^7$  kelvins), fusion reactions will begin to occur, which will produce 14-MeV neutrons. Some of these neutrons will escape the device and can be detected. Since the rate at which DT fusion occurs increases dramatically as temperatures rise above  $10^7$  kelvins, the rates at which neutrons are produced, escape, and are detected are also very sensitive to the temperature at the location where the detected neutrons were produced. Consequently, from measurement of the escaping DT neutrons, a temperature can be inferred.

Radiochemistry is a diagnostic technique that employs the effects of the neutrons emitted from the device. Small amounts of material (radiochemical tracers) that readily transform to different isotopes when exposed to a flux of neutrons are positioned in various places throughout the device. These isotopes subsequently decay radioactively, but the decay time is long compared with the time required to recover material from the explosion. The relative abundances of the products after the nuclear explosion compared with the initial amount of material are a measure of the time-integrated neutron flux at the position of the radiochemical tracer. Another important measurement provided by radiochemists is known as  $\Delta P$  and does not rely on additional radiochemical tracer materials. It measures the change in the ratio of plutonium isotopes. That change is a sensitive measure for the number of fissions that occurred in the plutonium. Knowing the number of fissions allows one to calculate the fission yield from the plutonium. Radiochemical samples were recov-

ered from an underground test through a process known as drillback. That is, core samples were drilled from the bomb residue left after the explosion and the collapse of the cavern. The samples were then chemically separated and radiologically counted to measure the relative abundance of all the material isotopes from the device. Tracers of different materials were used to prevent cross contamination from tracers in different regions of the

*One cannot emphasize strongly enough that simulations cannot be undertaken with meaningful expectations unless the diagnosis of the physics that occurred in the device and the modeling of the physics are both understood in great depth.*

device. The radioactive decay products and beta or gamma energies, are a unique signature of the specific isotope of an element. These types of data are generally referred to as integral measurements.

Many more diagnostic techniques were used to assess how a device operates, but those described above are generally emphasized in present-day comparisons of simulations with archival test data. Ultimately, all diagnostic results from a nuclear test contribute to our understanding of a particular device. The acid test of this understanding is whether our numerical simulation agrees with the experimental signals or data. We must understand whatever differences exist between the simulation and the experiment if we are to gain

confidence in our ability to predict device performance. One cannot emphasize strongly enough that simulations cannot be undertaken with meaningful expectations unless the diagnosis of the physics that occurred in the device and the modeling of the physics are understood in great depth.

## Getting to the Nuclear Test

Placing a test on the nuclear test schedule was a complex process not always governed by a quantifiable set of reasoned criteria. We were always facing a limited budget to address a seemingly limitless set of questions. Therefore, placing a test on the testing schedule was a balancing act between the slate of questions and our priorities. What diagnostics will be needed to obtain the data necessary to answer the question? How much volumetric real estate in the rack will the necessary diagnostics take and not interfere with other diagnostics? How much of the test budget can we spend on this shot and still do the other necessary nuclear events? To explain how a proponent for a test worked through all the politics in the above set of questions is worth a paper in its own right. For this article, we will assume that a test (consisting of a nuclear device with both a primary and a secondary) is on the schedule and then sort the remaining questions by considering the physical and technical needs required to obtain the necessary data.

In general, after a nuclear test had been officially placed on the testing schedule, a nuclear test team would be set up. This team would consist of personnel from the design division (for design), physics division (for development and deployment of diagnostics), engineering division (responsible for providing actual bomb parts and the assembly of the parts into a usable test object), and testing divi-

sion (the people responsible for overseeing anything happening at the test site and supplying Nevada Test Site (NTS) support people such as crafts people, crane operators, and stemming teams). Interactions among all these organizations were necessary to ensure that a successful and safe test would take place. This team would develop a test plan addressing what could be done within the allowed budget and time constraints.

The design division team would generally consist of a primary designer, a secondary designer, a diagnostician, and any additional team members needed to support this group. This team was responsible for developing the total nuclear design of the device to be used for the test. Members would work closely with the engineering and physics team. The engineering team would generally consist of a primary engineer, secondary engineer, and an assembly engineer, as well as an assembly team. The primary engineer was responsible for producing the necessary primary parts, just as the secondary engineer was responsible for the secondary parts. The assembly engineer was charged with building the whole collection into a working nuclear device with the help of the assembly team. The physics team would generally consist of a diagnostic physicist for each required diagnostic experiment fielded on the nuclear test and any additional experimenters needed to support that work. The physics team worked very closely with the testing division to ensure everything came together correctly at the NTS.

Staff of the design and engineering divisions would get together to determine what the nuclear device would be and the features or properties that would be needed to address the goals for the test. Then staff from the design and physics divisions would determine the best diagnostic experiments required to obtain the necessary data



**Figure 4. Tower and Cables before Lowering the Rack**

**This aerial photograph shows a diagnostic rack tower in the distance. Next to the tower is the crane that would lower the rack into the hole drilled for the event. Cables from the rack were snaking a long distance to a trailer park, which contained the instruments recording the information from the diagnostics. Once the diagnostic rack had been fully prepared, miles of cables, literally, were used to connect the downhole diagnostics to the recording trailers aboveground.**

for addressing those goals. These people would also define the size of the nuclear test rack necessary to hold the test device and the accompanying diagnostics.

In designing the total experiment, one had to decide which detectors, instruments, and recording devices should be up close and which ones should be far away. How close, how far away, and how to connect the two determined the geometry of the experiment. The diagnostics for the HE do not raise this question because the detectors must be adjacent or buried in the HE, and fortunately the signals can be transmitted in ordinary coax cable (like TV cable) or fiber-optic cable to oscilloscopes or digital recorders in a bunker or trailer that can be far away—in some cases, miles away. This signal (current versus time) travels at two-thirds the speed of light in coax cable. With a typical time of about 100 microsec-

onds between the HE detonation and the nuclear yield, there was plenty of time for the HE signals to escape the radiation from the bomb and safely reach the recording bunker. The cables carrying later signals must be shielded against the radiation from the explosion (see the box “Getting Out the Signals in a High-Radiation Environment” on page 41). The attenuation in the ground for underground tests or in air for the atmospheric tests also helps shield the signal cables. All these factors determined the geometry or distance and LOS for the detectors in the racks underground (or in the air, for aboveground testing).

To prevent further pollution of the environment by atmospheric tests, nuclear testing was finally confined to the underground at the various test sites around the world. At the NTS, a hole, similar to a large-diameter oil well, was drilled into the alluvial sediments, and its depth depended on yield. The device was placed in a rack and lowered to the bottom of the hole. The many signal cables from the rack led to trailers of recording instruments. These trailers were located far away from the hole to prevent their falling, with recorded data and all, into the large crater that sank into the earth after an explosion. That subsidence crater marks the collapse of the underground cavity created by the explosion. Figure 4 shows the trailers of equipment and the many signal cables snaking around on the surface. The cables were fed downhole as the rack, with its detectors and bomb, was being lowered into the ground.

The diameter of the hole was generally determined by the type and number of diagnostics and their individual complexity, as well as by how difficult it would be to isolate (shield) the individual diagnostics from the other diagnostics within the rack. The diameter of the hole could vary from 4 feet (for a relatively simple shot) to 12 feet. The depth of the hole was a

function of the predicted total device yield. When a nuclear device exploded in an NTS rack, many neutrons and gammas that escaped the device were examined by diagnostic experiments. Generally, diagnostics have collimated LOS pipes looking from the experiment position to a particular device position (Figure 2). The particular particle or ray being investigated comes up the LOS. However, there are many neutrons and gammas scattering within the rack, producing additional particles that can interfere with the collection of the desired data. Isolating or shielding the individual experiments from the crosstalk induced by original bomb neutrons and gamma or secondary particles induced by scatter within the rack was therefore of major importance. The design division's diagnosticians would also play a big part in these decisions by calculating the crosstalk between the proposed diagnostic LOS. Once a rack had been lowered into a hole (Figure 5), the hole would be stemmed to contain the exploded bomb debris after the shot was fired. This stemming consisted of layers of magnetite, sand, concrete, and epoxy. The exact stemming process was experimentally determined from a large number of NTS shots and was dependent on the location of the hole within the NTS.

### What Is Done with These Measurements?

The analysis of the numerous measurements collected results in a deep understanding of how the device operated. Typical questions that test diagnostics answer and that can later be compared with simulation results are the following: Did the multiple detonation points of the HE initiate a correct detonation wave? Is the arrival time of the first neutrons or the time from HE initiation to the



**Figure 5. Lowering the Rack**

**The rack and a large number of extremely long cables were carefully lowered downhole from the surface. During the rack's emplacement, care had to be exercised to ensure that the cables maintained connection between the downhole equipment on the rack and the recording instruments in the trailer park. The cylinders on the cables are gas blocks that would prevent the flow of downhole gases through the cables into the atmosphere.**

time that the fissile material reaches criticality correct? What is the multiplication rate  $\alpha$  of the fission criticality? What is the peak of the alpha curve before boost? When does boost occur? What is the boosted yield of the primary? What is the time between primary and secondary operation? What are the temperatures measured in the device? What is the multiplication rate  $\alpha$  in the secondary? What is the total yield measurement from ground shock? Does the radiochemistry indicate the same yield? Does the radiochemistry indicate the predicted distribution of neutron fluxes?

Many other measurements contribute to the understanding of a device. The total number of measurements for each test, when combined

with the possible judgments regarding each of these measurements, yields an astronomical number of permutations. Designers must be aware and able to speak to all the realistic possibilities, using their informed judgment. For certification of a device, designers will choose to simulate a suite of nuclear tests that encompass the body of relevant data associated with the device. Ideally, simulations are generated that reproduce the diagnostic measurements for each nuclear test. In practice, this may not always be true, and subjective judgments are made regarding the validity of calculations that may not fully reproduce the experimental data. However, when designers believe that a set of satisfactory calculations exists for the suite of tests, a certification judgment of the device is made. This process generally takes years and undergoes peer review. The peer review process assesses whether designers may have made obviously incorrect assumptions about the physics intricacies associated with the device. Designers must convince a peer group that the device operates as they understand it does. Although the focus of attention is on the responsible designer, it takes a cast of many from groups across the Laboratory to certify a device for the stockpile. Their work, in addition to that of the designers, ultimately leads to the Laboratory director's signature on a weapon certification statement.

Today, without nuclear testing, we rely increasingly on simulation tools to provide the necessary answers to maintain a safe and reliable stockpile. Advanced Simulation and Computing (ASCI) is developing state-of-the-art computing facilities and a new generation of simulation tools to mitigate the effects from the moratorium on nuclear testing. Although currently less mature in capability and usage than the suite of tools (legacy codes) that gained general acceptance up to

the end of nuclear testing, the new codes have contributed to some significant accomplishments. The various ASCI codes have demonstrated capabilities beyond those of the legacy codes in various milestone calculations. They have been and continue to be used as a tool in the resolution of current stockpile issues. Ultimately, the success and acceptance of these new codes will depend on their ability to match the diagnostic information from previous nuclear tests, as well as experimental data from today's ongoing experiments. As these new tools gain widespread use and are tested on more complex and challenging problems, their relative importance will evolve. Weapon designers will continue to use the legacy codes to solve current and future stockpile problems. The newer ASCI codes will supplement the legacy codes until the ASCI codes are validated. The validation will be done against past NTS data as well as newer data from ongoing experiments. Without new nuclear tests, the most difficult problem will be to develop, using available experimental facilities, the physical models that describe behavior consistent with the conditions found in a nuclear device. As part of the model development process, designers will draw on valuable diagnostic information from nuclear test data to help confirm a model's validity.

The purpose of the diagnostic measurements is to develop an understanding of all the physical processes that conspire to make a nuclear explosion possible and reliable, including processes that make a device safe. These very complicated measurements were performed many times in the past. Archived data from them have been the basis for the development of the most sophisticated, lightweight, high-yield devices currently imaginable.

In summary, data of many types

taken on over a thousand U.S. nuclear tests are essential to the understanding of how nuclear weapons work. The physics taking place within a thermonuclear weapon during its implosion and explosion is an extensive, highly nonlinear, closely coupled set of processes. Understanding these processes by numerical modeling requires that the modeling be able to reproduce the measured data.

We have discussed how certain types of data are used in the attempt to understand the workings of weapons (currently in the stockpile). Acquiring additional data from small-scale experiments and nonnuclear integral tests is currently the only way to answer some questions for which no specific NTS data exist. Confirming that those new data are accurate and applicable to weapon issues is a very difficult procedure. The Laboratory is applying that procedure today. Accurate and complete archiving of those data (old and new) is vital to the continuing effort to maintain a safe and reliable stockpile in which we have confidence. Those data are the cornerstones of the calculational effort needed to continue certification into the near future.

The Laboratory has taken on the challenge to maintain and continue to certify the U.S. nuclear stockpile, and the Laboratory staff works daily toward that goal. Without nuclear testing, however, weapons performance cannot be demonstrated as in the past. ■

*For further information, contact  
Stirling Colgate (505) 665-5254  
(colgate@lanl.gov).*



# QMU and Nuclear Weapons Certification

## *What's under the hood?*

David H. Sharp and Merri M. Wood-Schultz

In the article “Science-Based Stockpile Stewardship,” Ray Juzaitis has described the main elements of the nuclear weapons program as it is taking shape today. The cornerstone of the program is nuclear weapons certification. Our purpose in this article is to explain our approach to certifying nuclear weapons in the posttest era.

Full-system nuclear tests and conservative designs have provided a high degree of confidence that stockpiled nuclear weapons will perform safely, reliably, and to specifications when their condition and use are within the tested envelope. Confidence was established through the certification procedure, whose outcome was a guarantee that the stockpiled weapon will achieve specific performance levels (military characteristics) under stipulated operating conditions (stockpile-to-target sequences, STS). Scientific judgment plays a critical role in determining the *sufficiency* of the criteria on which a certification is based and, to some extent, in determining whether the criteria have been met.

The methodology for certifying nuclear weapons has always included aboveground experiments, nuclear tests, and simulations of weapons operation. These elements were tightly interwoven, and no single element was sufficient by itself. Full-system tests were particularly important, however, in that such tests swept away many,

although not all, uncertainties about the performance of nuclear weapons.

The need to answer questions about the stockpile has not gone away with the cessation of nuclear testing. For example, aging can alter the state of a weapon, and although not all observed aging defects are serious, some may be. Evaluating the effects of aging becomes increasingly important as weapons are kept in the stockpile well beyond their designed lifetimes. Similar questions arise concerning the effects of manufacturing or design flaws that may come to light, as well as the effects of planned refurbishments and modifications. Questions arising from the possible need for weapons of new design are looming.

Thus, there is a compelling need for assessments of how weapons will perform in an untested configuration. That is the problem. Plainly, there is no complete substitute for nuclear tests as a source of confidence in such assessments. Developing predictive capabilities that can support certification in the posttest era is therefore a tremendous challenge. Can this challenge be met with an improved scientific understanding of the behavior of nuclear weapons derived from a new generation of large- and small-scale nonnuclear experiments, better physics modeling, and more powerful computing? In this article, we will look at what needs to be done to answer this question, starting in the

next section with a discussion of quantification of margins and uncertainties (QMU), a methodology created to facilitate analysis and communication of confidence in an assessment or certification.

Confidence is so central to certification that the use of predictive simulations in this context needs to be discussed first. Confidence in predictions of nuclear weapons performance, as with all scientific predictions, will be based on the track record, that is, on the scope and success of past predictions. But in matters concerning health, safety, or security, the cost of incorrect predictions can be very high, and one will often have just one chance to get the right answer. In such cases, the issue of confidence in prediction comes up with particular force, as compared with cases in which predictions are used mainly to guide the development of science. In both cases, one wants correct predictions; it is the consequences of incorrect predictions that are different.

An analogy can be drawn between nuclear weapons certification and Food and Drug Administration (FDA) approvals of new drugs. Just as the FDA requires demonstration of actual efficacy before approving a new drug, we require *positive evidence* that a nuclear weapon will work; absence of evidence that it will not work is not sufficient. Likewise, just as the FDA requires documentation of contraindi-

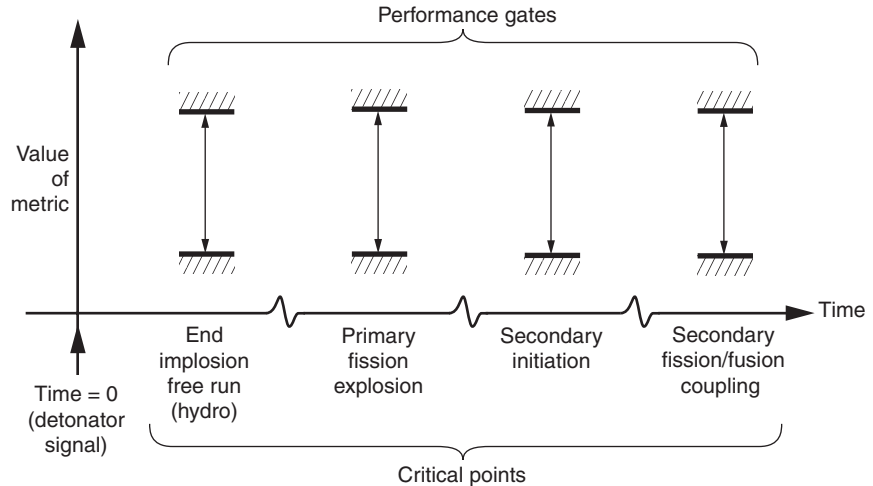
cations and side effects, leading to a lot of fine print in drug advertisements, our certification and validation studies come with some fine print. All this is not mere fussiness; in both cases, the driving force is the need for high confidence in predictions about the behavior of very complex systems.

### The Role of QMU in Maintaining Stockpile Confidence

QMU, currently under development at Los Alamos and Livermore National Laboratories, is a framework that captures what we do and do not know about the performance of a nuclear weapon in a way that can be used to address risk and risk mitigation. The QMU framework is explained here in its simplest form, for example in a deterministic rather than a probabilistic form. Like any other part of science, QMU will evolve on the basis of experience gained through its use in actual applications.

The basic idea of QMU is to evaluate confidence in terms of the degree to which the operation of a weapon is judged to lie within “safe” bounds on judiciously chosen system or operating characteristics. A useful operating characteristic might pertain to the system configuration at a critical juncture in its operation, or it could relate to a time-dependent or time-integrated characteristic of the system.

QMU does not determine predictions or their uncertainties *per se*; these are inputs to a QMU analysis. It is designed to analyze and communicate the confidence in a conclusion or decision based on those predictions. Confidence—not the rigorous statistical determination of confidence intervals but the intuitive concept—is intrinsically hard to quantify. It is typically determined through a mental “calculation,” weighing factors that may well differ from person to per-



**Figure 1. An Illustrative Timeline for the Operation of a Thermonuclear Weapon**

son. The person producing an assessment may not even realize all the factors that were considered. This degree of fuzziness can confound any attempt at analysis of confidence, its quantification, or its communication. It is clearly necessary to identify well-defined characteristics of a system on which discussions of confidence can be based.

“Characteristic” is a broad term. It can mean any function of the physical variables determining the performance of the system. Such characteristics can be static or dynamic, measured or calculated, intuitively clear or obscure. To give just one example, the amount of fissile material in a primary and its peak compression during operation are characteristics of a nuclear weapon.

In QMU, the characteristics of a system that are used to evaluate confidence are termed “metrics.” Metrics and the other basic concepts necessary for QMU will be discussed with the help of Figures 1 and 2. Figure 1 shows a schematic timeline for the operation of a thermonuclear weapon, along with a schematic application of QMU. Four metrics are shown. The first one—the pit energy—is an operating characteristic of the system at the time when the kinetic energy of the imploding pit is at its maximum, so this metric is based on a snapshot

of device behavior. In contrast, the system yield clearly is a characteristic that depends on the entire history of the device operation. What all metrics have in common is that they are high-level indicators of some aspect of the system’s operation.

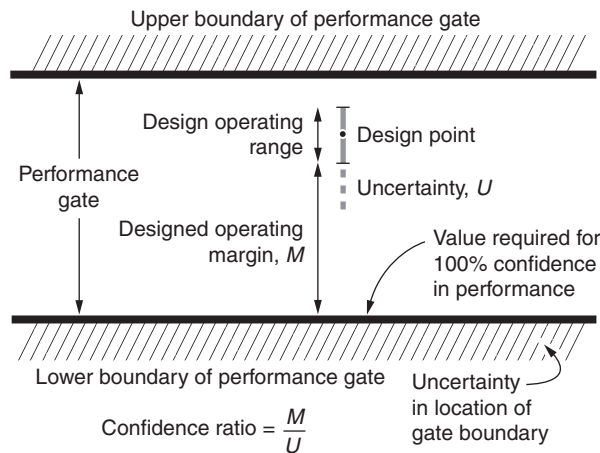
Defining useful metrics requires an understanding of the strengths and weaknesses of the tools used to evaluate the metrics. We define a complete set of metrics as one that, taken as a whole, is sensitive to *all* the important and potentially inadequate aspects of the simulations and measurements used in the evaluation. Data must be available to validate a useful gate for each of these metrics. Together, these requirements will affect the scope of weapons issues that can be addressed.

The process of evaluating a metric is conceptually straightforward for both measured and calculated metrics. Determining the uncertainty in the metric being evaluated is also relatively straightforward for a measured metric, but not for a calculated metric. In the latter case, each aspect of the calculation—databases, physics models, and numerical methods—may have errors. Error and uncertainty in predictive simulations are thorny problems, discussed in “Estimating Uncertainties” below.

The vertical bounds associated with each metric in Figure 1 represent the range of values for that metric that are judged to be acceptable. This range is termed a “gate” (see Figure 2). Metrics and gates are intended to delineate safe parameter regimes. Each metric used in QMU must be assigned an appropriate gate. It is clear that setting an appropriate gate is crucial to using QMU successfully.

The procedure for setting a gate is to evaluate the metric for a set of successfully tested configurations. It is important that this set include variations in whatever parameter is being considered to ensure that the effects of variations in that parameter are represented in setting the boundaries of the gate. The set of metric values used to set the gate boundaries is thus known to correspond to successful performance, and the conservative presumption is made that systems having metrics outside this range—systems for which this metric does not fall within the gate—will not work. Nuclear test data are absolutely essential in defining valid gates. There is no substitute within the constraints of existing predictive capabilities. As much nuclear test data as possible are used to maximize confidence in the location of a gate. The criterion for successful device operation supplied by QMU is that system performance must lie within *all* the defined gates. Confidence that this criterion has been met derives from the “safety” margin at each gate.

Margin is simply a measure of how much “room” is left between a metric at the limit of its operating range and the boundary of its gate. The details of the uncertainty in a metric, its gate, and the resulting margins are shown conceptually in Figure 2. We note that a range of metric values, called the designed operating range, results from



**Figure 2. Key Elements of a Performance Gate: Boundaries, Margins, Operating Range, and Uncertainties**

the effects of different STS environments and manufacturing tolerances on device operation and from the effects of any intrinsic variability on device performance.

The confidence ratio (CR), defined at the bottom of Figure 2, summarizes the situation at each gate. The CR is simply the ratio of the margin in a metric to its uncertainty. The gate where the CR is smallest is the aspect of performance most likely to be or to become problematic—the weakest link. If we were very sure that we had not underestimated the uncertainty  $U$ , then 1.0 would be an acceptable CR. However,  $U$  is generally known imprecisely. An acceptable CR, therefore, will depend on scientific judgment as to the accuracy of  $U$ . The use of scientific judgment is common throughout science, and its role in weapons certification will be discussed in more detail below.

The CR is an example of a figure of merit, and other figures of merit could easily be defined. Confidence in overall device operation is represented by the CRs for the entire set of gates, and any CR approaching 1—a weak link in the chain—is a warning flag.

Three QMU functions—defining metrics, setting gates, and evaluating

uncertainty—have been discussed, and acceptable confidence was quantified in terms of the CR. We stress that this confidence derives not from the QMU formalism itself but from the quality of the science used in applying it.

Even if QMU gives an acceptable CR, a fundamental question remains: Is this procedure *sufficient* to guarantee acceptable performance? This issue was touched upon above as the requirement that *all* the important vulnerabilities be adequately constrained by the QMU

metrics. Whether or not this has been done can only be based on expert, or scientific, judgment. Scientific judgment is always essential in reaching a decision on the basis of incomplete or inconclusive evidence and therefore has always played a significant role in certifying nuclear weapons. The foundation of science is that experiment is the sole judge of truth, but the use of expert judgment is legitimate when it is *provisional*, in the sense that it is subject to challenge and correction through the scientific process itself. In a posttest era, we must continue to rely on expert judgment. Expert judgment can still be challenged on the basis of nonnuclear tests, predictive simulations, and peer review, although the standards are softer than the ones set by full-system tests.

QMU is designed to facilitate such challenges to expert judgment. It is also flexible enough to incorporate all the criteria on which a certification might be based, and that is why it can be used as the methodology for certification. It is important to keep in mind that QMU, like other tools, does not determine the adequacy—or in this case, the sufficiency—of the product it is used to create. That aspect still depends on the craftsman (the designer) and the raw materials (the data).



## Determining the Behavior of a Complex System

There are basically two ways in which to learn about the integrated behavior of complex, real-world systems: One is full-system testing (observation), and the other is full-system simulation. Full-system tests and simulations are complementary but are not interchangeable. Full-system tests always provide more confidence than simulations because they give a definitive answer as to *whether* a particular device worked under the specific conditions realized in the experiment. Also, they typically provide some detailed data on the internal conditions during the experiment. However, full-system tests are not equally definitive in telling *how* the device works, and it can be prohibitively expensive to explore device behavior over a broad range of operating conditions in this way. In contrast, full-system simulations are usually cheap (compared with hardware), have zero risk, are controllable, and allow access to the details of the physical processes. The price of these advantages is the need for complete and precise knowledge of the operation of the system. Because such knowledge is often not available for complex systems, simulations come with myriad opportunities for errors.

Increasing the scope of stockpile-related questions that can be answered with confidence and without nuclear testing requires that the boundary between what can be reliably established by full-system simulation and what must be proved by a full-system test be shifted. Correct and reliable prediction using a simulation presents two core issues: One concerns data, and the other concerns the integration of information pertaining to subsystems into full-system simulations.

**Experimental Data.** Data are needed to define initial conditions and parameter values for specific prob-

lems and to validate or constrain models. For complex problems, a lot of detailed data are needed to validate models for predictive purposes, and the data requirements go well beyond what is needed for interpolation. Even in a laboratory setting, detailed quantitative data about fluid motions, for example, are often hard to come by. The problem of getting well-diagnosed, accurate data is very much more difficult for nuclear weapons because they operate in a regime that is far from laboratory conditions.

For complex systems, data are usually sparse, relative to the need, so it is often necessary to combine data from multiple, diverse sources when testing a model. Some of the data requirements are being met through integral experiments at facilities such as the Dual-Axis Radiographic Hydrodynamic Test (DARHT) Facility at Los Alamos, the Z-machine at Sandia National Laboratories, the Omega Laser at the University of Rochester, and eventually the National Ignition Facility (NIF) at Lawrence Livermore National Laboratory and through subcritical underground experiments conducted at the Nevada Test Site (NTS).

Laboratory-scale experiments can play a vital role in building predictive models, although their usefulness is occasionally underestimated. Sometimes, model parameters can be determined from a more basic theory. For example, the viscosity coefficient appearing in the Navier-Stokes equation could be calculated from kinetic theory. However, archival nuclear test data remains the crucial core of the data used for certifying weapons because it provides the only faithful integration of the interactions among the various parts of a functioning device.

**From Subsystems to Full Systems.** The task of building an understanding of full-system behavior from a knowledge of component subsystems is one of the most difficult

aspects of modeling complex systems. Multiscale science, that is, consistent representations of physical processes that extend over more than one length (or time) scale, is often a problem. A completely adequate “microscopic” model, which can include properties that profoundly influence large-scale behavior, is often not feasible for use in full-scale studies, so a method is needed to incorporate the essential fine-scale information into macroscopic simulations. Areas of weapons physics where this issue arises include modeling the initiation of high explosives, materials damage modeling, and the fluid-mixing problem.

The next part of the integration problem is to model full-system performance of a complex device starting from models of the individual components or processes. We refer again to Figure 1, this time to illustrate how one might decompose a complex system into pieces that can be studied independently or, at least, conditionally. The initiating event in a nuclear weapon (seen at the far left of the timeline in the figure) is the detonation of a high explosive (HE). The physics and chemistry of detonations are extremely difficult subjects, which have been studied at Los Alamos since World War II. Fortunately, the HE detonation is unaffected by the physics occurring in the nuclear regime of device operation, so HE can be studied and modeled using information from aboveground (nonnuclear) experiments. Doing so allows an HE detonation model to be developed and tested independently of the downstream physics.

The next step in Figure 1 is the pit implosion, during which the flow of dense materials occurs. Because the material flow is driven by the HE, the pit implosion is conditional on the HE simulation, yet it is, at this point, still independent of nuclear-phase processes. Like the HE model development, laboratory experiments and large-scale

experiments (for example, those at DARHT) can supply extremely useful guidance in modeling the flow of dense materials.

Two observations are appropriate at this point. The first (and rather obvious) observation is that an accurate full-system simulation must be built from accurate, reliable models of the individual processes that are occurring: detonation, material flow, neutronics, and so on. There is no reason to think that coupling together poorly modeled processes or subsystems will produce anything but a poor model of the whole system. The second observation is that accurate modeling of the coupling of different physical processes, for example, of neutron and radiation transport to material flow, can itself be very difficult to achieve.

The timeline in Figure 1 continues to the time when criticality is achieved. A calculation must now couple material flow to nuclear and thermonuclear processes, for which the data is not as detailed or systematic as that available for the earlier processes. We are left without a guarantee that predictive models can be validated for this late-stage operation, but steps can be taken to improve our understanding.

The first step has already been stated: Begin with a good model. A good model, say for fluid mixing, will be internally consistent and will agree with a broad range of results from large- and small-scale nonnuclear experiments, with few if any adjustable parameters. The validation experiments must include all the important aspects of the weapons process—for example, change of flow from laminar to turbulent—and at a variety of parameter values demonstrate predictive capability.

A model that works well in the laboratory regime is not necessarily correct in the weapons regime. But one can still test the model postdictively in the weapons regime, using comparisons with a portion of the NTS data-

base to constrain any free parameters and the results of applying the model, with no additional parameter adjustment, to the remaining NTS data as evidence of the model's predictive power. If sufficient data exist, this procedure will provide a fairly good means for establishing confidence in models for the explosion phase of operation. However, if a bottom-line result reflects a sensitive dependence on initial conditions or other problem parameters, then its reliability may be subject to question.

How far will all this take us toward meeting the goal of a predictive capability for assessment and certification? This will certainly depend on the question one is trying to answer; we will be able to deal with some questions using predictive science but not with others. The boundary will be set by the scope and power of the predictive models that we are able to develop—an explanation that requires an explanation.

Scope refers to the number and variety of cases in which the theory has been tested. Knowing the scope is important in building confidence that one has identified the factors that limit the applicability of the theory. Power is judged by comparing what is put into the model with what comes out. Theories that correctly predict a wide range of phenomena with just a few input parameters are powerful; phenomenological models—those that are calibrated to data and hence closely tied to specific problems in their formulation and predictions—are less powerful. Nevertheless, they are extremely useful, and are in fact the default solution to the problem of producing assessments when adequate fundamental models are not available. Monitoring progress toward predictive capability is the job of validation and uncertainty quantification, the topics of the next section.

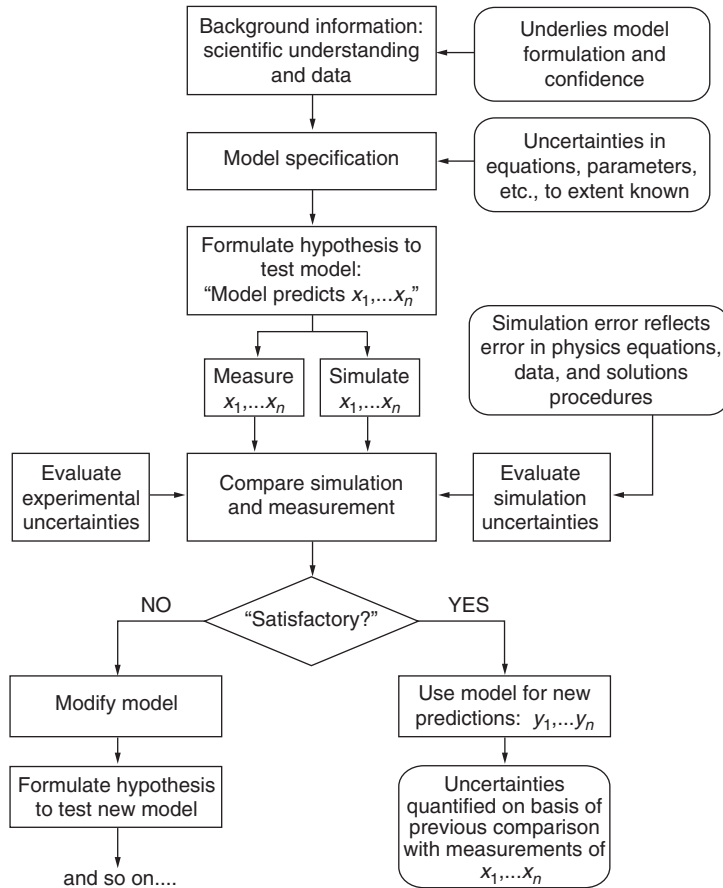
## Estimating Uncertainties

Nuclear weapons performance is calculated using complex computer programs, or codes. These codes combine databases for various physical quantities (equations of state, opacities, and so on), multiple physics models, and algorithms for solving the physics equations to calculate the operation of the weapon, given its initial state. Like all codes, weapons codes are approximate representations of reality. As we call on them for actual predictions, as opposed to interpolations or small extrapolations, to help answer questions about weapons that deviate from their tested condition, knowing how accurately the codes describe the real world, that is, knowing the error in code predictions, becomes of paramount importance.

Validating a code is not like proving a mathematical theorem. Nuclear weapons simulation codes must simulate coupled, nonlinear, multiscale physical processes, and the most important and difficult-to-model aspects of weapons behavior (which occur during the explosive nuclear-energy production phase) are not accessible to laboratory experiments. This leads to reliance on integral data from nuclear tests and to the additional complication of having only indirect inferences about weapons behavior from this data.

Nevertheless, determining uncertainties in simulation-based predictions revolves around the answers to a few basic questions: What do you need to predict? What factors can lead to errors in the predictions? How can you get a handle on these errors?

Errors in predictions can come from poor-quality input data, incomplete or insufficiently accurate physics models, and inaccurate solutions of the governing equations. Some of these, for example, equation-of-state errors, error models for material damage or fluid mixing, instrumental



**Figure 3. The Scientific Method—the Basis for Quantifying Uncertainty in Predictions**

errors (as in DARHT or NIF), and errors in numerical solutions can be determined through standard experimental methods with sufficient resources. Solution errors are a distinctive feature of predictions made using large-scale simulations. They contribute to the total error in a prediction and must also be considered when drawing conclusions from comparisons of data with predictions. As discussed in “The Role of QMU in Maintaining Stockpile Confidence,” the explosion-phase physics is problematic. In addition to the calculations being exceptionally difficult and the physical regime being inaccessible in the laboratory, there are significant uncertainties about some aspects of the physics.

Determining the error in a simulation is directly analogous to determining the error in an experiment. A direct determination of error is sometimes possible for simple experiments and for simple simulations by comparing a measurement or prediction with a standard or an analytical solution. In contrast, errors in complex experiments and simulations must be calculated by breaking down the end-to-end operation into components (subsystems) amenable to separate error analysis. The subsystem results must then be painstakingly combined to produce the overall uncertainty in the specific quantity of interest. Although this procedure is obligatory in complex situations, it has the virtue of showing which sources of error are

most influential and of providing guidance for reducing the errors one by one. Incremental progress will be made at mitigating the effects of errors, but significant uncertainties in predictions based on simulations will remain for the foreseeable future.

The analysis of uncertainties has many aspects, but they can be combined into a simple, coherent framework as shown schematically in Figure 3. This figure simply displays the main steps in the scientific method but in a probabilistic setting in order to include uncertainties. It shows a “forward step,” which goes from a hypothesized model to predictions that are compared with experiment, a “backward step” that consists of model improvement as a result of the comparison with data, and then new predictions. We note that Figure 3 has an alternative, and equivalent, interpretation in terms of the steps in Bayesian statistical inference.

The approaches to error analysis and uncertainty quantification discussed above pertain to prediction of events within or very near the regime of the data set used for validating the model. Outside this regime, uncertainties cannot be assessed, and predictions may be wrong. “Known unknowns”—that is, recognized phenomena for which adequate models are lacking—are a common source of error in simulations. Then there are “unknown unknowns.” By definition, they cannot be dealt with directly, but an attempt is made to address them through “What if?” exercises and, most important, through conservative design. In the QMU framework, that means ample margin. Certainty is still not in the cards, so our highest priority is to avoid catastrophic systemic failures, rather than failures resulting from isolated low-probability events.

## The Future of Certification

The demands on certification procedures derive from our responsibility to identify and remediate factors, including obsolescence, that could place the nation's nuclear deterrent at risk. Managing these risks will present a broad range of challenges to our certification and assessment capabilities. Some of these challenges can be dealt with confidently, by using improved predictive capabilities, but others will stress these capabilities to the point at which further nuclear testing may be needed to maintain confidence. Questions in the latter category may include certification of weapons of new design and assessment of severe weapons defects because they deal with weapons behavior well outside the tested range. How a particular question is dealt with is a matter of judgment, and QMU should be helpful in explaining the basis for confidence in such judgment.

Risk mitigation for the nuclear stockpile problems will be accomplished through (1) surveillance to monitor the actual condition of the weapons, (2) predictive assessment of the impact of changes observed in the surveillance program, especially the identification of possible failure modes (flagged in QMU by loss of margin at a gate), and (3) the ability to refurbish, remanufacture, or modify a weapon system to remedy defects (diagnosis is hollow unless followed by treatment). The last option requires a functional manufacturing capability, which is an extremely complex and expensive undertaking. Other options restrict the weapons' potential use, such as changes in the STS.

One sometimes hears that the problem of recertification can be obviated simply by "making them [warheads] the way we used to." This is appealing, but it cannot address design flaws, new designs, or the simple fact that, for all practical purposes, we

cannot make them the way we used to. Thus, the need for a more predictive scientific understanding of weapons operation cannot be side-stepped so easily.

Nor would the need for better predictive capabilities be completely obviated by a return to testing. The stockpile questions that need to be answered would inevitably outstrip the number of tests authorized or conducted to answer them, as it has occurred in the past. It is a fact of life that larger political considerations affect, and sometimes override, technical needs. Moreover, human and institutional factors will continue to profoundly influence the stockpile stewardship program. An example of such a factor is the need for an integrating goal that can focus both questions and efforts within the weapons program. Ironically, one of the critical roles of nuclear testing was to provide exactly this focus. The challenges posed by weapon assessment and certification will be met through a combination of the currently recognized steps of science, each held to higher standards of control and error analysis than is customary, and through integration of the various parts of the weapons program so that they effectively support the development of comprehensive predictive capabilities. Successful stockpile stewardship will produce tight estimates for the outcomes of critical events and will identify corrective actions where necessary. Failure, in terms of inadequacy, will be recognized as estimates that are too loose—that is, too uncertain or too unreliable—to be useful.

Predictive science applies to phenomena resulting from understood or acknowledged causes. In time, an increasing number of such causes can be studied and brought within the predictive framework with a corresponding increase in confidence in our ability to identify the factors that limit the use of our models to assess

the behavior of untested weapons. True failure could still occur in those cases in which the unrecognized cause and unanticipated effect are significant. The fundamental question of the sufficiency of our certification procedure—"How will we know if we have made a mistake?"—will never go away. Nevertheless, we believe that, with effort and determination, the nuclear weapons community can go a long way toward meeting the challenge of certification as it is presented today. ■

*For further information, contact David Sharp (505) 667-5266 (dcs@lanl.gov).*

# Weapon Certification

## *A personal view*

*Donald R. McCoy*

Historically, before the September 1992 moratorium on nuclear testing, a nuclear weapon would be placed into the stockpile only after it had undergone several hydrodynamic and nuclear tests over a period of years. Computer simulation codes were used to set weapon design parameters and to estimate both the energy generated by the weapon and the weapon's design margins. Weapons designers knew that the simulation codes were not predictive and gave the wrong answer for weapons safety and performance. To convert simulation code results into predictions of nuclear tests, they would use scaling factors based on nuclear test results. Those nuclear test results would attest to or refute the weapons designers' understanding and judgment of weapons safety and performance. Simulation codes, however, were used to certify yields of weapons placed in the stockpile when the yields were higher than the limit of 150 kilotons established by the Threshold Test Ban Treaty (TTBT). Simulation codes were also used to determine weapon design margins and uncertainties and thus ensure that weapon yields certified for nuclear testing at the Nevada Test Site (NTS) did not exceed the TTBT limit.

Modernization of weapon delivery systems required that new, robustly manufactured designs enter the stock-

pile on a regular basis. The schedule for weapon development, testing, and production was driven by the planned deployment of Department of Defense (DoD) delivery systems. Today, we certify nuclear weapons performance and safety without additional nuclear testing but with new tools and capabilities provided by the Stockpile Stewardship Program established in 1993.

### **The Past**

I joined Los Alamos under a post-doctoral appointment in the Theoretical Division in 1980 to conduct research in numerical solutions for neutron transport problems with applications to nuclear reactor design and operations. I came to Los Alamos because the Laboratory had the fastest and most capable computers in the world. I enjoyed using this capability to develop improved numerical methods and to publish several papers. I joined the weapons program in the Diagnostic Physics Group in late 1982. The group was responsible for predicting diagnostic measurements for nuclear tests and interpreting the measurements after a test. Nuclear tests were supported by a multidisciplinary team of scientists and engineers from many Laboratory groups and divisions. The weapon development and test program was planned

on a multiyear schedule and was highly visible inside and outside the Laboratory. Delaying a nuclear test would bring your name and your supervisor's name to the attention of the Laboratory director, so the team of scientists and engineers felt enormous pressure to meet the planned nuclear-test schedule. The teams supporting nuclear tests were the only organizational unit I have seen at the Laboratory that were stronger and more focused than groups in line organizations. In the early 1980s, Los Alamos and Lawrence Livermore National Laboratories and the Defense Nuclear Agency were conducting 15 nuclear tests per year at a rough cost of \$30 to \$40 million dollars for each nuclear test.

I was assigned to my first nuclear test, code-named Tortugas, in early 1983. Los Alamos named nuclear tests after towns or places in New Mexico in the early to mid 1980s and towns or places in Texas in the late 1980s and in the 1990s. As far as I know, I am the only weapons scientist that got to work on a nuclear test code-named after the county in which he was born—Bexar County, Texas. I was responsible for predicting the diagnostic measurements that would be fielded on this nuclear test. I learned how to run the simulation codes that predicted the signals measured by the diagnostics on a nuclear test. The timing involved—nanoseconds—and the

magnitudes of neutrons, gamma rays, and x-rays were very different from the ones I had seen in the nuclear reactor business. In the 1980s, most recording of nuclear test diagnostics was done with oscilloscopes. In order to provide the nanosecond time response required for the diagnostic measurements, the oscilloscopes had only limited dynamic range for recording and were set for nominal predicted current, one-half the predicted current, and twice the predicted current. The challenge in making these predictions was not limited to running simulation codes in order to predict diagnostic signals. I also had to ask myself if I could believe the results knowing, as I did, that the weapons simulation codes I used as a source term gave the wrong answers.

I remember traveling to the Nevada Test Site for the first time in 1983 to provide the predicted currents to the Physics Division experimentalists. I had learned that it takes nine months to a year to design the numerous measurements for a nuclear-test diagnostic rack (for a description of those measurements and the rack, see the article “How Archival Test Data Contribute to Certification” on page 38), set up the detectors in the rack, and set up and test the recording equipment in the aboveground trailers. The experimentalists took me aside and told me, “If your predictions are a factor-of-2 incorrect, high or low, we don’t produce useful data. If this happens, we will take you to the nearest subsidence crater and beat you up and leave you.” I suddenly realized that I wasn’t performing theoretical research in my new job.

Stringent test schedules combined with simulation codes that were not predictive forced me and everyone else in the nuclear testing program to manage risk. Because I couldn’t perform all the sensitivity calculations I thought were reasonable in time for each scheduled test, I learned to focus

on those sensitivity calculations that would yield a factor-of-2 difference in predicted detector current and worked on smaller sensitivities only if I had time. The weapons designers worked under similar constraints. Fortunately, after eight years of typically three nuclear tests a year, I was never rolled in a subsidence crater—but I did I work with some technical staff who came close. The nuclear test and weapon development programs provided the most enjoyable work experience I have had at the Laboratory by lending a strong sense of mission and value to my job.

### How Things Changed

It has been over 10 years since the inception of the stockpile stewardship program, whose mission is to develop the means to maintain confidence in the nuclear weapons stockpile without additional nuclear testing. The main driver for stockpile stewardship was to support the nonproliferation policy of the Clinton administration. At that time, the thinking was that, if the United States did not conduct nuclear tests, other countries would not test and develop nuclear weapons. It turned out that some member countries of the nuclear club continued nuclear weapon development and testing and some nonnuclear countries have since announced intentions to develop nuclear weapon capabilities using nuclear testing.

I applied for a change-of-station position and was fortunate to be accepted to work for the Nuclear Testing Division at the Department of Energy (DOE) Defense Programs Office in 1992 and 1993. As it turned out, 1992 was the last year of U.S. nuclear testing. The Congress passed a bill with the Exon/Hatfield/Mitchell amendment, and in October 1992, President George Bush signed the bill that allowed the United States to con-

tinue nuclear testing for three years under the following restrictions: There could be only five tests per year. Four would test safety improvements to existing stockpile weapons, and one would test reliability. The Congress directed the DOE and the DoD to prepare and submit to the Congress, in early 1993, a three-year plan for those last 15 nuclear tests.

The planning was an interesting exercise. The DOE, with the help of its legal staff, interpreted the law precisely as written—nuclear testing was for safety and reliability. The DoD viewed the law as less specific. It hoped to use those last tests to obtain nuclear test data that would improve the predictive capability of our simulation codes and to conduct nuclear tests at the extremes of the weapon design margins. I was a member of the technical staff that, with input from the Los Alamos and Livermore, proposed several three-year plans and presented each one for discussion and debate among an interagency group chartered by the National Security Council. We proposed so many test schedules that the joke in Washington, D.C., at that time was, “What is the nuclear test schedule of the day?” By the summer of 1993, the DOE and DoD had not settled on a definition of the last 15 nuclear tests, and the Clinton administration extended the nuclear test moratorium indefinitely.

I returned to the Laboratory in late 1993 and continued to participate in the formulation of the U.S. nuclear testing policy. Once the moratorium was extended, the administration pursued the goal of a worldwide comprehensive test ban treaty. The debate shifted to defining exactly what is allowed under a nuclear test ban and what is verifiable under a test ban treaty. An interagency group debated whether hydronuclear experiments—defined by the National Security Council staff as the generation of less than 4 pounds of generated nuclear

energy—should be allowed, or whether limited nuclear testing—less than 1 kiloton of energy generated—should be allowed because at that time that yield was thought to be the verification threshold.

The question of limited nuclear testing was addressed by DOE and DoD representatives at a nuclear weapons symposium held in Omaha, Nebraska, in June 1995. The symposium was hosted by Admiral Henry Chiles, then commander in chief of the U.S. Strategic Command. Energy Secretary O’Leary and the nuclear weapons laboratory directors attended the symposium. The laboratories presented the technical benefits of limited testing, and representatives of the nuclear weapons complex plants discussed the required physical-plant infrastructure and capabilities to maintain and refurbish nuclear weapons. The DoD proposed a strategy that implemented the new simulation and experimental capabilities of stockpile stewardship while allowing up to 10 years of limited (less than 1 kiloton of generated energy) nuclear testing to validate the stockpile stewardship capabilities. The DOE proposed a strategy of continued implementation of stockpile stewardship capabilities without conducting limited nuclear testing or hydronuclear experiments.

The outcome of the symposium was a statement from the laboratory directors to the secretary of energy that limited nuclear testing was not needed at that time. The DoD agreed to the DOE’s strategy, with the additional safeguard that the laboratory directors would provide the secretaries of energy and defense with an annual assessment of the stockpile and of the need for nuclear testing. The administration announced in August 1995 that the United States would pursue a zero-yield nuclear test ban treaty. At this point, stockpile stewardship was the only option

under U.S. policy to maintain confidence in the performance and safety of the nuclear weapons stockpile.

### The Present

Using stockpile stewardship tools and capabilities in place of nuclear tests requires a greater predictive simulation capability than was available in the past to certify changes or modifications to the nuclear weapons stockpile and to develop modified weapon designs. A main goal of the present nuclear weapons program is to manage risk across an aging nuclear weapons stockpile by making informed decisions about nuclear weapon design margins and uncertainties and about ways in which those margins and uncertainties change over time. This goal can be accomplished by continuous surveillance, that is, sampling of weapon components in the nuclear weapons stockpile, evaluating and assessing the condition of those weapon components, and establishing the lifetimes of those components. Decisions to replace weapon components, which sometimes cannot be manufactured exactly like the original components, must be based on the best technical assessment and evaluation of the current weapon design margins and uncertainties and on ways to improve the weapon design margins and reduce their associated uncertainties. Los Alamos is currently restoring the nation’s capability to manufacture pits. The question to be answered is, “Will a pit manufactured at Los Alamos produce the required nuclear weapons performance that pits manufactured at Rocky Flats used to produce?” In the past, nuclear testing verified that yield. At present, testing is not available. However, I believe the answer to the previous question is “yes,” but proving this assertion in the absence of nuclear testing is a dif-

ficult technical challenge. An underlying concern that has always been an issue with stockpile stewardship is that certifying nuclear weapons without nuclear testing will not address “unknown” issues that could arise in the nuclear explosion phase of a nuclear weapon’s operation. Since this nuclear explosion phase of operation is not accessible without a nuclear test, technical judgment by weapons scientists and engineers will underpin our statements concerning weapon certification.

Quantitative understanding of design margins and uncertainties requires the development of new simulation tools and capabilities because the nuclear explosion phase of a nuclear weapon can no longer be accessed directly. We started the development of these simulation tools and capabilities with the creation of the Accelerated Strategic Computing Initiative (ASCI) program in 1996. We currently have new codes that can run nuclear weapon implosion and explosion problems in three dimensions on terascale computing platforms that could never be simulated in the past. The most difficult elements, verification and validation of these simulation tools and capabilities, are under way.

**Certification.** The most challenging technical problem to solve for stockpile stewardship is proving or certifying that the results of these simulation tools are believable and can be used to produce meaningful design margins and uncertainties for our aging nuclear weapons or for replacement components that cannot be manufactured exactly like the original ones. That is why we are developing new models that can more accurately capture the details of physical and material behavior. In turn, the new models have led to new experiments in the radiation flow and static and dynamic behavior of materials. The

results of these experiments will provide the empirical basis for the models and a means to validate the models.

For the Los Alamos–manufactured pit, we have developed a certification strategy that requires development of weapon simulation baseline models that should match past nuclear test data for both the implosion and explosion phases of experiments (see the article “How Archival Test Data Contribute to Certification” on page 38). These models are built with legacy and Advanced Simulation and Computing (the new ASCI program name) tools and will be used to predict results of subcritical implosion-phase experiments conducted on assemblies whose geometries mimic those of nuclear weapons. These subcritical experiments are planned for the next few years (see the article “The New World of the Nevada Test Site” on page 68).

We will ask such questions as “Which baseline simulation model gives the best prediction for the experiments?” and “Which simulation model gives the smallest uncertainties in the nuclear weapon design margins?” Of course, the hard part in answering these questions is relating implosion experiments to the explosion phase of a nuclear weapon when the only tool available is a simulation model. This certification strategy will be one of the first big tests of the success of the stockpile stewardship program at Los Alamos.

Another challenge in stockpile stewardship has developed over the past 10 years; it was not anticipated when the program started. Originally, there was concern that entry-level weapons scientists and engineers without nuclear weapon development and testing experience would have too much confidence in these new stockpile simulation tools and capabilities. Actually, this concern has not yet materialized. The general view of

entry-level staff has held that the current legacy and advanced simulation tools do not have the right, physically based modeling capabilities or are not sufficiently validated. Therefore, when a stockpile issue arises and an assessment and evaluation of the issue are undertaken, no timely resolution of the issue occurs because of the need for a perfect simulation tool with which to address the problem. The technical judgment developed in the old weapon development and testing program, which included management of risks when weapons were certified and tested, does not have an equivalent development of technical judgment in the Stockpile Stewardship Program. That is why managers in the current nuclear weapons program must assume more risk concerning certification than managers had to assume in the past. I believe this situation is a key management challenge to stockpile stewardship.

The new approach to certifying a nuclear weapon is very different from that used in the past and requires new tools and capabilities that are currently under development. Over the next several years, many parts of the nuclear weapons stockpile will be refurbished. That process will require certification of both replacement components manufactured by new methods and modified components. We may also have to certify modified weapon designs. The test for stockpile stewardship will occur during this period, and I am waiting for the answer to the question, “Is stockpile stewardship succeeding or failing?” ■

*For further information, contact  
Donald McCoy(505) 667-4224  
(dmccoy@lanl.gov).*





# The Pit Production Story

*Douglas D. Kautz, David B. Mann, Richard G. Castro,  
Lawrence E. Lucero, and Steven M. Dinehart*

In 1993, Los Alamos National Laboratory was asked to return to an important part of its roots. It would once again tackle the manufacturing of the plutonium-bearing pit, a major component of nuclear weapons. The Manhattan Project pioneers had learned to work with plutonium, perhaps the strangest and most reactive of all the elements on Earth, and they had built the first pit, testing it in the Trinity experiment. After World War II, Los Alamos continued pit manufacturing until 1952, when that mission was transferred to the newly completed Rocky Flats Plant in Golden, Colorado—see Figure 1.

Rocky Flats produced thousands of pits year after year until 1989, when the Department of Energy (DOE) abruptly ended the manufacture of plutonium components because of environmental concerns. Two years later, DOE changed the plant's official mission from defense programs to environmental remediation and began the search for an interim location where pit manufacturing could be continued on a small scale. Although the United States would eventually have to build a new pit-manufacturing facility to replace the Rocky Flats Plant, the projected time for its completion was 2017. Consequently, when Rocky Flats began environmental remediation in 1991, DOE asked Los Alamos to take on the mission of pit surveillance. And in 1993, the Laboratory was asked to take on pit manufacturing since our facility at Technical Area (TA) 55 was the only fully functional plutonium facility in the DOE–Defense Programs complex—see Figure 1.

At the time of transfer, the Laboratory could nominally perform almost all plutonium-processing steps needed to manufacture most pits in the enduring stockpile; however, close inspection revealed a host of issues to be solved. Some machining and welding equipment, as well as dimensional inspection capabilities, were absent; several processes needed improvement to meet the quality requirements for the manufacture of war reserve components; and several processing methods had to be converted to fit existing equipment or to meet new regulatory mandates that disallowed the use of Rocky Flats technologies. In addition, several pieces of equipment,

although functional, required replacement or additional backup capability so that the mission could be viable at Los Alamos. Substantial upgrades were also needed in the processing capability for nonnuclear components. Despite many challenges, the pit-manufacturing effort began but as a very small project. Many of the participants performed several functions, and they focused on developing processes rather than product. As the project matured and needs were better understood, the emphasis shifted toward manufacturing pits that would be certifiable, that is, that would meet all the specifications required for inclusion in the enduring stockpile. The successes described here reflect the dedication of a large number of people in many organizations across the Laboratory. The largest contributors were the Nuclear Materials Technology, Engineering Sciences and Applications, Materials Science and Technology, and Chemistry (formerly Chemical Science and Technology) Divisions.

### Early Decisions on Materials and Processes

Early in the project, we made several major decisions that would influence the entire manufacturing effort—from preparing the plutonium metal to fabricating the components and assembling the pits. First, we would reduce the use of various process chemicals to meet environmental and waste-processing concerns. Next, we would develop new welding processes for various joints, and finally, we would develop methodologies to ensure that highly reactive plutonium did not exhibit undue corrosion upon assembly into the pit. An overriding factor in all our decisions was, and remains, the necessity to produce pits that are equivalent to those manufactured at Rocky Flats.

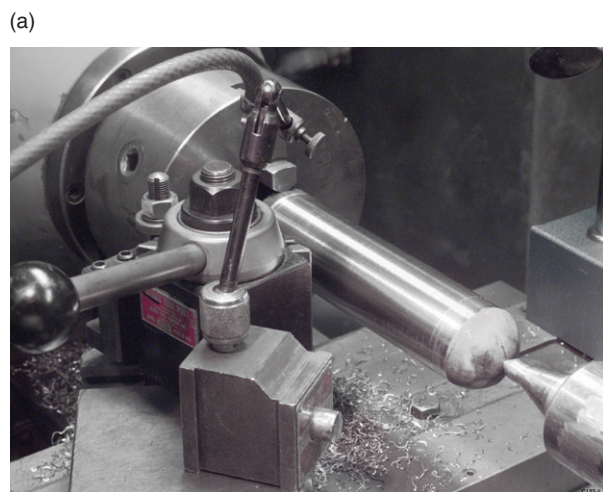


**Figure 1. The Rocky Flats Plant and TA-55**

These pictures are aerial views of the Rocky Flats Plant (top) and TA-55 at Los Alamos (bottom).

The choice of solvents posed a particularly thorny problem. At Rocky Flats, carbon tetrachloride and 1,1,1-trichloroethane had been used in large quantities, but their use at Los Alamos was prohibited by

modern environmental and waste-processing constraints. To develop processing strategies that employ different solvents and minimize the amount needed, we launched several compatibility studies with plutonium



**Figure 2. The Dry Machining Process**

(a) The new dry machining process avoids heavy use of lubricants, which are expensive and difficult to dispose of when becoming waste. In this process, no chemical changes take place on the surface of the plutonium parts, and thus all the plutonium shavings can be collected for reuse. The dry machining process, therefore, generates no plutonium waste. (b) Star machinist Dean Martinez is programming the T-base lathe to machine a component.

and other materials. We not only developed those strategies but also continued to reduce the waste stream by purifying and reusing the solvents while ensuring the cleanliness of components assembled into pits.

The heavy use of lubricants at Rocky Flats posed two additional problems that we had to avoid. First, the lubricants themselves generated hazardous waste streams. Second, because plutonium is highly reactive, each time a lubricant is used, a cleaning step involving large quantities of solvent must follow to ensure that the plutonium does not reduce to plutonium hydride, in which case it must be scrapped and reprocessed. Our solution was to reduce the use of lubricants through development of a “dry” machining process. Unlike traditional processes, dry machining requires lubricant only during the finishing of parts—refer to Figure 2. We also reduced the lubricants in other operations that had used them during Rocky Flats processing.

Creating the dry machining process took approximately 18 months and involved development of new tools, procedures, machining parameters, and airtight gloveboxes (Figure 3). We altered materials processing. At Rocky Flats, wrought processing techniques had produced the plutonium. But installation of the equipment for that process at the Los Alamos facility would have forced major facility changes with consequent lengthy delays in acquiring a revised facility operating permit. Instead, both our pit-manufacturing and certification staff compared the properties achieved through wrought processing and casting and concluded that cast material could indeed meet the needs of the weapons community—refer to Figure 3.

To study all replacements for process chemicals and materials used in pit manufacture, as well as their effects on the materials used in pits, the War Reserve Materials Compatibility Board was convened. Once the board found that the new

materials were compatible with the pit materials, an extensive quality-control program was instituted to ensure that no changes occurred in the formulation or processing of those materials. That same quality-control program daily ensures that all materials in the manufacture of pit components meet the established standards of uniformity and high quality (Figure 4).

A simple example of quality control concerned rubber bands that hold a marking mask on the pit. The rubber bands for our first pits were made from pure rubber and left no residues on the pit. A second batch received from the vendor contained an extra ingredient that would have left an unacceptable residue on the pit. The materials compatibility board studied the new material and rejected its usage for the war reserve product.

This description gives only a glimpse of the many decisions about materials, processing, and quality control. Both the production staff and those in advisory roles worked together long and hard to develop and

approve changes in numerous processes involved in pit manufacture, paving the way for us to meet our production milestones.

### The Road to Quality Pits

Early on, the production staff and the Los Alamos weapons design groups decided that the major changes in processing should be tested by manufacturing a series of development pits and then checking whether the processing changes reduced their functionality. We manufactured nine pits for this purpose.

With the first pit, we tested an important welding process imported from Rocky Flats. That test was a success. For the second pit, we used processes and tracking systems that were available at that early stage and achieved only mixed success, taking three tries to complete fabrication satisfactorily. The finished pit was subjected to an environmental test, and the results were compared with those from pits manufactured at Rocky Flats.

We used the third pit to compare the surface reactivity of plutonium fabricated at the Laboratory with that of pits made at Rocky Flats. During manufacture of the fourth pit, we tested the effectiveness of new cleaning materials chosen to meet new waste-generation regulations. We exposed the plutonium to larger-than-normal quantities of various processing materials known to be difficult to remove and then showed that the new cleaning material, trichloroethylene, could successfully remove the materials. Postfabrication testing indicated no significant differences from pits that had undergone conventional processing.

Then, in 2000, two major problems caused significant delays in our fabrication schedule. First, during the Cerro Grande fire in May, the TA-55

(a)



(b)



**Figure 3. Cast Plutonium**

(a) Shown here is an induction furnace used in the casting process. Induction heating provides good stirring of molten plutonium and a clean atmosphere for processing this highly reactive metal. (b) Casting technician Anthony Valdez is setting up the crucible used in casting.

plutonium facility was entombed for the first time since its opening in December 1978 for fireproofing upgrades. After the fire, all pressurized gas and fluid lines were tested for leaks in response to a corrective action from a contamination incident that had occurred earlier in the year. After resolving those issues, we returned to manufacture a pit that tested the effects of glovebox atmospheres on the plutonium material and on the processing used to remove any reacted material from the plutonium before final fabrication.

Although by that time significant work had been done to qualify materials and processes, our development pits still deviated from any pit that would ever be allowed to enter the enduring stockpile. We decided to produce a series of “standard pits” that were as close as possible to war reserve specifications and processing. We made the first in this series mainly to exercise the newly formulated systems for tracking data and parts. Although several difficulties were noted during the processing of this pit, the lessons learned helped

the project mature greatly. The next in the development-pit series was purposely manufactured with several defects to test the capabilities of our major nondestructive testing processes. We then manufactured a second standard pit, which tracked much better than the first, but the experience showed that we still had several challenges to overcome before we could successfully meet all required product specifications and quality standards.

The next development-pit test assessed the integrity of Los Alamos components by directly comparing each one with the corresponding Rocky Flats component. After the pit was successfully fabricated, it was tested to ensure that no reactivity differences could be discerned between the different materials. Finally, we built a third standard pit to check our quality control and assurance systems and to demonstrate the efficacy of the last remaining nondestructive testing process.

A long hiatus in fabrication then ensued as the project went through major restructuring. To bring opera-

tions back up, we built two standard pits whose status we compared with that achieved for the last fabricated pit. Although the quality control and assurance systems had matured greatly during the restructuring activities, the long period of inactivity had, as expected, some negative effects on both the equipment and the process operators. Even after the second pit was fabricated, not all the processing

problems had been fully resolved.

On the other hand, during fabrication of the second pit, we successfully instituted a major, new inspection process. That process, although very difficult to install and prove out, is necessary for certifying the quality of fabricated components. Its implementation and that of another inspection process allowed us to finally fabricate a product that was fully compliant with the product specifications. All our work on honing and documenting our processes came together when we manufactured the next standard pit. It became the precursor to a major milestone for the project, namely, a pit produced with fully qualified processes and quality systems as specified in the DOE QC1 quality control policy. The final development pit was a calibration unit fabricated with several documented defects; it will be stored and periodically tested to ensure that nondestructive evaluation processes are performing in the manner expected.

### Qualification Pits

The next major milestone was the delivery of a certifiable pit—a pit that



**Figure 4. Quality Assurance Audit**  
Los Alamos and National Nuclear Security Administration quality specialists verify that quality assurance systems support pit manufacturing.

met all the manufacturing specifications required for placement in the stockpile. Although the pits fabricated at Los Alamos must still undergo several engineering and physics tests before they can be fully certified and actually placed in the stockpile, all quality control and assurance systems and process qualifications associated with manufacturing will be in place.

In May 2003, the Laboratory completed the first nuclear weapons pit that meets specifications for use in the U.S. stockpile. The newly made pit, called QUAL 1 because it was built with fully qualified processes, is for use in the W88 warhead, which is carried on the Trident II D5 Submarine-Launched Ballistic Missile, a cornerstone of the U.S. nuclear deterrent.

The pit production project is restoring our nation's ability to make nuclear weapons, a capability that had been lost when the Rocky Flats Plant was shut down in 1998.

### Looking Forward

Members of the project are already working on ways to improve both the yield and the efficiency of processing.

We plan to reduce the waste generated during plutonium casting by replacing single-use fabrication tooling with reusable tooling. We are studying ways to remove a common machining defect encountered during turning operations. We are instituting real-time monitoring on several pieces of equipment so that process holds now encountered while waiting for batch results can be minimized. We are also embarking on an in-process monitoring strategy to gain much needed process-performance data. Having moved beyond the certifiable-pit milestone, we will institute this type of robustness initiative to provide a more consistent and higher-quality product to our DOE and Department of Defense customers. ■

**Douglas Kautz** graduated from the Colorado School of Mines in 1982 with a bachelor's degree in metallurgical engineering and received a master's degree in metallurgical engineering from the Colorado School of Mines in 1987.

Doug worked for Rockwell International from 1982 to 1987, specializing in the process engineering of materials used in nuclear weapons. From there, he moved to Lawrence



Livermore National Laboratory, working on high-energy-density welding research and materials issues with nuclear weapon components until 1994, when he became a staff member in the Nuclear Materials Technology Division at Los Alamos, where he is now a deputy leader of the Weapon Component Technology Group.

# Strategy for Small-Lot Manufacturing

## *In-process monitoring and control*

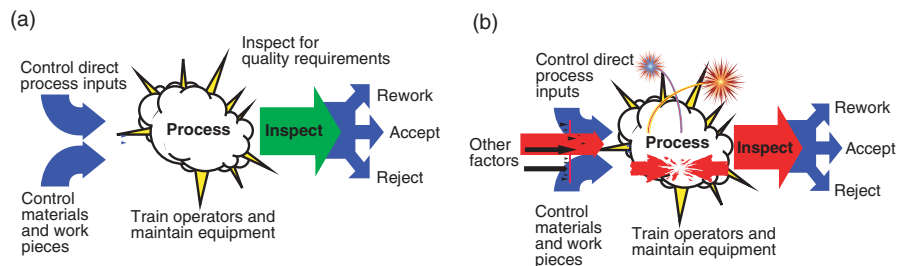
Vivek R. Davé, Daniel A. Hartman, William H. King, Mark J. Cola, and Rajendra U. Vaidya

During the Cold War, the nuclear weapons complex produced thousands of components each year to support the stockpile. The manufacturing process stream—a unique combination of equipment, people, “qualified” processes, and plant idiosyncrasies—had high production capacity but not always high yields. Figure 1(a) shows the elements that define a qualified manufacturing process: Inputs are controlled, all procedures are followed, and the end product is found to be satisfactory through statistical sampling involving inspection and destructive testing. Products made by use of qualified processes were then “certified” as capable of entering the active stockpile, if they met military characteristics (demanding in-service requirements) and stockpile-to-target requirements (that is, they would operate as expected from the time they were taken out of the stockpile to the time they would reach their target) when tested. However, the manufacturing process was treated as a series of black boxes whose internal process dynamics were poorly understood and not monitored. Nevertheless, this method of process qualification and product certification served the nation well for four decades.

The current Los Alamos approach to pit manufacturing follows this old paradigm. It tries to recreate as closely as possible the original manufacturing stream used at Rocky Flats but on a smaller scale because the production volume is much lower. Here, we explain why this

### Challenges Faced by Los Alamos Manufacturing Processes

- Overall mission scope has shifted after the demise of the Soviet Union.
- Manufacturing operations suffered a long period of inactivity.
- Tremendous upheaval was felt in transferring operations to new sites.
- More than 90 percent of the key personnel have changed.
- Many remaining process experts are retiring.
- Significant changes in equipment, processes, and process flow have been implemented.
- Plant features and layout have been significantly changed.
- Prior continuous operations were fragmented.
- Quality requirements are the same as in the past.



**Figure 1. A Qualified Process and What Can Go Wrong**

(a) Elements of a qualified process are shown. (b) Qualified processes can be adversely affected (red areas) by hidden factors such as human error caused by insufficient process knowledge; inadequate procedures or incomplete documentation; material variations resulting from changes in processing history, minor element-composition differences, and changes in surface condition and oxidation state; dimensional deviations or residual stresses in work pieces; equipment and tooling degradation; inadequate maintenance or calibration; tooling wear and fixture distress; and marginally stable parameters of process qualification.

approach is problematic for small production volumes and outline, through a real example, a modern approach to quality manufacturing by process monitoring and control in real time.

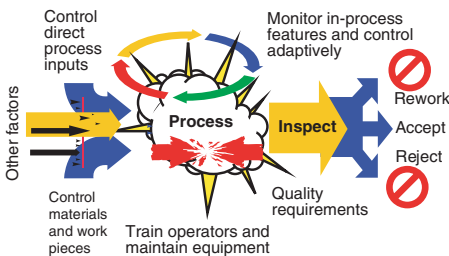
### The Problem: Small-Lot Manufacturing

Figure 1(b) outlines problems that can arise when one tries to develop qualified processes with lots that are

down, say, from thousands per year to tens per year. Key interaction terms and intermittent or sporadic process dynamics may be missed entirely. The resultant processes, which are supposedly qualified, could manifest spurious process dynamics in seemingly unpredictable patterns over time. Such processes may therefore be incapable of holding the product in a state of statistical process control in the absence of further process understanding. Manufacturing at Los Alamos has already reached this position.

### The Solution: In-Process Monitoring and Control

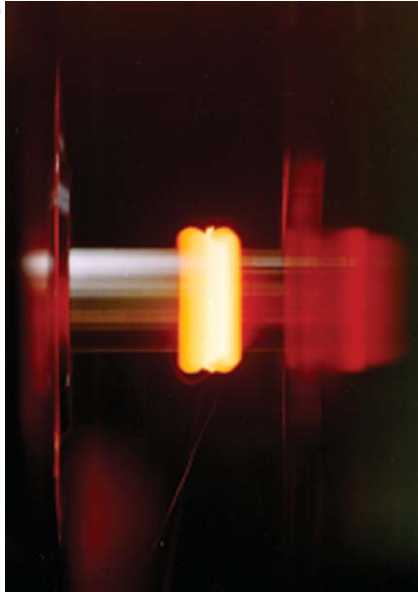
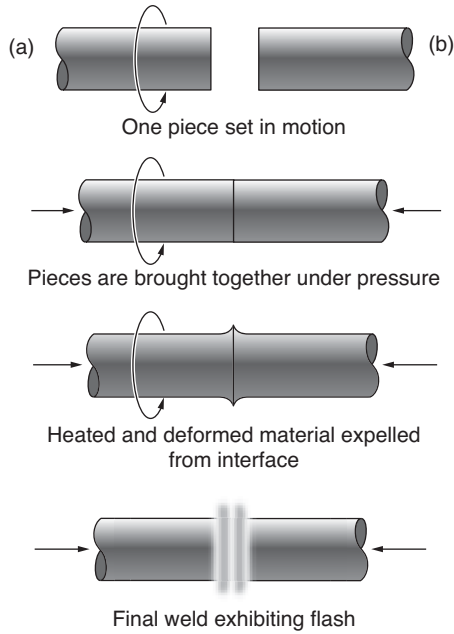
In the new approach, we shift our frame of reference from an outside view, in which the operations are items on a work instruction sheet, to an inside view, in which physical processes are interrogated and controlled as they happen. In-process data



**Figure 2. In-Process Monitoring and Control**

are interpreted through pattern recognition and classification algorithms that are trained not only to identify processing faults but also to classify the root causes of those faults. The shift is somewhat analogous to going from alchemy to chemistry, from a black art to a predictive science based on underlying physical principles.

Various steps are required to create in-process assurance of part quality.



**Figure 3. Steps in Inertia Welding**  
**(a) The steps of the welding process are listed, and a photo of the final welding step is shown in (b).**

First, we identify critical in-process physical behaviors determining part quality and the means to measure them. We then find out how those behaviors are correlated to specific attributes that constitute quality.

Typically, in-process raw data cannot be directly correlated to specific faults in part quality or process integrity. We must therefore employ data reduction methods to find those key signatures that might identify the presence of specific faults. We then use those signatures to develop learning algorithms that not only identify the faults but also classify their causes. We train the algorithms during process development by intentionally creating fault conditions. We then establish an operating window, or range of values of allowed in-process behavior. Finally, we deploy an in-process control system based on this operating window. The results of this methodology can be spectacular: In the F-22 Advanced Tactical Fighter Program, there are engine components that have never been inspected after having been man-

ufactured and are flying as built. The elements of in-process quality assurance shown in Figure 2 can be compared with those of traditional process qualification shown in Figure 1(b).

### Practical Example: A Welding Problem

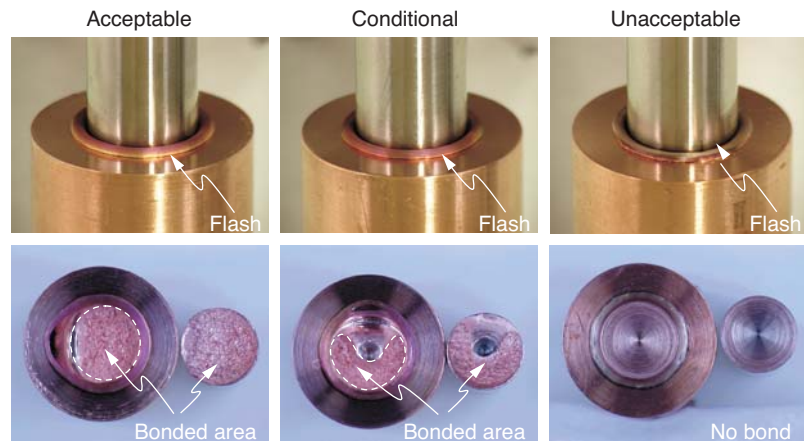
Inertia welding, or the solid-state friction welding of two parts, is a process used on certain defense and aerospace components (see Figure 3). Several defects (hidden factors) are of concern in inertia welding: insufficient or excessive speed or pressure resulting in inadequate joint strength, angular offset in grips or at bond plane resulting in variations in residual stress, and machining defects, handling damage, or contamination at the bond plane.

Here, we describe our in-process approach for detection of bond plane contamination in inertia welds made between copper and stainless steel. Contamination is the most difficult

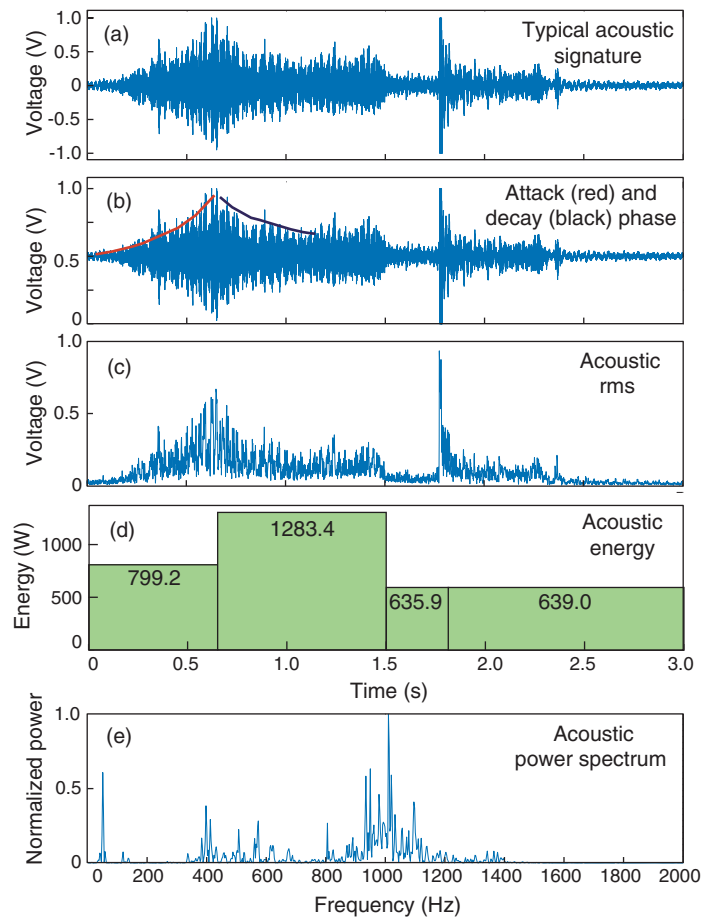
problem to diagnose because thermo-mechanical material flow during the weld expels the original interfacial material. Nevertheless, as shown in Figure 4, even minor amounts of contamination can have dramatic effects on the bond. The three very different weld qualities shown were produced with identical process parameters (knob settings on the welding machine). Thus, the only means to detect conditional or unacceptable welds without destructive testing is an in-process sensing approach. Other forms of nondestructive evaluation have proved to be inconclusive.

To detect bond-plane contamination, we collect acoustic and vibrational signals emitted during the welding process—see Figure 5(a). Because those signals are not useful in their raw form, we have applied various data-reduction procedures to extract key features, as illustrated in Figures 5(b) to 5(e). First shown is the so-called attack and decay descriptor, an analytical tool typically used in speech recognition, describing attack phases, or regions of increasing sound intensity, and decay phases, or regions of decreasing sound intensity. Next is the root-mean-square (rms) intensity of the acoustic signal. Third is the total acoustic energy for different portions of the signal, or simply the total accumulated acoustic counts for given portions of the signal. Finally, the frequency content of the signal is shown. It is obtained from the Fourier transform of the time-domain data. The resulting acoustic power spectrum showing the relative signal intensities at various frequencies will, when linearly superimposed, reconstitute the time-based signal.

In the next phase of in-process data analysis, we want to use the key features to make inferences about weld quality. We need an analytical method that assigns a unique set of in-process “signatures” to a good weld and at the



**Figure 4. Acceptable, Conditional, and Unacceptable Welds**  
 An acceptable bond (90% to 100% of the area is bonded) requires that surfaces be machined, cleaned, and immediately welded. A conditional bond (50% to 90% of the area is bonded) results if trace amounts of contamination accumulate on the surfaces, and an unacceptable weld (less than 50% of the area is bonded) results if the surfaces are not well cleaned and therefore residual organic contaminants are left at the interface.



same time enables us to classify the probable root cause for a bad weld.

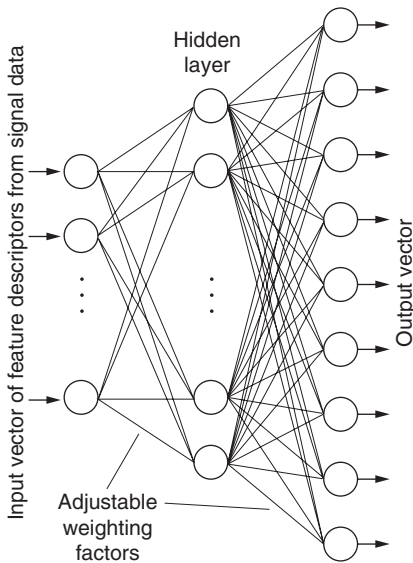
**Figure 5. Typical Acoustic Signature and Reduced Signatures**



**Table I. Ability of Neural Network to Identify Weld Quality**

Feature Used	Identification of Weld Quality (Accuracy, %)		
	A or U	A or C	A, C, or U
Attack and decay	85	63	54
Acoustic rms	74	47	50
Acoustic power spectrum	100	100	100
Acoustic energy	100	32	50
<b>Number of Training Instances</b>	19	18	25

A = acceptable bond, C = conditional bond, and U = unacceptable bond or = exclusive “or” (xor)



**Figure 6. The Neural Network for the Welding Process**

Many statistical and nonstatistical approaches are possible. In this work, we have used an artificial-intelligence algorithm known as an artificial neural network. A neural network is a collection of heuristic computational architectures and algorithms that are biologically inspired, nonlinear, massively parallel, distributed, composed of simple computational elements, and ideal for pattern identification and classification in complex data sets. These architectures and algorithms acquire process knowledge through a learning process that uses data sets. The

knowledge is stored in a neural network through interneuron connection strengths or synaptic weights.

We train our neural network (see Figure 6) using feature data representing good and bad welds, and the network stores this knowledge implicitly in its weighting factors. The network can infer the quality of the bond and thus distinguish among acceptable, conditional (or marginal), and unacceptable welds, and it performs root cause analysis of faults, deducing the cause of the faults. Both capabilities are very important for small lots that require precision welding. Table I summarizes the neural network’s ability to identify the quality of the bond from the various feature descriptors in Figure 5.

In this example, we did not take the additional step of constructing a process window because the objective was to detect and diagnose the occurrence of a very problematic defect, namely, bond-plane contamination.

In carrying out this work, we made a conscious tradeoff between less expensive sensors coupled with sophisticated data analysis versus expensive but more capable sensors. We used an array of low-cost thin-film piezopolymer sensors and miniature acoustic-emission sensors that cost less than a dollar a piece in place of a quartz sensor costing \$10,000.

## On the Verge of a New Quality-Control Revolution

By emphasizing in-process dynamics and control, we can significantly increase our ability to characterize and control manufacturing processes of small precision lots. The potent combination of inexpensive and virtually limitless computational power, inexpensive sensors, and algorithms capable of dealing with large, complex data sets has set the stage for a revolution in our approach to manufacturing quality. This effort has sufficient intellectual scope to be a “grand challenge” for Los Alamos, in particular, and the weapons complex of the National Nuclear Security Administration, in general. If successfully implemented, this new approach could shift the present conformance and inspection mindset to a predictive approach that would emphasize fundamental process understanding. The new approach would bring tangible and quantifiable benefits to Los Alamos manufacturing. Here is a list of the most significant ones: process characterization based on what the part experienced and not just on knob settings on a machine tool that may be obsolete within a decade, manufacturing recipes that can be easily moved from one machine tool to another, reduced scrap and rework, automated analysis of root causes, targeted process improvements, reduced cycle time to bring new processes online for new products, and less work to qualify a new piece of equipment or process.

For the past 80 years, we have used final inspection to verify product conformance to specifications and statistics and to quantify the consistency of the process in meeting those specifications. In-process dynamics, therefore, represents the first major new concept in quality control in almost a century. Its impact may well be as far-reaching in this century as statistical process control has been in the last century. ■

## Historical Timeline Leading to In-Process Dynamics Approach

<b>B. C. to 1700s</b>	Manufacturing is dominated by the artisan and the guild structure.
<b>1750s to 1850s</b>	The Industrial Revolution reaches both the Old and New Worlds.
<b>Late 1700s</b>	First use of interchangeable parts in manufacturing assemblies.
<b>1880s to 1900</b>	Thomas Edison literally electrifies America, having profound influence on industry.
<b>Early 1900s</b>	Both the National Bureau of Standards in the United States and the British Institute of Standards in England are founded—standards drive improved inspection methods.
<b>1900s to 1920s</b>	Ford establishes the production line, the manufacturing paradigm for the next 100 years.
<b>1920s to 1930s</b>	Alfred P. Sloan at General Motors formulates the management structure for the twentieth-century manufacturing firm.
<b>1920s</b>	Walter A. Shewhart invents statistics for process control.
<b>WW II</b>	The U.S. military-industrial complex helps win the war by using mass production methods, together with inspection to ensure conformation to specifications.
<b>WW II</b>	Stan Ulam, John von Neumann, Nicholas Metropolis, and others at Los Alamos form the basis for modern digital computers as well as a scientific computation.
<b>1947</b>	Bell Laboratories invents the transistor.
<b>1950s to 1980s</b>	W. Edwards Deming promulgates the modern approach to statistical process control. The Japanese eagerly adopt it and experience a manufacturing revolution.
<b>1958</b>	Texas Instruments invents the integrated circuit.
<b>1974</b>	Intel launches the 8080, the first successful commercial microprocessor.
<b>1960s to 1990s</b>	Development of heuristic algorithms, John Holland's genetic algorithms, Lotfi Zadeh's fuzzy logic, the Hopfield model of neural networks, and data mining and complexity sciences.
<b>1960s to 1970s</b>	ARPANET, the ancestor of the Internet, is developed.
<b>1970s to 1980s</b>	The first personal computers arrive on the market.
<b>1980s to 1990s</b>	Cheap computing and sensors following Moore's Law.
<b>1990s to 2000</b>	Six Sigma is widely implemented, representing the culmination of 80 years of statistical measurement and control of conformance to specification.
<b>1990s to 2000</b>	First implementations of the in-process approach (for example, military engine parts).
<b>21st century</b>	A new revolution in quality control is made possible by the existence of key ingredients: cheap computing and sensors, as well as advanced data processing algorithms for large and complex data sets that represent in-process behavior.

*For further information, contact  
Vivek Dave (505) 663-5625  
(vivek@lanl.gov).*

# The New World of the Nevada Test Site

*Ghazar R. Papazian, Robert E. Reinovsky, and Jerry N. Beatty*

The Nevada Test Site (NTS) has been an integral part of many Los Alamos National Laboratory programs for more than 50 years. In 1951, Los Alamos conducted the first nuclear event at the NTS. It was the atmospheric shot Able, an airdrop of 1 kiloton in Area 5 (referred to as “Frenchman Flat”). Many weapon tests, both atmospheric and underground, were conducted until the 1992 moratorium on nuclear testing. The Divider event carried out by Los Alamos in September 1992 was the last underground nuclear test. The moratorium challenged us, the Laboratory staff, with the task of maintaining the capability to return to nuclear testing, should that be necessary, and of identifying and nurturing a niche where we could be relevant to the emerging stewardship mission while maintaining close ties with our proud past.

The story of the present underground complex in Area 1 (refer to Figure 1) of the NTS started in the late 1960s, when U1a, the first shaft<sup>1</sup> in that area, was mined. The idea of the shaft was part of Bill Ogle’s larger vision for Area 1 (Ogle was leader of the Test Division at Los Alamos between 1965 and 1972). Very much in tune with “thinking big,” which was characteristic of the time, Ogle had envisioned having not one but two shafts mined at Area 1. The two would be connected with a drift,<sup>2</sup> a

line-of-sight pipe, and a series of fast closures in which to conduct nuclear testing. According to Ogle’s idea, a nuclear explosive would have been contained in one shaft. Upon detonation, the explosive would have delivered an electromagnetic pulse to both a missile and its warhead located in the other shaft. Called the Flashlight Program, this idea, however, was never implemented. Only later, in 1986, was a 458-foot drift mined south from the shaft of the 1960s vintage in preparation of the Ledoux nuclear test of 1990.

When Jay Norman became leader of the Test Division at Los Alamos (1988), he started a long-term investment strategy for the development of a low-yield nuclear experiment research (LYNER) facility. The vintage shaft U1a became its site. In 1992, the LYNER facility was ready for its new role in support of the stewardship mission. Shortly thereafter, a new shaft, U1g (1100 feet north and 50 feet east of U1a), was mined and connected to U1a by a series of drifts and alcoves housing experiments and equipment for the stewardship mission. The purpose of the new shaft was to allow adding infrastructure into the LYNER facility. It became possible to install diagnostic cables to surface-located recording trailers, provide power to the underground complex, and most important, provide a second emergency egress to the surface through a pipe with a diameter of 48 inches (much like the shaft sunk during the Pennsylvania coal mine accident of 2002).

Because of increased experimental activity in the underground complex at Area 1, yet another shaft, referred to as U1h, was mined and connected

<sup>1</sup>A shaft is a deep excavation used for mining, conducting experiments, lowering men and materials, or ventilating underground workings. Shafts are typically vertical or nearly vertical.

<sup>2</sup>A drift is a long alcove that has a plug behind which multiple experiments can be conducted in drilled holes.

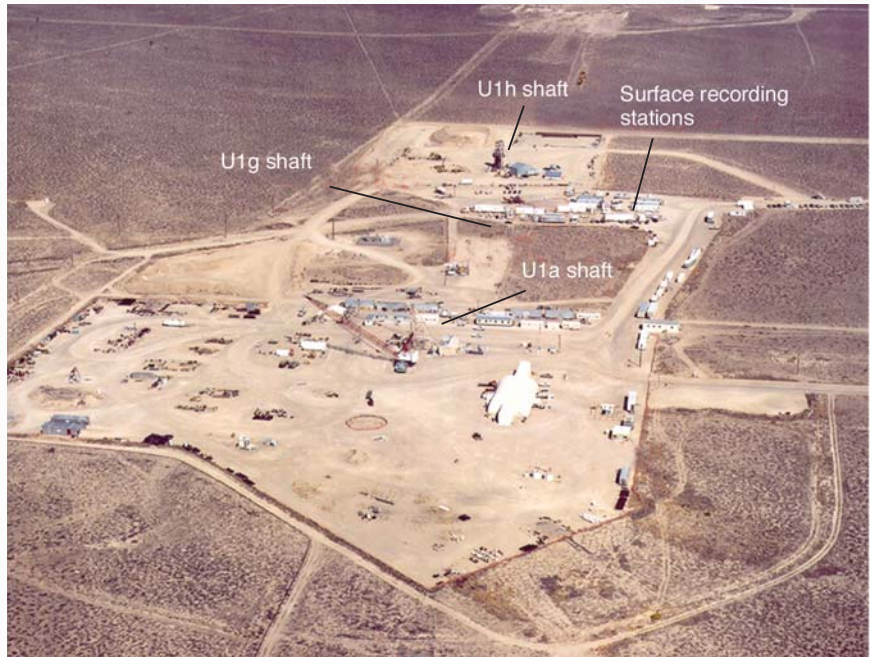
to the complex network of drifts. The U1h shaft was commissioned in 2001 and is located 1490 feet from U1a. Its primary purpose is to ensure worker safety because it provides additional egress from the complex during an emergency. A special lift basket is available to expedite rapid removal of underground workers during an emergency. Both U1g and U1h are within a few hundred feet of the experimental alcoves.

In over a decade since the moratorium on underground nuclear testing, the nature of testing at the NTS has changed considerably. To maintain the existing infrastructure in case of a return to nuclear testing and obtain data for the stewardship mission, we conducted high-consequence subcritical experiments. We then used results from those experiments to test models for computer simulations. In a subcritical experiment, high explosives (HEs) and special nuclear materials are used, but the experiment never achieves criticality, or a self-sustaining chain reaction.

This article highlights past and future subcritical experiments conducted at the NTS by Los Alamos with the operating partner Bechtel Nevada, the Atomic Weapons Establishment (AWE) in the United Kingdom, and the Lawrence Livermore National Laboratory. It also discusses the new Atlas pulsed-power facility residing at the test site and a possible future site for criticality experiments.

## Subcritical Experiments

Kismet was the first experiment at U1a after the 1992 moratorium on nuclear testing. It was really a proof-of-principle test for determining the most functional layout plan for underground cavities, known as alcoves, that would house subcritical experiments supporting the readiness pro-



**Figure 1. Aerial View of Area 1 at NTS**

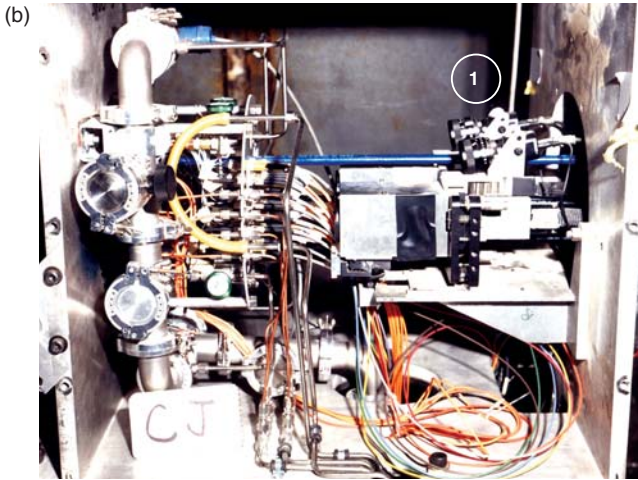
gram and stewardship mission. In Kismet, we used only a small amount of HE to revive studies of downhole methods—for example, recovering data over very long lines. From the test, we obtained relevant information that helped us plan and prepare the test alcoves in the U1a complex.

A whole series of subcritical experiments followed Kismet. The first few were carried out in dedicated alcoves mined at the U1a complex for containment purposes, the traditional way of conducting experiments in an underground environment. Rebound and Stagecoach were the first and second subcritical experiments fielded by the Laboratory. For these and other past subcritical experiments described below, please refer to the pictorial summary on the next two pages. Rebound and Stagecoach were aimed at providing information about the behavior of plutonium alloys when compressed by high-pressure shock waves. Two different alloys were used in the experiments: new alloy in Rebound and aged alloy in Stagecoach (up to 40 years old).

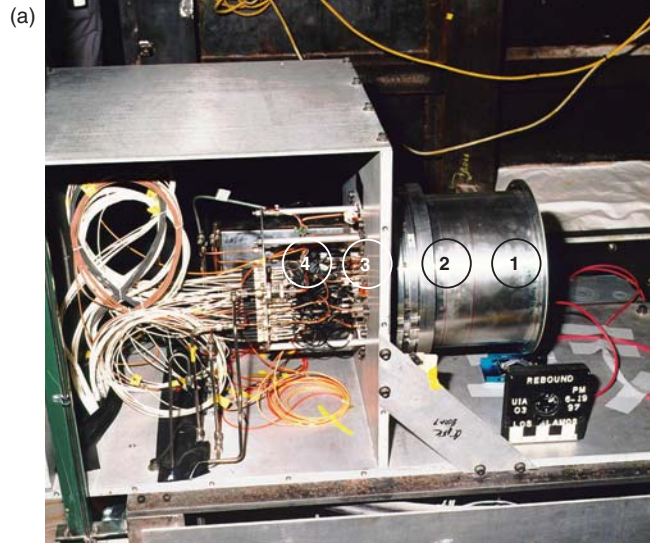
Diagnostic techniques derived from Rebound were refined in Stagecoach. The valuable data obtained on the equation of state of plutonium provided input to our modeling codes for certification of existing weapon pits. At the same time, those data gave useful information about aging effects and manufacturing site variability on plutonium alloys. During these experiments, we also developed diagnostics to be used in future experiments.

The next two subcritical experiments, Cimarron and Thoroughbred, continued our effort in support of the stewardship mission and readiness program and contributed to the development of diagnostics to study the dynamics of pit performance. These two experiments were conducted on mockup pit geometry inside mined alcoves. Relevant data were obtained on the performance of plutonium produced by different manufacturing methods and sources. In addition, an extensive list of lessons learned from the Cimarron experiment was implemented in the Thoroughbred experiment.

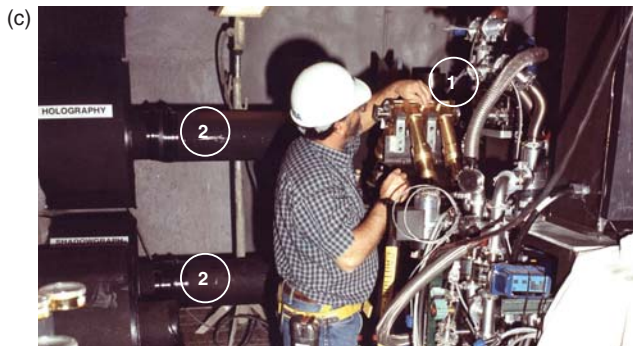
# Alcove Subcritical Experiments



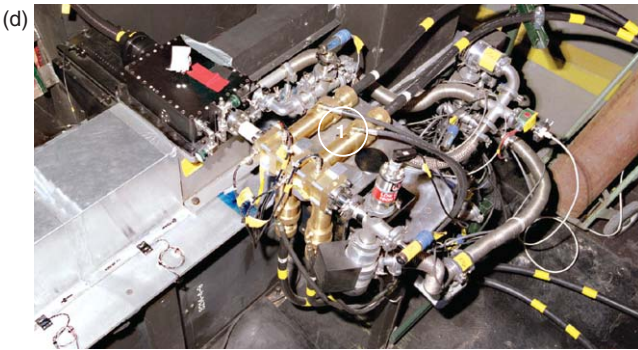
(1) Optic alignment gear



(1) HE package, (2) flyer plate, (3) sample plate, and (4) diagnostics and cabling



(1) X-ray diagnostics and (2) optical diagnostics (shadowgraphy and holography)



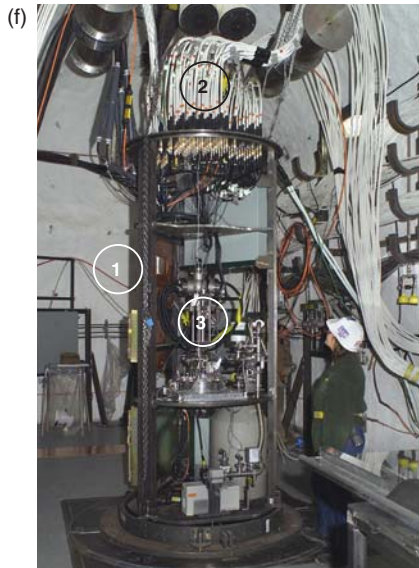
(1) Brass heads



(e)

Since the moratorium on underground testing, Los Alamos has been conducting important subcritical experiments, whose results help validate our computer modeling capabilities. Data about the equation of state of plutonium from Rebound (a) and Stagecoach (b) provided input to our modeling codes that contribute to the certification of existing weapons systems. For the Cimarron and Thoroughbred experiments, shown in (c) and (d), we developed techniques to measure pit performance. These two experiments were primarily intended for ejecta studies, or studies of particles propelled from a material's surface when the material is compressed by a powerful shock wave. The x-ray and optical diagnostics measured ejecta from a shocked plutonium surface. The black pipes in the background in (c) are line-of-sight pipes for the optical diagnostics shadowgraphy and holography, which can image ejecta particles in two and three dimensions, respectively. The x-rays generated in the four brass heads shown in (d) are directed through the plutonium ejecta. X-ray intensity is transformed into optical signals, which are then transferred to the recording system. High-frequency data were captured underground and were transferred to computers in a trailer (e). The data retrieved on the surface included timing and plutonium ejection characteristics.

# Drilled-Hole Subcritical Experiments

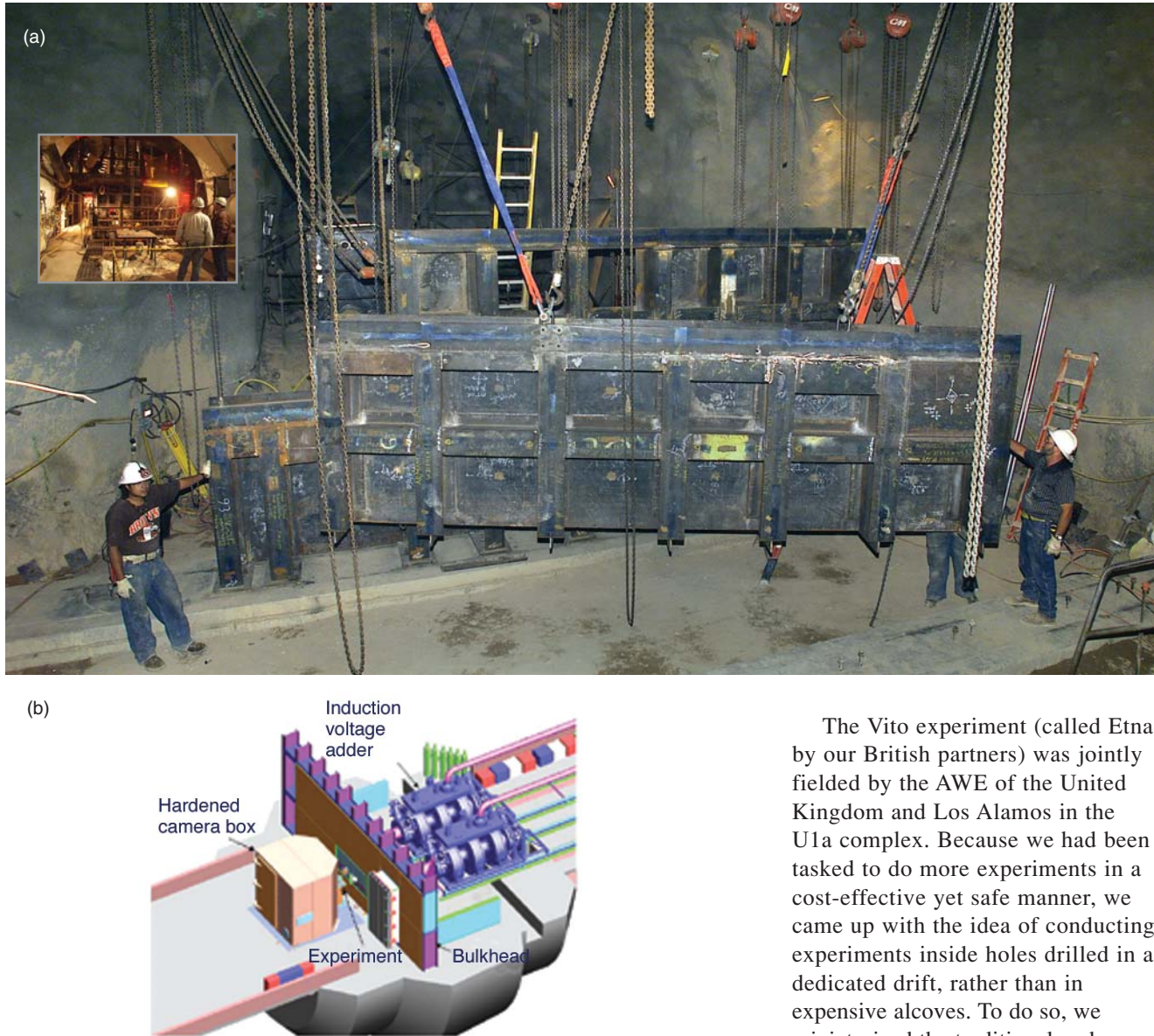


(1) Racklet, (2) fiber-optic cables, and (3) AWE package



Like Cimarron and Thoroughbred, the Vito (Etna) experiment (f) was also primarily intended for ejecta studies. Conducted in a drilled hole at 35 feet below the drift floor (called “invert” in mining jargon), Vito tested our readiness capabilities. Shown here is the 10-ft racklet with the experiments, diagnostics, and vacuum equipment in place. At this point, we are ready to insert the experimental physics package, the last operation before emplacement. The Mario (g) and Rocco (h) experiments followed Vito and were primarily intended for studies of surface properties. They contained optical diagnostics that looked at spall. The racklet shown in (g) is ready to receive the subcritical package. In (h) the racklet is shown resting on the support collar while the emplacement hardware is being prepared for lowering it into the hole. The series of photos from (i) through (l) shows the steps observed for emplacing and sealing (“stemming” is the word used at the site) the racklet into the drilled hole. Before being emplaced, the racklet is carefully lowered into a canister (i). In (j), the racklet is shown almost inside the canister. Once the racklet is inside, technicians bolt it down, for safety, and lower it into the drilled hole (k). The racklet and canister are then stemmed, a process shown in progress in (l). The workers, each wearing a yellow safety harness, are pouring stemming materials into the hoppers, which have a hose connected to the spout and direct the materials where needed.

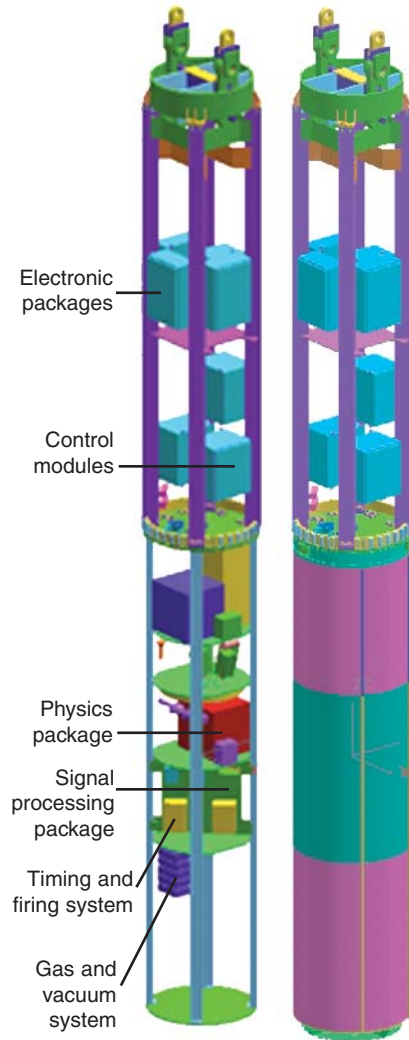




**Figure 2. Armando**

Spall measurements are the focus of the Armando subcritical experiment. In (a), ironworkers are positioning a bulkhead for the Armando alcove, and the inset shows the almost completed alcove. The area facing the bulkhead corresponds to the right part of the schematic in (b), where two x-ray systems are placed. The induction voltage adder increases the electron energy and is a technically sophisticated part of those systems. The area to the left of the bulkhead in (b) includes the physics package containing the HE. Typically, the experimental complex is destroyed by the blast. For cost-effectiveness, we propose to contain the package in a specially designed vessel that will protect the experimental complex and permit multiple uses of the equipment. Together with the Sandia National Laboratories, Albuquerque, we developed the x-ray prototype, which is being replicated commercially by PSI-TITAN, our industrial partner. Naval Research Laboratory staff have configured the x-ray diode.

The Vito experiment (called Etna by our British partners) was jointly fielded by the AWE of the United Kingdom and Los Alamos in the U1a complex. Because we had been tasked to do more experiments in a cost-effective yet safe manner, we came up with the idea of conducting experiments inside holes drilled in a dedicated drift, rather than in expensive alcoves. To do so, we miniaturized the traditional racks used in the days of nuclear underground testing and placed the subcritical experimental package, diagnostics, timing and firing equipment, and cabling into these new structures called “racklets.” The racklet and its cargo would then be lowered 35 feet below the drift floor into a drilled hole, whose top would be stemmed (or sealed). That is how Vito (Etna) was conducted, and it allowed us to exercise our readiness capabilities. It also allowed our British partners to conduct studies of actual pit dynamics, including timing, HE



**Figure 3. Unicorn**

The Unicorn subcritical experiment will measure early-time behavior in a pit. The data and information obtained from this experiment will be integrated with those from previous experiments to enhance our modeling codes as part of the stewardship mission. Unicorn will also allow us to exercise our readiness capabilities. Shown at left is a cartoon of the Unicorn rack (30 ft in height) and its canister.

performance, and plutonium ejecta characteristics.

The Mario and Rocco subcritical experiments were also placed in drilled holes, and they measured the early-time hydrodynamic behavior of plutonium mockup segments manu-

factured at different facilities and machined by different techniques. Wrought plutonium from Rocky Flats was used in the Mario experiment, and cast plutonium from Technical Area (TA) 55 at Los Alamos, in the Rocco experiment. The two experiments provided comparison data for the shift of pit manufacturing sites from Rocky Flats to Los Alamos.

The Armando (Figure 2) and Unicorn (Figure 3) experiments will be conducted in 2004 and 2005, respectively. Armando will enable comparative studies of the performance of plutonium pits of actual geometry manufactured by the Los Alamos and Rocky Flats methods. A sophisticated x-ray system, built by Los Alamos and Sandia National Laboratories staff in collaboration with Bechtel Nevada, was tested at Los Alamos and will be transported to the test site. The x-ray diode was configured by the Naval Research Laboratory, and the whole system is being replicated by PSI-TITAN, our industrial partner. The two radiographic systems will be installed in a special alcove and prepared to measure spall characteristics from each pit simultaneously. For worker safety, the x-ray system must be properly integrated with the underground environment, a crucial but difficult task.

A subcritical experiment as well, Unicorn will be lowered from the surface down a 600-foot-deep hole. It will thus more closely resemble the physical conditions of an underground test and give a better measure of our readiness capabilities.

In addition, four other subcritical experiments are planned in support of the W88 Certification Project. As we complete this project, we will develop the next series of subcritical experiments, including fundamental physics experiments intended to support the enduring stockpile.

## Activities of Lawrence Livermore National Laboratory

The subcritical experiments conducted by our sister laboratory, the Lawrence Livermore National Laboratory, in the U1a complex and at aboveground complexes such as the Big Explosives Experimental Facility parallel Los Alamos work at the NTS. Similar to Los Alamos studies, Livermore studies have focused on smaller scale tests (but Livermore conducts more such tests than Los Alamos) to obtain data on plutonium spall, ejecta, and other dynamic properties. The variances in aging, manufacturing methods, and changes in plutonium production facilities are also part of Livermore's program. Livermore is also developing the Joint Actinide Shock Physics Experimental Research Facility, referred to as JASPER, in Area 27 of the test site. JASPER (refer to Figure 4) is a two-stage light gas gun that fires projectiles at plutonium samples at speeds of up to 8 kilometers per second. As a result, very high pressures (6 megabars) are generated in the sample.



**Figure 4. JASPER**

This two-stage light gas gun is a significant scientific achievement because samples can reach very high pressures.



## Atlas, the Pulsed-Power Facility

The last few years have seen the emergence of a new capability for exploring material behavior under the unique conditions associated with the operation of a nuclear device. Called pulsed-power hydrodynamics, this capability has become an essential tool for stockpile stewardship.

In a pulsed-power facility (Figure 5), very high magnetic fields produced by very large electrical currents implode a relatively thin-walled conducting cylinder, called a liner, to high velocity while maintaining the imploding material at near-solid density— and largely unmelted, as shown in Figure 6(a). Pulsed-power hydrodynamics produces implosions of unprecedented precision. When liners are imploded, their circularity can be maintained to far better than 1 percent of the initial radius, and their axial uniformity can be preserved to the limit of the imaging resolution. This level of precision



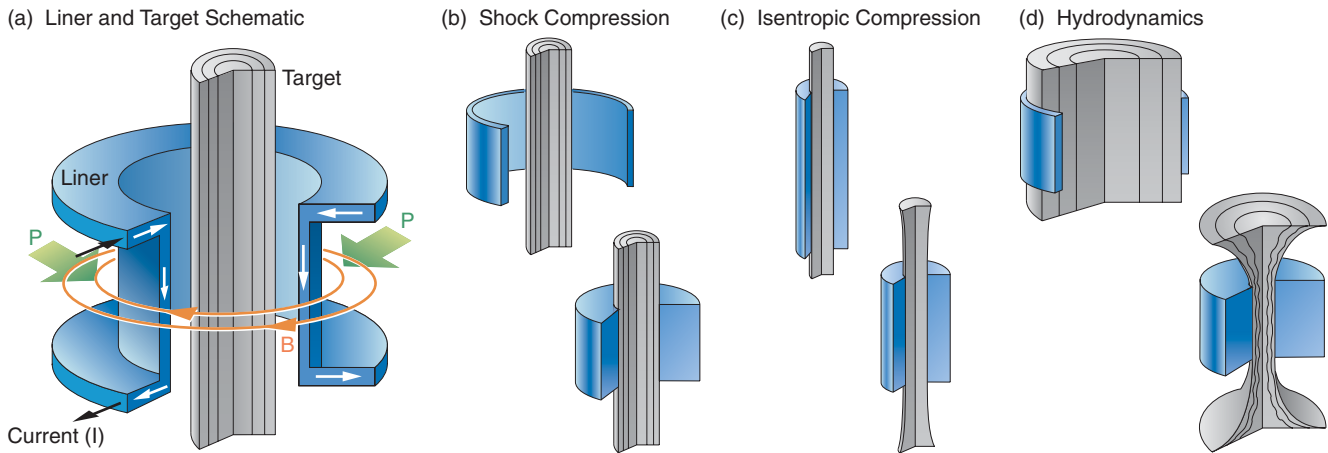
**Figure 5. Atlas**  
The Atlas pulsed-power system is essential to the Laboratory's stockpile stewardship mission.

allows liners to be used as drivers for materials properties and hydrodynamics experiments that are in a converging rather than a planar geometry (planar geometries—for example, those in a light gas gun—have long been the standard ones).

The most attractive pulsed-power system for driving such experiments is an ultrahigh-current, low-impedance, microsecond-time-scale source that is both economical to build and reliable

to operate. The Atlas system, shown in Figure 5, is the world's first pulsed-power system to be specifically designed and optimized for pulsed-power hydrodynamics experiments. Atlas was designed and built at Los Alamos and entered experimental service in September 2001. Within one year of having completed shake-down experiments, Atlas was disassembled and is being moved to a new facility in Area 6 at the NTS, where it is scheduled for recommission in 2004. Atlas will resume experiments shortly thereafter. Atlas is capable of delivering 30-mega-ampere currents in a heavily damped sinusoidal waveform with a 5- to 6-millisecond rise-time and is ideal for driving liners up to 10 centimeters in initial radius and up to tens of centimeters in initial length.

Magnetically imploded liners offer unique advantages as drivers for pulsed-power hydrodynamics applications. Because energy is delivered to the liner from the magnetic field at the



**Figure 6. Magnetically Imploded Liners**  
(a) This schematic shows a metal liner surrounding a target cylinder. A very large current sent through the body of the liner (black arrows) creates a very strong magnetic field (orange field lines). The interaction between the current and the magnetic field produces an inward-directed force that implodes the liner, driving it toward the target. Data about the behavior of the target as it is being compressed are used to validate modern computer codes. We have precise

control of the implosion process, and can (b) drive the liner at extremely high velocities to deliver a strong shock to the target, (c) compress a target at nearly constant entropy (isentropic compression) to reach states of matter not accessible from a single shock, and (d) compress targets hydrodynamically to study instabilities and interfaces. Because of the cylindrical geometry of both the liner and the target, we also have good diagnostic access to the target transverse to and down the cylinder's axis.

speed of light, magnetically imploded liners can reach velocities higher than those available from gas guns or planar explosive systems. Higher velocity in the liner (or impactor) means higher pressures and temperatures in a strong shock delivered to the target located at its center, extending the range of traditional Hugoniot data to well above 1000 gigapascals—see Figure 6(b). Because the parameters of the electrical drive can be continuously adjusted over a wide range, the liner acceleration profile, and hence final velocity, can be continuously and controllably varied to meet experimental requirements. With appropriate design, the acceleration delivered by the field to the liner is nearly shockless, allowing full characterization of the liner's condition as the liner strikes the target. Furthermore, magnetically imploded liners can shocklessly pressurize a material that is the liner itself or that is initially in contact with the liner to reach off-Hugoniot states at pressures approaching 100 gigapascals—see Figure 6(c). The size of magnetically driven liners is naturally associated with centimeter-sized targets. At these scales, the target is many times the characteristic size of grains in the material of which it is made, allowing reliable probing of continuum properties. The fundamentally cylindrical geometry permits good diagnostic access both transverse to and down the cylinder's axis. Because liners can also hydrodynamically compress fluid structures, the size and geometry of the target permit studies of interface behavior and of the growth of instabilities—refer to Figure 6(d).

### Future Site for Criticality Experiments

Historically, TA-18 at Los Alamos has been used for criticality and safety studies of various materials used in the weapons programs. This nuclear



**Figure 7. The Device Assembly Facility**

**This aerial view shows the facility proposed to house future criticality and safety studies.**

facility is the nation's only remaining one for general-purpose nuclear materials handling for various experiments, measurements (to determine the presence of nuclear materials), and training. Under consideration is a proposal to relocate this facility and capabilities to the NTS at the Device Assembly Facility shown in Figure 7. Integrating these capabilities with those already in place will provide more efficient use of our NTS resources as we meet the challenges of the stewardship mission.

*For further information, contact Ghazar Papazian (505) 667-0403 (raffi@lanl.gov).*

### Outlook

The NTS continues to be a testbed for experiments that return unique and crucial data in support of the enduring stockpile and fundamental weapons physics. The location and geology of the site, coupled with a traditional “can-do” attitude, serve the laboratories and the nation well. Conducting subcritical experiments not only supports certification but maintains an operational test-readiness infrastructure. The transition of Atlas and of operations at TA-18 will augment the mission space in which the NTS conducts activities today. ■



# The Development of Flash Radiography

*Gregory S. Cunningham and Christopher Morris*

*The Manhattan Project  
PHERMEX and DARHT  
Proton Radiography at LANSCE*

In many ways, flash radiography is to nuclear weapons what medical x-radiography is to the human body: It allows one to see inside a complex structure without disturbing it. Starting in the Manhattan Project and continuing to this day, flash radiography has been used to take stop-action pictures of dynamic events: from the detonation of high explosives to the implosion of a mock weapon assembly containing a surrogate material for the nuclear core.

In this article, we will explain the basic principles of flash radiography and trace decades of progress in improving image quality. A major goal is to follow the hydrodynamic implosion, or hydrotest, to the point at which the surrogate core is maximally compressed. Los Alamos state-of-the-art x-ray hydrotests at the Dual-Axis Radiographic Hydrodynamic Test (DARHT) facility and facilities envisioned for the future will produce time-sequenced images of the implosion dynamics and provide views of the implosion along multiple lines of sight. Images from hydrotests at DARHT are already playing a crucial role in solving stockpile issues, including those related to the certification of remanufactured parts. A recent Los Alamos invention, proton radiography, is now fielded at the Los Alamos Neutron Science Center (LANSCE), the Laboratory's medium-energy accelerator facility. Proton radiography is also providing important data for the nuclear weapons program. A higher-energy proton radiography machine has the potential for providing a new level of quantitative precision and quality to the data from hydrotests.

## Historical Origins

Quite remarkably, the best analog to our current experimental program in science-based stockpile stewardship is found in the program to develop the

plutonium implosion bomb during the Manhattan Project. It was known from the start that enough fissionable material for a single bomb, either uranium-235 or plutonium-239, would not become available until late in the project. Consequently, the gun device—which propels one subcritical piece of fissionable material into another at high speed—was the favored method for assembling a supercritical mass. Because it was straightforward, this approach had the highest probability of success, whereas the spherical implosion of a subcritical configuration would present major technical challenges.

At that time, the nuclear properties of plutonium-239, the new manmade isotope, had been only crudely determined. When a barely visible speck of plutonium produced at Ernest O. Lawrence's cyclotron at the Berkeley Radiation Laboratory arrived at Los Alamos, scientists from the Physics Division used the material to measure more definitively the neutron number per fission and the cross sections for fission neutron capture and scattering. These were the first nuclear experiments completed at Los Alamos, and the results were encouraging. Because the neutron number for plutonium was indeed higher than that for uranium-235, plutonium would likely yield a more efficient nuclear explosion.

These measurements were fed into computational bomb design models as soon as they became available. Later, however, when reactor-produced plutonium arrived, the scientists detected a high neutron background, which, Enrico Fermi quickly showed, derived from the spontaneous fission of the isotope plutonium-240, a reactor byproduct present in the sample. Gun assembly of plutonium pieces containing plutonium-240 would be too slow to prevent premature initiation of the chain reaction by the neutrons from spontaneous fission, and therefore the likely outcome would be a fizzle rather than an efficient nuclear explosion. This finding forced the project to

switch goals and aim for an implosion device.

In the implosion device, high explosives surrounding a spherical assembly would be detonated at many points, and the resulting converging spherical detonation wave would compress the nuclear material to a supercritical configuration. One had to measure the velocity of the implosion and the state of the metal during assembly in order to determine the optimal timing of the neutron initiators needed to achieve successful device performance. Director J. Robert Oppenheimer called this endeavor “one of the most urgent of the project's outstanding problems.” Its solution occupied a talented group of physicists, chemists, and electrical and mechanical engineers. Details of the dynamic response of materials subjected to high-explosive drive were studied in small-scale experiments in GMX Division (predecessor of the present Dynamic Experimentation Division). To test the overall performance of the implosion device, the implosion group performed so-called “integral” experiments on mock assemblies, which had the correct geometry and components except for a nonfissionable, surrogate core in place of the plutonium pit.

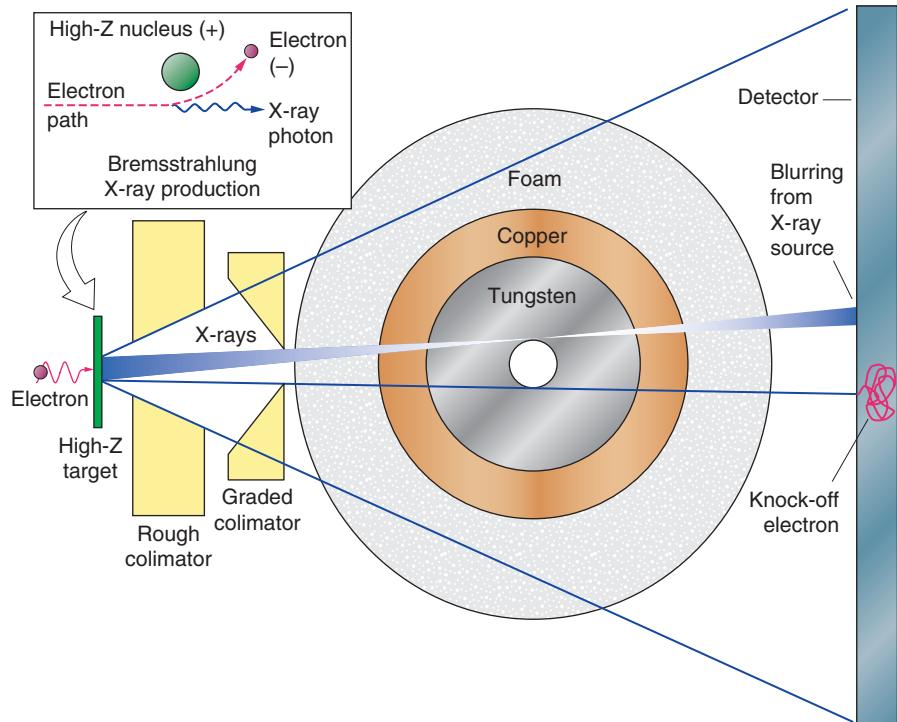
As we discuss below, radiography with x-rays was a key diagnostic for those small-scale and integral tests and has remained so through the decades. In fact, many of the experimental tools developed in the 1940s to study the evolution of high-explosive-driven systems—electrically charged metal pins, optical framing cameras, and flash radiography—are still the standard diagnostics for monitoring weapons implosion. Today, through the use of vastly improved equipment and modern data acquisition and analysis methods, these tools are still helping us quantify important phenomena (such as the details of material failure and high-explosive detonation) and develop accurate, predictive physical

models to describe them. Before we trace the history of flash radiography for studying weapon implosion, we explain the basic principles of this technique, as well as some limitations that we hope to circumvent with advanced techniques.

### Attenuation Radiography for Stockpile Stewardship

In modern hydrotests, we replace the fissionable pit of a nuclear primary with a mock pit made of a surrogate material such as natural uranium, lead, or tantalum. This nonnuclear system is imploded and its dynamics studied to provide constraints for the physics models used in numerical simulations of weapons performance. The primary diagnostic of hydrotest experiments is point-projection flash radiography. At the times of most interest, the experiment is illuminated with a short pulse of x-rays, and the transmitted flux (number of x-ray photons per unit area) is recorded on a suitably shielded detector.

Figure 1 is a diagram showing the geometry of a static experiment on the French test object (FTO), which was designed to allow French and U.S. experimenters to collaborate on flash radiography methods and analysis. High-energy x-rays are produced through the bremsstrahlung interaction of energetic—10 to 30 million electron volts (MeV)—electron beams with high-Z targets (that is, targets made of materials with high atomic numbers). Interaction with a positively charged nucleus causes an electron to bend (accelerate) and therefore radiate, or emit, photons. The loss of energy brakes its speed, hence the term bremsstrahlung (or braking radiation in German) for both the process and the emitted radiation. Although the emitted photons have a continuous energy spectrum, most of the photons that are transmitted through a



**Figure 1. Basics of Flash X-Radiography**

This schematic shows (from left to right) the production of x-ray pulses, their transmission through the so-called French test object (FTO), and their detection. At left, high-energy electrons (red) hit a high-Z target and interact with the heavy nuclei to produce x-rays (blue). The x-rays are attenuated by a rough collimator, which defines the field of view and then by a graded collimator that flattens the transmission profile to reduce scattering into the center of the image. The FTO consists of an inner spherical shell of tungsten (inner radius = 1 cm and outer radius = 4.5 cm) and an outer shell of copper (outer radius = 6.5 cm) surrounded by a shell of foam (outer radius = 22.5 cm). The location of material interfaces recorded at the detector are blurred because the x-ray source has a finite extent and because electrons knocked into motion by arriving x-rays have a finite range in the detector.

hydrotest assembly have an energy near 4 MeV. That energy is near the minimum in the absorption cross section for the materials in the assembly.

A crucial performance parameter that we would like to be able to measure with flash radiography is the density of the surrogate material at nuclear time—the time a “real” system would start producing a significant amount of nuclear energy. As we will discuss below, by measuring the attenuation of the x-ray flux transmitted through the center of the assembly, we can determine the integrated quantity  $\rho_A$ , the areal density, or line-of-sight mass, of the implosion system:

$$\rho_A = \int_0^L \rho(x, y, z) dz, \quad (1)$$

where  $\rho(x, y, z)$  is the volume density of the hydrotest object and  $z$  is the longitudinal coordinate through the assembly. For an object with constant density,  $\rho_0$ , the areal density is  $\rho_0 \times L$ , where  $L$  is the longitudinal thickness of the object.

Because the areal density of the system at nuclear times is very large, we need an enormous dose (that is, intensity  $\times$  time) of x-rays to make the transmission measurement. The doses currently available, even from the lat-

est flash x-ray machines, are insufficient to provide the quality of data that weapons scientists will require to adequately constrain their calculations for future certification. In pursuit of better data quality, flash x-ray machines are being constantly upgraded and improved; at the same time, new data analysis technologies are being developed and implemented at Los Alamos and other weapons laboratories.

The most basic physics of transmission radiography is contained in the Beer-Lambert law, the solution to the differential equation that describes the number of particles surviving transport through a medium without interaction. The Beer-Lambert law can be derived from investigating transmission through an infinitesimally thin piece of material with constant density  $\rho_0$  and thickness  $l$ . In this case, the probability that a particle goes through with no interaction is  $(1 - \sigma\rho_0 l/A)$ , where  $\sigma$  is the cross section for interaction in centimeters squared and  $A$  is the atomic mass in grams. For a material with finite thickness  $L$ , the probability of no interaction is

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\rho_0 L}{\lambda n}\right)^n = e^{-\frac{\rho_0 L}{\lambda}}, \quad (2)$$

where  $\lambda = A/\sigma$  is called the interaction length in grams per centimeter squared ( $\text{gm}/\text{cm}^2$ ). Thus, if  $N_0$  photons impinge on the material, then, on average, the number  $N$  that will make it through without interaction is given by

$$N = N_0 e^{-\frac{\rho_0 L}{\lambda}}. \quad (3)$$

This equation can be generalized to materials with varying density as

$$N = N_0 e^{-\frac{\rho_A}{\lambda}}, \quad (4)$$

where  $\rho_A$  is the areal density defined

in Equation (1).

Equation (4) tells us that we can determine the areal density of the object in units of the interaction length  $\lambda$  of the incident radiation by measuring the ratio of incident to surviving particles:

$$\frac{\rho_A}{\lambda} = -\ln\left(\frac{N}{N_0}\right). \quad (5)$$

This simple analysis leaves out important details. For example,  $\lambda$  depends on energy, and the x-ray source is not monoenergetic. Also, scattered x-rays produce background “fog” in the image. Nevertheless, this simple analysis provides an important guide for evaluating and developing radiographic tools.

For example, we can calculate the uncertainty in the measured value of areal density,  $\Delta\rho_A$ , under the assumption that the only source of noise is the Poisson (counting) statistics of the transmitted beam. That uncertainty is given by

$$\frac{\Delta\rho_A}{\lambda} = \frac{1}{\sqrt{N}}. \quad (6)$$

The optimal interaction length  $\lambda_{\text{optimal}}$  would be one that minimizes  $\Delta\rho_A$  for a given object. Setting to zero the derivative of the uncertainty with respect to  $\lambda$  and solving for  $\lambda$ , we find that  $\lambda_{\text{optimal}} = \rho_A/2$ , or the optimum equals half the thickness (or areal density) of the object. For the FTO,  $\rho_A = 182.5 \text{ gm}/\text{cm}^2$ , so the optimum interaction length is  $91.75 \text{ gm}/\text{cm}^2$ . Can we achieve such a long interaction length?

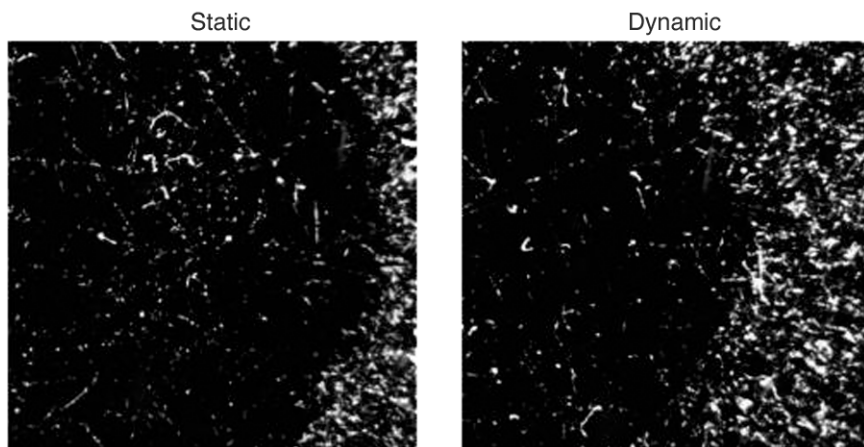
In the case of x-rays,  $\lambda$  varies strongly with x-ray energy. The interaction length reaches a maximum value at the energy at which the cross section for producing electron-positron pairs (that cross section increases with increasing energy) becomes comparable to Compton scattering (which decreases with increas-

ing energy). The maximum interaction length of x-rays is weakly dependent on atomic number,  $Z$ , and therefore for all high- $Z$  materials,  $\lambda$  is maximum at about the same x-ray energy, namely, near 4 MeV. The interaction length  $\lambda$  in uranium for 4-MeV x-rays is about  $22 \text{ gm}/\text{cm}^2$ , or a little over a centimeter in natural uranium, much smaller than the thickness of a hydrotest assembly. The relatively short interaction length of x-rays in heavy elements implies a large uncertainty in the areal density measurements, even when extremely high doses are used.

## The First Radiographs of Implosion

Let us now go back to the origin of weapons radiography. In the spring of 1944, the Manhattan Project shifted its main focus to implosion after the discovery that reactor-produced plutonium had a high neutron background that was due to the presence of plutonium-240. The previous fall, John von Neumann had suggested that, with enough high explosive driving an implosion of a fissionable metal core, one could ignore the strength of the material and assume that the solid material behaved like a fluid. In this case, partial differential equations could be written and solved in a numerical program on IBM machines to determine the velocity of implosion. However, the input to the equations, that is, the high-explosive drive and the equation of state (EOS) of the metal and explosives needed to be determined. Flash radiography was one of the important diagnostic techniques used in quantifying the spherically converging high-explosive drive.

In the initial experiments conducted in 1943, Seth Neddermeyer’s group tried surrounding a small metal sphere with weak explosive charges and detonating it at many points. The scientists expected the diverging spherical



**Figure 2. Radiographs of an Explosively Driven Implosion Experiment**  
 These radiographs were taken in 1944 to image the outside edge of the surrogate pit during a hydrotest. The x-rays used to image the implosion were generated with the 15-MeV electron beam from the betatron borrowed from the University of Illinois. The detector consisted of a lead glass converter and a Wilson cloud chamber to detect the recoil electrons. The dark area in each image is the shadow cast by the pit. The radius of the pit is smaller in the right image.



**Figure 3. The Radiolanthanum Experiments**  
 This photo taken at Los Alamos during the Manhattan Project shows the “remote handling” of a kilocurie source of radiolanthanum located inside a lead container. This strong gamma-ray source would be placed at the center of a hydrotest assembly to measure the areal density of the pit (from the center outward) as a function of time.

waves from each detonation point to cancel each other out upon interaction, creating the desired converging detonation wave. The results were disappointing. The recovered ball of metal had been compressed but showed many

asymmetries. Later experiments with multiple points of detonation around a cylindrical shell showed that high pressure develops where the detonation waves collide, which could result in the formation of “jets.” The group then

modified commercial x-ray machines to achieve precision timing of the x-ray flashes to about 1 microsecond so that the x-ray flashes could be coordinated with the explosive shots. Indeed, the resulting images confirmed that jets did form at the interaction between multiple detonation waves. That diagnosis led to the design of explosive “lenses,” which shape the individual detonation waves so that such high-pressure areas do not form. Flash radiography of small imploding metallic spheres revealed two other reasons for nonideal implosion: density variations in the explosives and asynchrony among the individual detonators. These discoveries led to improvements in the manufacture of explosives and to “electric detonation” for more reliable timing.

To interrogate implosion of a full-scale device (with a surrogate pit material), they would need more penetrating radiation—x-ray photons with energies near 4.0 MeV. Oppenheimer decided to acquire the University of Illinois betatron, an electron accelerator that produced 1-microsecond-long pulses of 15-MeV electrons. As already discussed, the high-energy electrons produced bremsstrahlung radiation as they passed through a high-Z target. The high-energy photons produced by the betatron penetrated the high explosive of a full-scale device but were stopped by the large areal density of the pit itself; therefore, a “shadow” of the pit’s outer contour could be observed by detection of the photons that made it through the device. These photons were detected when a sheet of lead glass was placed on the other side of the device. Interactions of x-rays with the atomic electrons in the lead glass produced energetic recoil electrons, and those recoil electrons made visible tracks in a vertical cloud chamber. That system provided the first flash radiograph of an “integrated” test (see Figure 2). The work on flash x-ray

imaging of full-scale devices was said to be “among the most impressive of several such achievements at Los Alamos” (Hawkins 1961).

After developing techniques to diagnose symmetric implosions, the Manhattan project pioneers wanted to measure what was happening inside the pit. For that purpose, they placed a small capsule of the radioactive isotope lanthanum-140 at the center of the pit. As the high explosive compressed the pit, the 1.46-MeV photons from the decay of the lanthanum-140 penetrated through to the outside of the pit, and their intensity was measured as a function of time. Those data provided a measure of the areal density of the pit as a function of time. Although the experiments were effective, the environmental hazards (see Figure 3) of both their production and their aftermath resulted in abandonment of the program in 1962.

## PHERMEX

The Los Alamos facility known as PHERMEX (for pulsed high-energy radiographic machine emitting x-rays) was commissioned in 1963 (see Figure 4). It was the first of a new generation of flash x-ray machines designed to produce enough flux to penetrate the center of a hydrotest experiment at “nuclear” time. The design of this high-energy pulsed x-ray machine, including techniques for recording the images, was the culmination of extensive Los Alamos studies led by Doug Venable and completed in the early 1950s.

Those studies defined the dose needed to penetrate the center of a hydrotest assembly and the feasibility of getting good images of the less-dense parts of the assembly. It was shown that systems up to  $4 \lambda$  in thickness could be imaged on film detectors. For objects up to  $10 \lambda$  in thickness, the large background of scattered radiation would obscure the film images.



**Figure 4. A PHERMEX Shot**

This photo shows an explosive shot at PHERMEX. The high-energy pulsed x-ray machine has served the weapons program for 40 years.

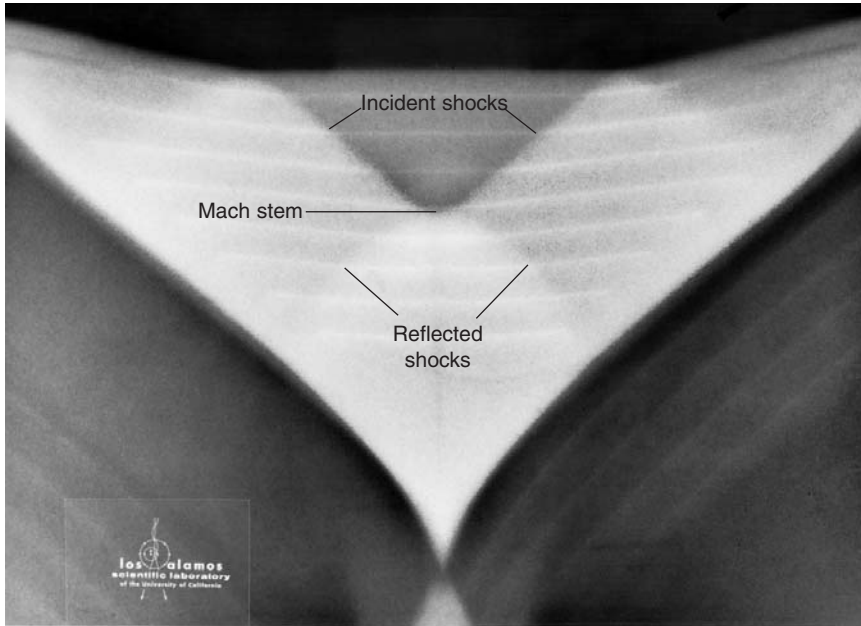
The studies were performed on static test objects that had been stretched in the beam direction to have areal densities commensurate with the high volume densities reached during hydrotest implosions. The doses needed to measure the internal densities of those test objects were determined as a function of scale. The idea was that, although PHERMEX could not see through a full-scale device, the hydrodynamics could be studied at quarter or half scale with surrogate materials. To minimize the obscuring effects of scattered x-rays in the thickest regions, the objects were radiographed through a graded collimator designed to be an approximate inverse of the object. Graded collimation dramatically reduced the scattered background while still allowing the thinner parts of the object to be seen. This technique has been crucial in enabling radiography across the full range of configurations reached during hydrotests.

At PHERMEX, three very large 50-megahertz radio-frequency (rf) res-

onators provide the energy needed to accelerate short pulses of electrons totaling 9 microcoulombs of electric charge to 30 MeV. Those pulses are then directed at a high-Z target to produce x-rays. When PHERMEX was commissioned, it produced 200-nanosecond-long pulses of x-rays exceeding 9 roentgens at 1 meter from the production target.

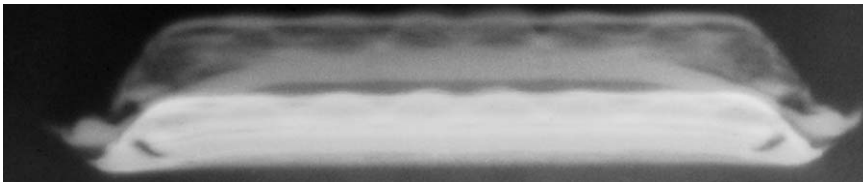
Thousands of experiments have been performed at the facility, including small-scale experiments to develop the physics of high-explosive-driven systems (see Figures 5 and 6), and a large number of major hydrotests. Much of this work has been compiled and presented in a marvelous summary of shock physics (Mader et al. 1980). PHERMEX has also undergone many upgrades during its lifetime. Currently, it can produce a single 200-nanosecond-long pulse with a spot size of 3 millimeters and a dose of 400 roentgens at a distance of 1 meter from the target, more than 40 times the





**Figure 5. PHERMEX Radiograph of Hydrodynamic Flow**

Dating from the late 1960s, this radiograph shows colliding shock waves in aluminum. The horizontal lines are thin foils of high-density metal interspersed in the aluminum to indicate the material flow. A Mach stem has formed at the intersection of the shock waves.



**Figure 6. Double-Pulse PHERMEX Radiograph of Spall in Iron**

An iron plate is driven by an explosive initiated at a number of individual points. The iron has spalled into a series of layers, and the effects of the initiation points are apparent. Successive transmission of two short x-ray pulses produces two sequential images on one film.

dose provided when it was commissioned.

For most of PHERMEX's history, a technique known as screen-enhanced film has been used to record images of the experiments performed in front of the facility. That is, the transmitted x-rays are converted through Compton scattering into moving electrons in a millimeter-thick sheet of lead. Then, as the electrons slow down in the sensitive photographic film located behind the lead sheet, their tracks are recorded (see

Figure 1). Recently, active cameras have been developed and fielded. They have higher sensitivity, higher quantum efficiency, and wider linear dynamic range than film. The latest version of these active cameras can take up to four sequential images and thereby tap the two-pulse capability of PHERMEX. For the first time, we can get high-quality images for each individual pulse (See the article "The DARHT Camera" on page 92.)

Ultimately, the amount of charge (number of electrons) delivered by

PHERMEX, and in turn, the x-ray dose, are limited by the stored energy in the rf resonators. The voltage in the resonators decreases as energy is transferred to the beam; the resulting spread in beam energy at high electron current leads to an unacceptably large spot size on the x-ray production target. Also, high-explosive-driven experiments generate material velocities in the range of several millimeters per microsecond ( $\text{mm}/\mu\text{s}$ ). Thus, the material moves appreciably during the length of the typical 200-nanosecond-long x-ray pulse. Reducing this "motion blur" to levels that do not interfere with the interpretation of the experiment requires pulse lengths of about 100 nanoseconds or less. In spite of the limitations in dose and pulse length, PHERMEX has been a workhorse for the weapons program during the 40 years of its operation.

## The DARHT Facility

A recent review (Ekdahl 2002) of the current state of electron accelerators for flash radiography reports that the United States, the United Kingdom, and France are all developing new flash-radiography capabilities to meet the challenges of maintaining nuclear weapons stockpiles under the restrictions of a moratorium on nuclear tests. This new generation of machines is designed to provide higher doses, better position resolution, shorter pulse lengths, and in some cases, data from a single experiment taken at multiple times and along multiple axes, so that the time-dependence and three-dimensional (3-D) aspects of an implosion can be elucidated.

As early as 1968, Doug Venable had proposed adding a second x-ray axis to PHERMEX, perpendicular to the first, to allow orthogonal tomography. And in 1981, long before the moratorium on testing, John Hopson and Tim Neal presented the first con-



**Figure 7. The DARHT Facility**

(a) The DARHT facility houses two linear induction accelerators set at right angles to each other and focused on a single firing point. (b) Each accelerator consists of a row of induction cells, each coupled to its own energy-storage device. The pulsed-power machine accelerates very large (kiloampere) electron currents, which produce very intense x-ray pulses that are 60 ns long.

cept for the DARHT facility at Los Alamos. It would be built at a new firing site and use two high-dose pulsed-power machines to provide orthogonal x-ray views of a single hydrotest, a capability similar to that at the Atomic Weapons Establishment (AWE) in Aldermaston, England.

The linear induction accelerator (LIA), rather than the pulsed-power diode machines originally proposed, is the technology being used for the two flash-radiography machines at DARHT (see Figure 7). LIAs were pioneered for flash radiography at Lawrence Livermore National Laboratory in the 1980s. Their operating principle is similar to the betatron technology used for the first high-energy flash radiography in Los Alamos in the sense that stored energy from a pulsed-power system is inductively coupled to a high-current electron beam. But in the LIA, the coupling is accomplished by a row of many induction cells rather than a single transformer. Because there are many transformers, each coupled to its own external energy-storage device, the electron currents can be much larger than the limiting currents at PHERMEX. In fact, kiloampere electron currents can be readily accel-

erated in an LIA.

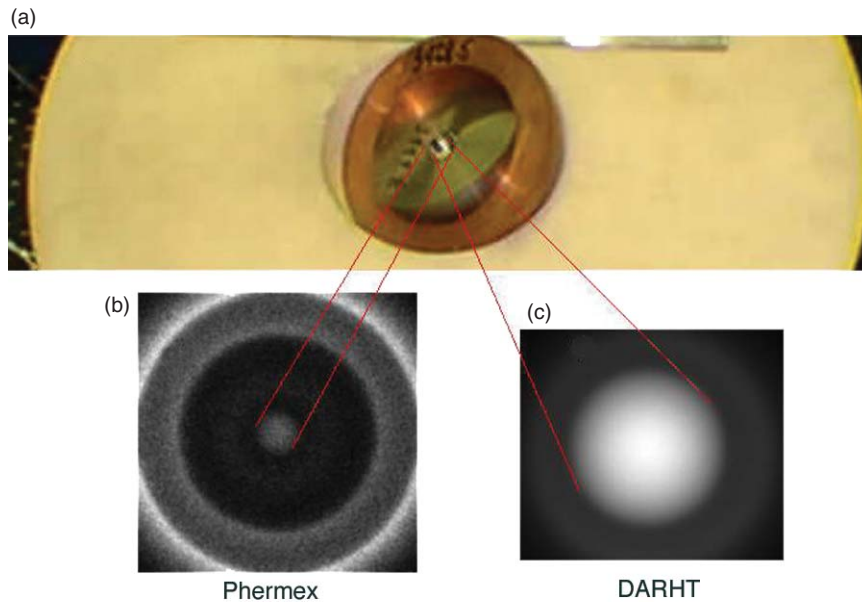
One performance feature to help rank machines that produce different x-ray spot sizes and doses is the root-square-mean (rms) error with which the radiographs from each machine can be used to determine the position of a material interface. This rms error is inversely proportional to the square root of the dose  $d$  and proportional to the radiographic position resolution  $\Delta x$ . Charlie Martin of AWE proposed the simple radiographic figure of merit  $FOM = d/\Delta x^2$ . The position resolution includes contribution from the x-ray spot size, the pulse length (which produces motion blur), and the detector resolution. The machine design determines the first two of these.

The first axis of DARHT, which has already been used for hydrotests, has delivered a dose of 500 roentgens in a 60-nanosecond-long pulse over a 2-millimeter spot. For the same pulse length, PHERMEX can provide a dose of only 120 roentgens in a spot of 3 millimeters. These performance parameters indicate that DARHT achieves about an order-of-magnitude improvement over PHERMEX in terms of Martin's radiographic figure of merit. The second axis of DARHT accelerated its first beam to full energy

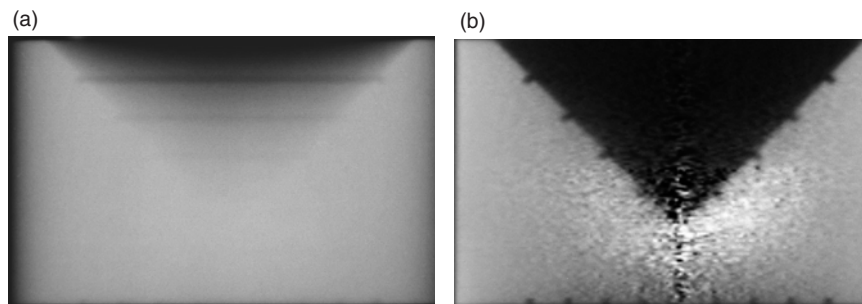
at the end of 2002 and will soon provide pulses with dose and spot size similar to those produced by the first axis. The new feature of this second axis is the production of four such pulses in 2 microseconds. That capability should be available to interrogate hydrotests by the end of 2005.

The first axis of DARHT is already providing weapons scientists with the clearest views ever seen of the inside of a hydrotest. The hydrodynamic data from those tests are used to validate new physics models that are being incorporated into weapons codes. Once the second axis becomes available, scientists will take four sequential radiographs along one axis and one radiograph along the perpendicular axis, thus providing the very first 3-D data from a single U.S. hydrotest.

Figure 8 compares radiographs of the unclassified FTO (refer to Figure 1) taken at PHERMEX in the late 1980s and at the first axis of DARHT. The FTO was designed to compare the performance of flash x-ray machines. The DARHT system—the combination of x-ray source and detectors—clearly demonstrates a very dramatic increase in performance. This facility is expected to be the centerpiece of the nation's hydrotest program for at least a decade.



**Figure 8. A Comparison of X-Radiographs from DARHT and PHERMEX**  
A photograph of half of the FTO is shown in (a), a radiograph of the FTO from PHERMEX is shown in (b), and another radiograph of the FTO from DARHT is shown in (c). The DARHT image reveals a dramatic improvement in quality caused by an improved source and better detector. The PHERMEX radiograph was taken with a graded collimator, whereas the DARHT radiograph was taken with a small, field-of-view, rough collimator, imaging less of the object with lower background. The boxes outline the approximate field of view of each of the radiographs. Comparing these radiographs with the early radiograph in Figure 2 reveals the progress made in flash x-ray radiography since the Manhattan Project .



**Figure 9. Improvements in Data Analysis**

(a) The radiograph of a steel cylinder with a diameter of 12 cm was taken with a cobalt-60 source. The cylinder has a conical section machined out of the top. Square grooves (2 mm by 2 mm) were machined at the bottom of the cylinder and on the inner surface of the conical section to assess the quality of the radiograph and subsequent image processing. (b) The reconstruction process, which first extracts the areal density and eventually the volume density of the cylinder, makes readily apparent many of the grooves that were almost invisible in the original radiograph. The enhanced noise at the center of the cylinder is an unavoidable consequence of this process. The reconstruction process was first implemented in the 1980s.

## Detectors, Collimators, and Data Analysis

The x-ray dose and spot size produced in state-of-the-art multipulse x-ray machines are limited by interactions between the electron beam and the high-Z target. When an electron beam carrying thousands of amperes interacts with an x-ray production target, it creates a high-density plasma of ionized target material and surface impurities. The electrons then interact with the plasma, dynamically changing the effective focal point and increasing the spot size. In addition, the beam causes material at the surface of the target to spall, which reduces the target thickness and, in turn, reduces the dose as a function of time. Both the destruction of the target and the beam-plasma interactions make the goal of multipulse high-dose x-ray radiography difficult to attain. Some progress is being made in techniques to mitigate these problems, but so far, improvements have been only incremental.

Difficulties inherent in increasing the dose have led researchers to search for optimal ways to extract the maximum information from the available doses. They have adopted and extended techniques first investigated at PHERMEX: graded collimation, advanced data analysis, image plate detectors, and multiframe active cameras. More sensitive detectors allow measurements to be made with less incident dose. Large-scale Monte Carlo calculations have led to improved experimental designs that increase signal-to-noise ratio by reducing scattered background. New data-analysis techniques have led to optimally estimating features of interest (refer to Figure 9). These advances in experimental design, detectors, and data analysis have been as important as the increased power and resolution of the x-ray beams for increasing the information that weapons scientists obtain from flash radiography.

## Proton Radiography

In spite of numerous improvements in high-energy flash radiography over the past 50 years, the dose limitations, position resolution, and backgrounds still limit the utility of the technology for obtaining adequate quantitative information from hydrotests for stockpile certification. Recently, a new idea, lens-focused proton radiography, has provided a potential solution to these problems.

As described earlier, electromagnetic scattering processes limit the maximum interaction length of an x-ray to a value far from the optimum for hydrotest experiments. Hadronic probes provide an alternative. Hadrons are fundamental particles, such as neutrons and protons, that interact with matter through the strong (nuclear) force. The absorption cross section,  $\sigma_A$ , for the strong interaction of hadrons with a nucleus with mass number  $A$  can be approximated as

$$\sigma_A = \pi r_A^2, \quad (7)$$

which is the geometric cross section of the nucleus, where  $r_A \approx 1.3A^{1/3}$  femtometers. (A Fermi, or femtometer, is  $10^{-15}$  meter.) This absorption cross section implies that the mean free path,  $\lambda^* = 1/n\sigma$  (where  $n$  is the number density of atoms), for a hadron in uranium of nominal density is a length of about 10 centimeters, or an interaction length of  $200 \text{ gm/cm}^2$ , an order of magnitude larger than that of high-energy x-rays. This interaction length is almost perfectly matched to hydrotest radiography. Consequently, a much lower incident flux of hadrons will produce the same statistical information now obtained from a higher flux of high-energy x-rays.

Of course, protons are charged, and therefore, when they interact with matter, the Coulomb force between the protons and the charge of the electrons and the nuclei in the material causes

the protons to continuously slow down and scatter into other directions.

However, for a proton with high enough energy, electromagnetic scattering processes produce only small changes in its direction and energy, even when it traverses a significant thickness of material. Thus, nuclear inelastic scattering remains the dominant mechanism removing protons from an incident high-energy proton beam. Consequently, high-energy protons have a large interaction length and are interesting as a radiographic probe. Moreover, current accelerators routinely produce high-intensity, short pulses of high-energy protons.

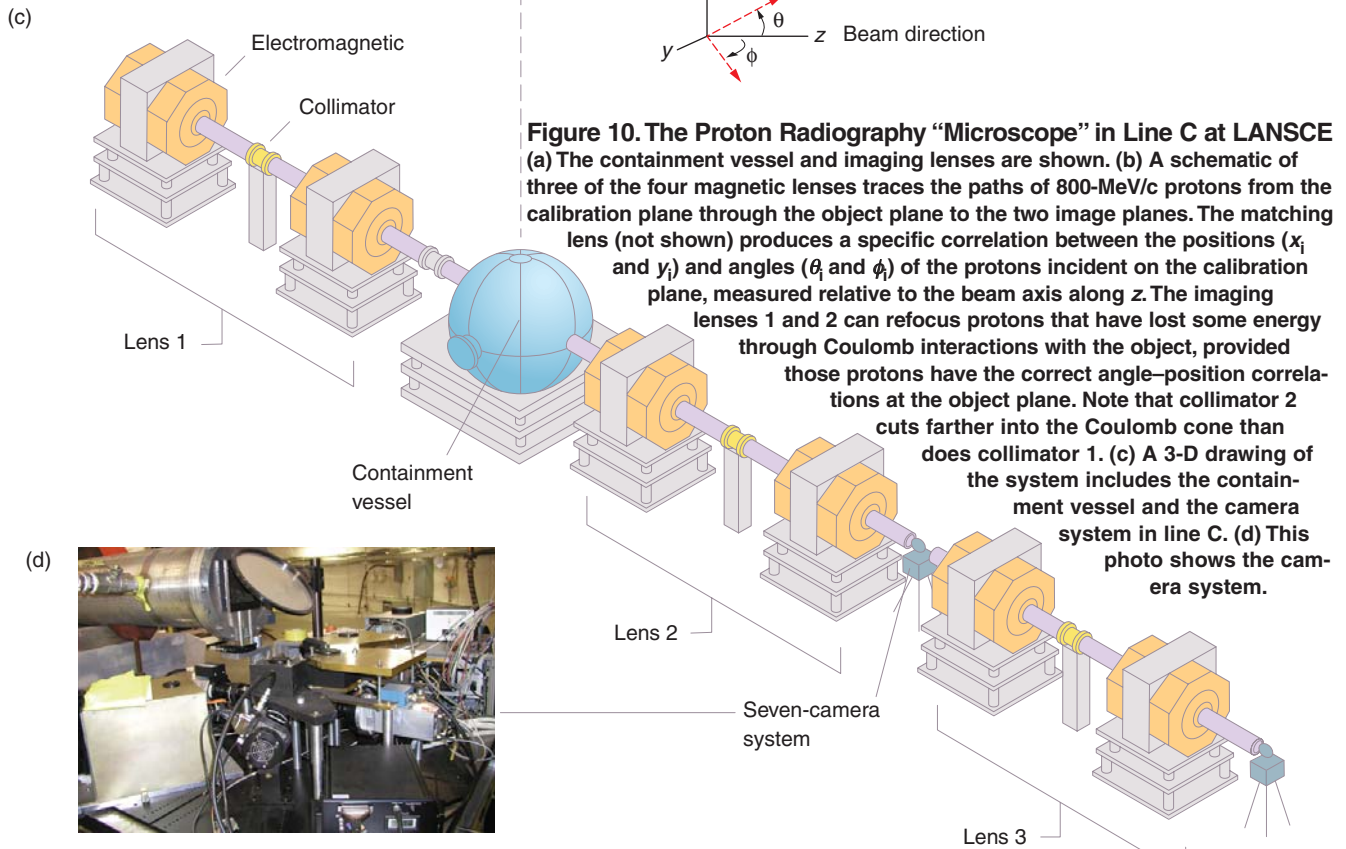
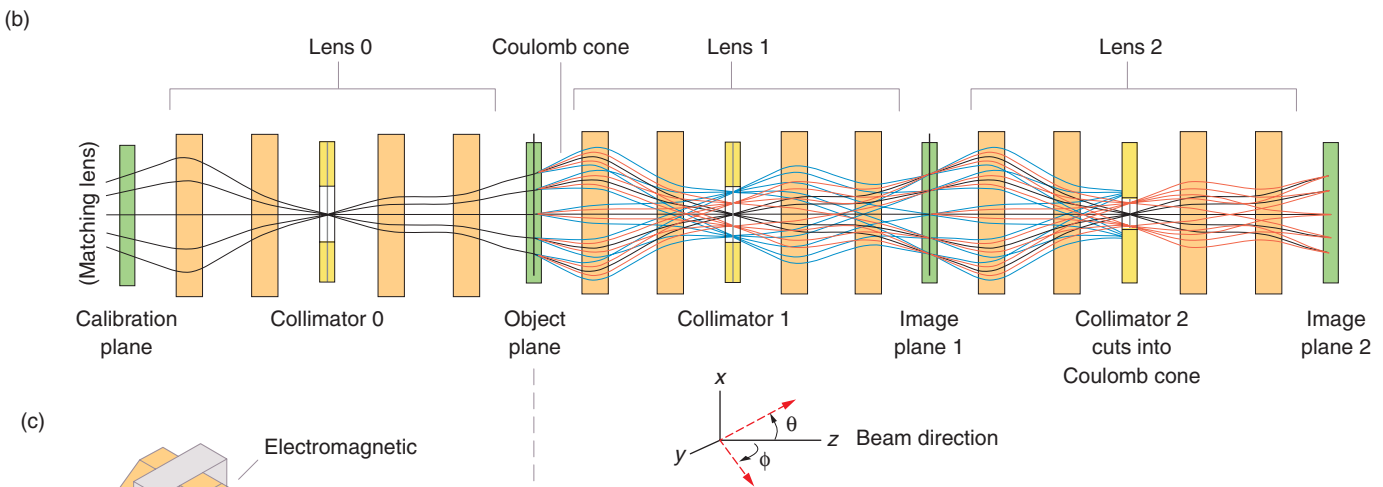
**Focusing Protons.** The charge on the proton allows using magnetic lenses to focus a proton beam of a selected energy or momentum (the two are almost equivalent at high energies, hundreds of times above the rest mass energy of the proton, which is 938.272 MeV). This focusing capability gives great flexibility to proton radiography and leads to many advantages over the standard point-source x-radiography shown in Figure 1.

Figure 10 shows the proton radiography line at LANSCE, which has four separate lenses: an angle matching lens (not shown), consisting of three quadrupole magnets, and three imaging lenses, each consisting of four quadrupole magnets. The initial beam goes through the angle-matching lens, which is tuned to focus the protons of a selected initial momentum  $P_i$  onto image plane 0 such that the protons (black rays) are spread over an area equal to that of the object and each proton's directional angles relative to the beam axis along the  $z$ -direction ( $\theta_1$  in the  $x$ - $z$  plane and  $\phi_1$  in the  $y$ - $z$  plane) are proportional to its distance from the beam axis, that is,  $\theta_1 = Ax_i$  and  $\phi_1 = -Ay_i$ . To calibrate our experiment, we measure the beam intensity at selected points on image plane 0. Lens 0 then refocuses the protons to have

the same position-angle correlation at the object plane that they have at image plane 0, just inverted from plane 0 to the object plane. Protons transmitted through the object are scattered by Coulomb forces into a cone of angles about their initial directions (represented by the red and blue rays). Even though the transmitted protons now have a spread in momentum, lens 1 can refocus the beam onto image plane 1 because the lens is designed to cancel the leading chromatic aberrations (the changes in image position caused by variations in momentum). A collimator in lens 1, like the f-stop of a conventional camera, selects the range of proton angles that can be transmitted to image plane 1. The contrast of the image can be increased by selection of a collimator that cuts into the Coulomb scattering cone.

**Correcting Chromatic Aberrations.** The largest aberration in the lens system is chromatic, or momentum dependent. That is, protons whose momentum varies from that for which the magnetic lens is tuned are the leading cause of image blurring. The angle-matching lens has been designed so that the proton trajectories incident at the object plane have a position-angle correlation that minimizes the chromatic aberration in the imaging lenses. In particular, that position-angle correlation at the object plane is such that the largest chromatic aberrations in image position cancel each other—the aberration proportional to  $\theta_1$  cancels the one proportional to  $x_i$ , and the aberration proportional to  $\phi_1$  cancels the one proportional to  $y_i$ .

**High-Efficiency Detection.** Protons are detected with high efficiency when a thin sheet of scintillator is placed at the image plane. As the protons pass through, the scintillator emits enough light for an image to be stored in a gated charge-coupled device camera, but it does not perturb



**Figure 10. The Proton Radiography “Microscope” in Line C at LANSCE**  
 (a) The containment vessel and imaging lenses are shown. (b) A schematic of three of the four magnetic lenses traces the paths of 800-MeV/c protons through the calibration plane to the two image planes. The matching lens (not shown) produces a specific correlation between the positions ( $x_i$  and  $y_i$ ) and angles ( $\theta_i$  and  $\phi_i$ ) of the protons incident on the calibration plane, measured relative to the beam axis along z. The imaging lenses 1 and 2 can refocus protons that have lost some energy through Coulomb interactions with the object, provided those protons have the correct angle–position correlations at the object plane. Note that collimator 2 cuts farther into the Coulomb cone than does collimator 1. (c) A 3-D drawing of the system includes the containment vessel and the camera system in line C. (d) This photo shows the camera system.

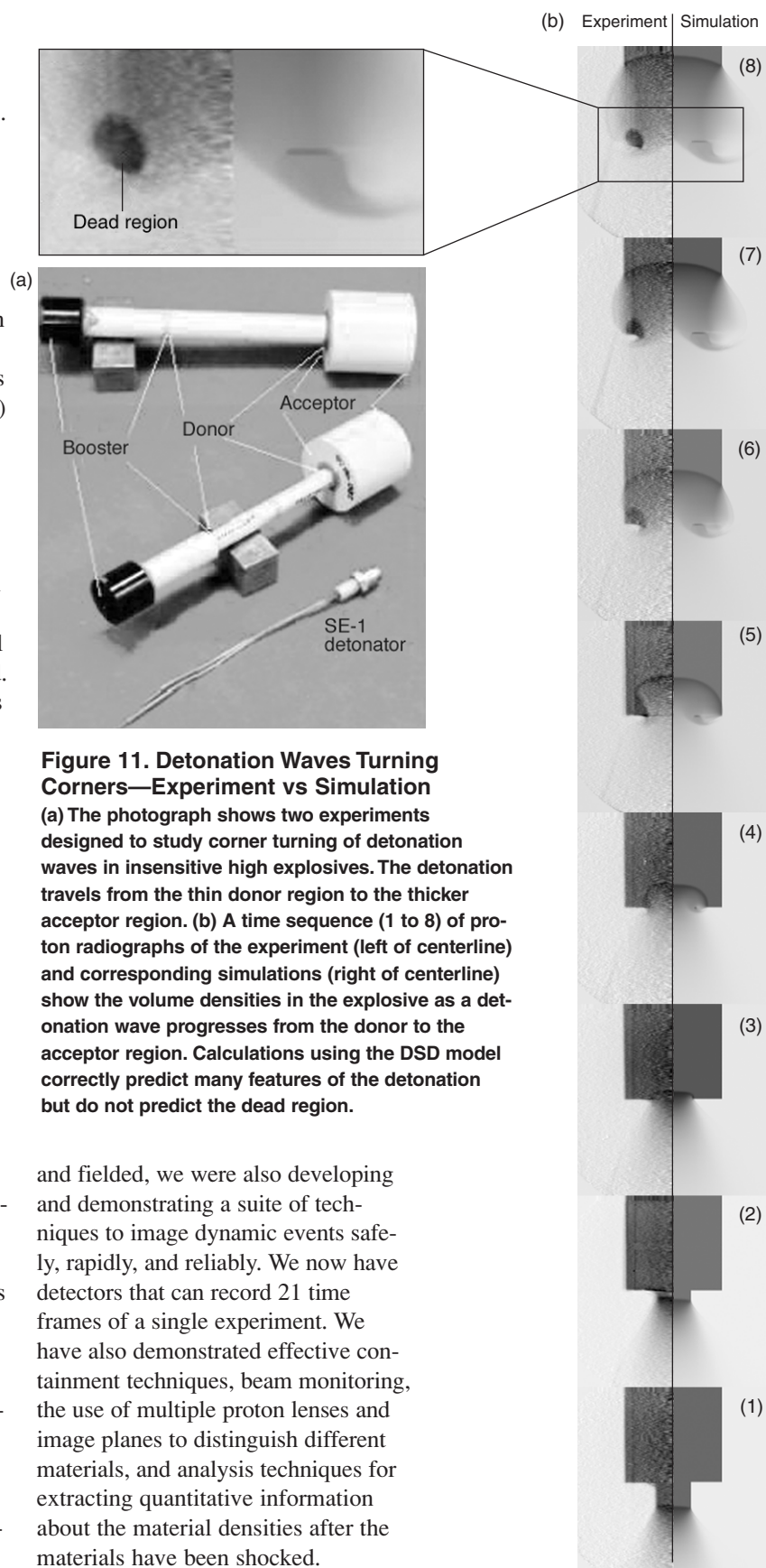
the proton beam significantly. In fact, so few protons are absorbed that the same pulse can be reimaged by lens 3. Lens 3 has a smaller collimator than lens 2 to cut farther into the Coulomb scattering cone (note that the blue rays are blocked). Since Coulomb scattering depends on the Z-number of the material, this double imaging process enables material identification from proton radiographs.

The ability to refocus protons means that the closest detector (image plane 1) can be at a long standoff distance from the blast. That characteristic, combined with the momentum selectivity of the lens, results in much lower backgrounds in proton radiography than in x-ray radiography. As a result, the need for background mitigation techniques such as graded collimation, an essential part of x-ray experiments, is eliminated. The combination of lower backgrounds and energy-independent cross sections allows a new level of precision not available in x-radiography.

Proton radiography is being studied with the 800-MeV/c proton beam provided by the LANSCE accelerator and with a 24,000-MeV beam provided by the alternating gradient synchrotron (AGS) at Brookhaven National Laboratory.

### Studies at LANSCE

On average, about 40 small-scale dynamic experiments are being performed each year at the proton “microscope” in line C at LANSCE (refer to Figure 10). Each experiment (or physics package) uses up to 10 pounds of high explosive. The physics package is contained in a vessel shown in Figure 10(c). These experiments are designed to study high-explosive detonation and dynamic material failure and to perform small integral tests to validate the models used in weapons codes. At the time that the first experiments were being planned, designed,



**Figure 11. Detonation Waves Turning Corners—Experiment vs Simulation**

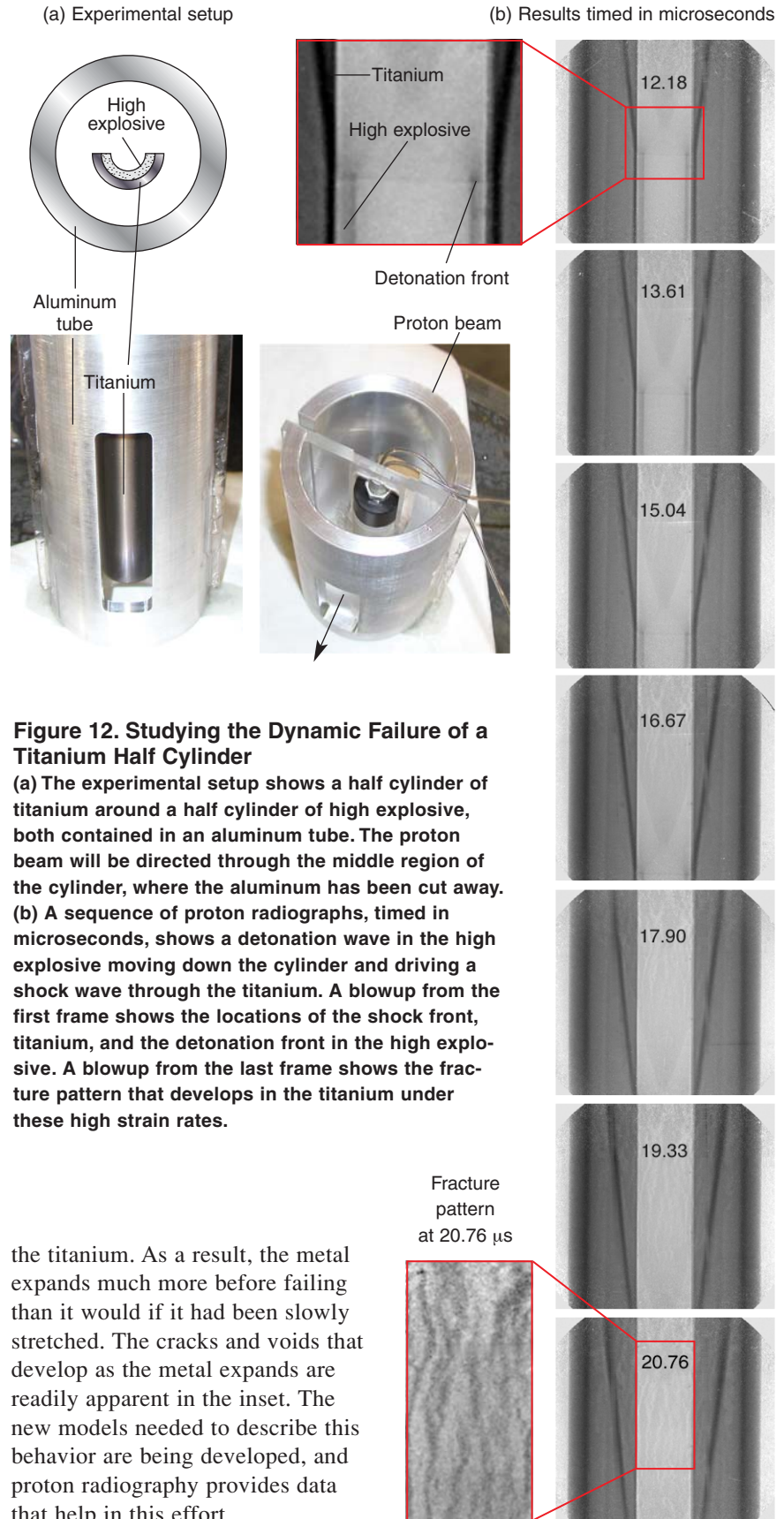
(a) The photograph shows two experiments designed to study corner turning of detonation waves in insensitive high explosives. The detonation travels from the thin donor region to the thicker acceptor region. (b) A time sequence (1 to 8) of proton radiographs of the experiment (left of centerline) and corresponding simulations (right of centerline) show the volume densities in the explosive as a detonation wave progresses from the donor to the acceptor region. Calculations using the DSD model correctly predict many features of the detonation but do not predict the dead region.

and fielded, we were also developing and demonstrating a suite of techniques to image dynamic events safely, rapidly, and reliably. We now have detectors that can record 21 time frames of a single experiment. We have also demonstrated effective containment techniques, beam monitoring, the use of multiple proton lenses and image planes to distinguish different materials, and analysis techniques for extracting quantitative information about the material densities after the materials have been shocked.

As an example, Figure 11 illustrates the experimental setup, the data, and the detonation shock dynamics (DSD) simulations of a set of experiments to study the corner turning of a detonation front in the high explosive PBX 9502. This insensitive, plastic-bonded material has been incorporated in some systems in the nuclear weapons stockpile to increase safety because it was specifically designed to be hard to detonate. The downside is that detonation waves do not propagate as well in PBX 9502 as they do in conventional high explosives. As a result, when the geometry of the explosive forces the detonation wave to turn a corner, dead zones (regions that do not detonate) appear.

During the experiment, a detonation wave is launched in a cylindrical stalk of PBX 9502 by a booster, and a sequence of proton radiographs is taken as the detonation wave propagates into a cylinder with a larger diameter. Figure 11(b) shows a frame-by-frame comparison of radiographs (left) and the corresponding simulations (right) for that sequence. The proton radiographs show that a dead region (dark) of unburnt explosive in the shape of a doughnut remains after the detonation wave expands in the larger cylinder. This phenomenon is not yet captured in the DSD model (the DSD model is presented in the article "High Explosives Performance" on page 96). The experimental data guide the development of models aimed at better predictions of this phenomenon.

The 800-MeV/c proton beam at line C has also been used to radiograph materials as they break apart, or spall, under the influence of high strain rates. Figure 12 shows proton radiographs of a half cylinder of titanium driven by the detonation of an embedded half cylinder of high explosive. The rapid propagation of the high-pressure detonation wave down the high-explosive half cylinder produces very high strain rates in



**Figure 12. Studying the Dynamic Failure of a Titanium Half Cylinder**

(a) The experimental setup shows a half cylinder of titanium around a half cylinder of high explosive, both contained in an aluminum tube. The proton beam will be directed through the middle region of the cylinder, where the aluminum has been cut away. (b) A sequence of proton radiographs, timed in microseconds, shows a detonation wave in the high explosive moving down the cylinder and driving a shock wave through the titanium. A blowup from the first frame shows the locations of the shock front, titanium, and the detonation front in the high explosive. A blowup from the last frame shows the fracture pattern that develops in the titanium under these high strain rates.

the titanium. As a result, the metal expands much more before failing than it would if it had been slowly stretched. The cracks and voids that develop as the metal expands are readily apparent in the inset. The new models needed to describe this behavior are being developed, and proton radiography provides data that help in this effort.

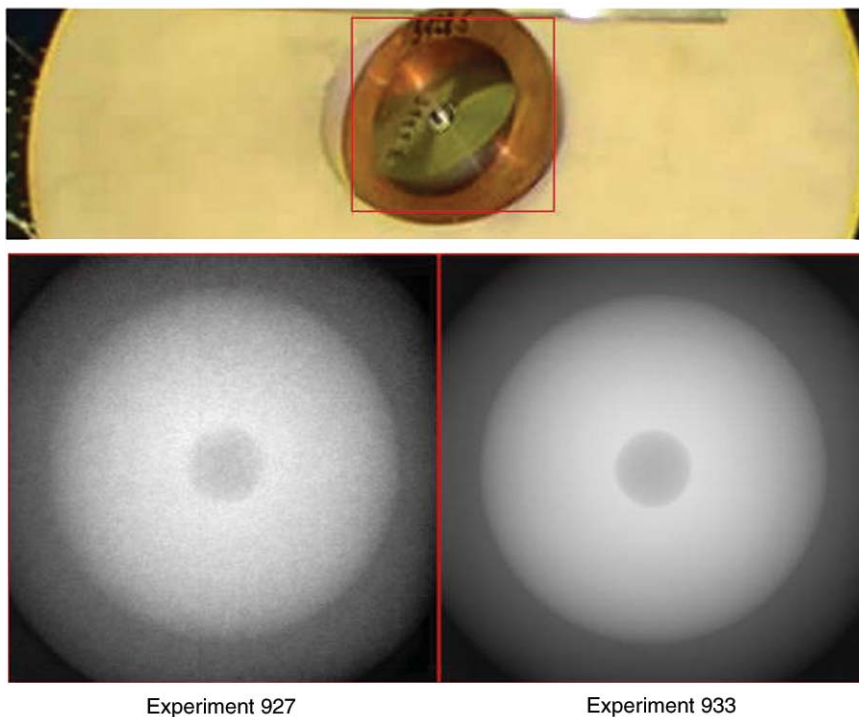
## High-Energy Proton Radiography

The experiments with the 800-MeV/c proton beam at LANSCE have been immensely valuable in demonstrating and developing proton radiography. For radiographs of the much higher areal densities involved in full-scale hydrotests, a higher-energy proton beam is required. We have conducted studies using high-energy proton beams that have exactly the format (beam size and intensity) needed to perform flash radiography for full-scale hydrotests.

Our first higher-energy experiments at AGS used secondary protons (whose energy is lower than that of the main beam), which are produced at very low rates. Although the exposures lasted for several hours, these first radiographs of the FTO, made with a quadrupole lens, showed great potential for the technique. In more recent experiments, we used a fast-extracted high-energy beam (30-nanosecond-long pulses) of up to  $10^{11}$  protons from the accelerator to radiograph various static test objects and to develop techniques for quantitative analysis of dynamic experiments.

Figure 13 shows the dramatic improvement in FTO radiographs obtained at the AGS. These experiments have demonstrated low backgrounds, multiple views, good statistics, and quantitative precision. We have also compared proton and x-ray radiography by radiographing the same thick, classified test object with the first axis of DARHT and with the high-energy proton beam at AGS. The results of this classified experiment demonstrate the dramatic improvement in the quality of radiography expected from this new probe.

Material identification has also been demonstrated for static experiments at the AGS in experiment 933. Material identification would be valuable in studying the properties of material interfaces in hydrotests, but moving objects present some new dif-



**Figure 13. Proton Radiographs of the FTO**

The radiograph on the left was recorded in a 4-h exposure in a secondary beam with about  $10^9$  protons with an energy of 10 GeV in energy. The radiograph on the right was made in a 40-ns exposure with about  $2 \times 10^{10}$  protons with an energy of 24 GeV.

ficulties that must be studied. We are developing Monte Carlo simulation codes to study the properties of the entire beam line, including the effects of test objects with complicated geometries.

Very recently, we conducted another series of experiments at the AGS on static test objects. These experiments were designed to allow assessing the quantitative accuracy of proton radiography for the study of criticality and mix, effects that are of the highest possible importance to stockpile stewardship. The test objects were carefully crafted to provide unprecedented fidelity to weapon design calculations. Some of the test objects constituted a weapon implosion “time-series,” reflecting very precisely the microsecond-timescale changes in device configuration that the design calculations predict.

Another aim of the recent experi-

ments was to demonstrate a capability to study mix and other stewardship-relevant phenomena in dynamic experiments at the AGS. Although the recent experiments used only static test objects, they were designed to pave the way for future experiments that would incorporate high explosives and be fielded in containment vessels. Even though a far cry from weapons hydrotests, these dynamic experiments would both provide unique data on hydrodynamic performance needed for the stewardship program and demonstrate that this type of imaging is fully compatible with the technology and infrastructure needed in the hydrotest regime—including, of course, processes and procedures required for the protection of the environment, personal safety, and health.

The success of proton radiography



has led Los Alamos to propose a new facility for hydrodynamic testing. The new proton-radiography facility will be used to make detailed quantitative movies of hydrotests that capture with unprecedented precision the time development of an implosion.

## Summary

Penetrating flash radiography has provided critical information to weapons designers since the inception of the Manhattan Project. Radiographic machines used during that time provided images of the outer pit surface to calibrate numerical models of the hydrodynamic performance of the device. Radiography was also used to identify several important early problems with the implosion device—for example, interaction of the high-explosive waves that caused jetting of the heavy metal. PHERMEX, commissioned nearly 40 years ago, provided the first radiographic machine capable of obtaining detailed data on the density distributions at the center of a primary in a radiographic hydrotest. In the intervening 40 years, there has been tremendous progress in x-ray machine and detector performance, scatter reduction, and quantitative analysis of flash x-ray radiography for stockpile stewardship. The second axis of DARHT, soon to be commissioned, will complete a state-of-the-art facility that will provide weapons designers with their clearest views of the inside of a hydrotest ever obtained.

Both PHERMEX and DARHT represented quantum leaps forward in our ability to peer inside implosions of mock nuclear weapon systems. However, the ultimate goal to produce highly quantitative, 3-D, time-evolving density maps, needed for certification without testing, has still not been reached. In the future, it is

likely that new radiographic machines with improved performance will be needed to certify the enduring stockpile. The recent invention of proton radiography at Los Alamos has the potential to meet this future need. Experiments have shown that proton radiography can provide high-quality radiographic information at many times during dynamic experiments using the 800-MeV/c proton beam from LANSCE. Experiments performed with the higher energy 24,000-MeV/c proton beam from the AGS accelerator have demonstrated low backgrounds, small statistical errors, and well-controlled systematic uncertainties. The combination of high-quality radiography, small-scale experimentation, and predictive modeling can form the foundation of a robust stockpile stewardship program without underground testing. Interestingly, more primitive versions of the same tools formed the design basis for the successful Trinity test in 1945. ■

## Acknowledgments

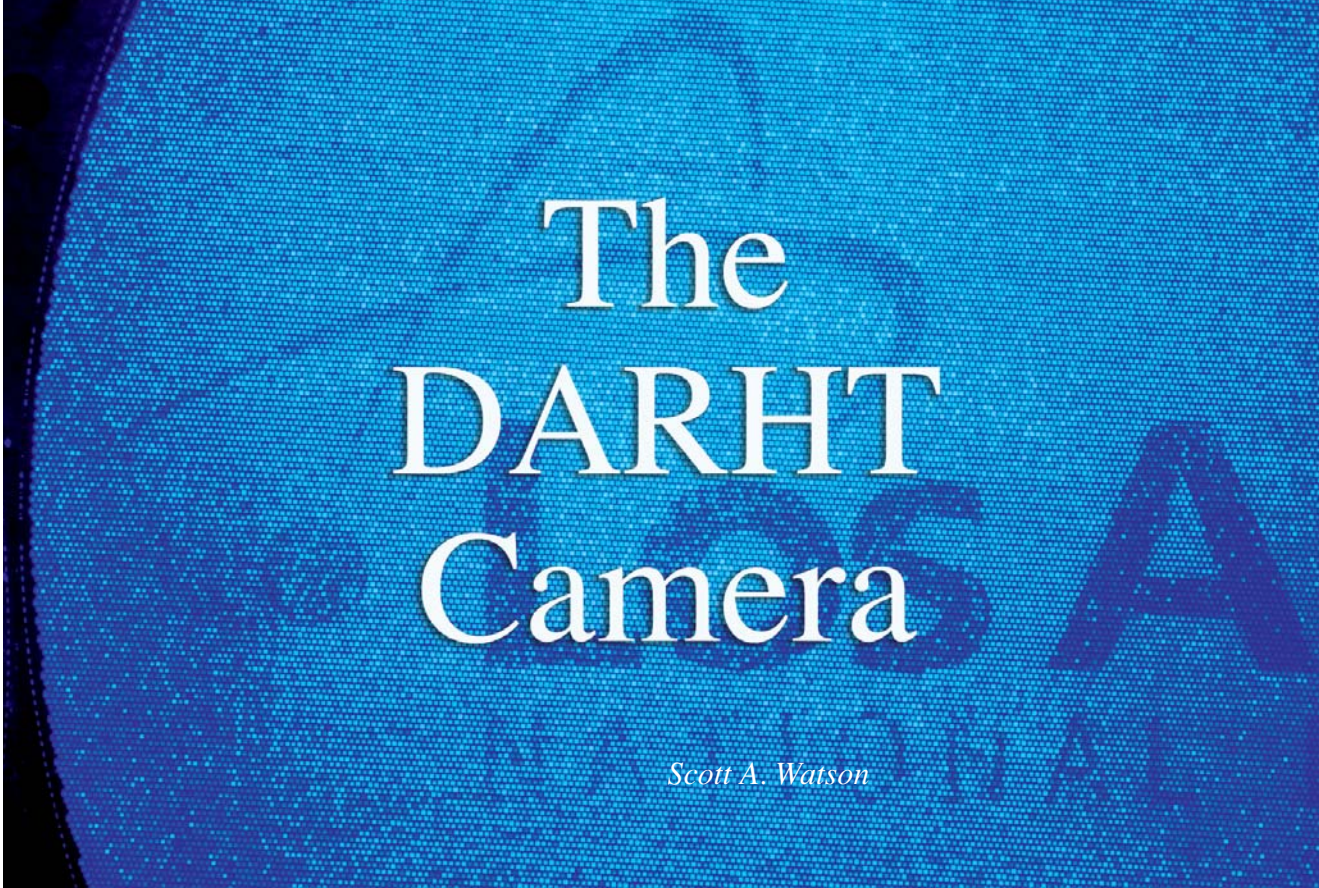
We are grateful for the time Doug Venable, John Hopson, and Tim Neal spent with us describing the history of PHERMEX and DARHT. Susan Seestrom's early editing was a great help in allowing us to get to a final manuscript. The Reports Library, Copy Center, and photo archive were invaluable resources for delving into the early history of flash radiography at the Laboratory. We would also like to thank Tim Neal, John Zumbro, Stephen Sterbenz, and Ken Hanson for contributing materials and providing helpful advice.

## Further Reading

- Anthouard, P., J. Bardy, C. Bonnafond, P. Delsart, A. Devin, P. Eyharts et al. 1998. AIRIX Prototype Technological Results at CESTA. In *Proceedings of the 1997 Particle Accelerator Conference*. (Vancouver, B. C., Canada, May 12–16, 1997). Vol. 1, p. 1254. Edited by M. Comyn, M. K. Craddock, M. Reiserr, and J. Thomson. Piscataway, New Jersey: IEEE.
- Alrick, K. R., K. L. Buescher, D. J. Clark, C. J. Espinoza, J. J. Gomez, N. T. Gray et al. 2000. "Some Preliminary Results From Experiment 933." Los Alamos National Laboratory document LA-UR-00-4796.
- Beer, A. 1852. Bestimmung der Absorption des Rothen Lichts in Farbigen Flüssigkeiten. *Ann. Phys. Chem.* **86**: 78.
- Boyd, T. J. 1967. "PHERMEX: A Pulsed High-Energy Radiographic Machine Emitting X-Rays." Los Alamos Scientific Laboratory report LA-3241.
- Burns, M. J., B. E. Carlston, T. J. T. Kwan, D. C. Moir, D. S. Prono, S. A. Watson et al. 1999. DARHT Accelerators Update and Plans for Initial Operation. In *Proceedings of the IEEE Particle Accelerator Conference*. (The 18th Biennial Particle Accelerator Conference, New York, March 27–April 2, 1999). Vol. 1, p. 617. Piscataway, New Jersey: IEEE.
- Christofilos, N. C., R. E. Hester, W. A. S. Lamb, D. D. Reagan, W. A. Sherwood, and R. E. Wright. 1964. High-Current Linear-Induction Accelerator for Electrons. *Rev. Sci. Instrum.* **35** (7): 886.
- Ekdahl, C. 2002. Modern Electron Accelerators for Radiography. *IEEE Trans. Plasma Sci.* **30** (1): 254.
- Eyharts, P., P. Anthouard, J. Bardy, C. Bonnafond, P. Delsart, A. Devin et al. 1998. Beam Transport and Characterization on AIRIX Prototype at CESTA. In *Proceedings of the 1997 Particle Accelerator Conference*. (Vancouver, B. C., Canada, May 12–16, 1997). Vol. 1, p. 1257. Edited by M. Comyn, M. K. Craddock, M. Reiserr, and J. Thomson. Piscataway, New Jersey: IEEE.
- Ferm, E. N., C. L. Morris, J. P. Quintana, P. Pazuchanics, H.L. Stacy, J. D. Zumbro et al. 2002. Proton Radiography Examination of Unburned Regions in PBX 9502 Corner Turning Experiments. *AIP Conference Proceedings* **620** (1): 966.
- Goldsack, T. J., T. F. Bryant, P. F. Beech, S. G. Clough, G. M. Cooper, R. Davitt, and R. D. Edwards. 2002. Multimegavolt Multiaxis High-Resolution Flash X-Ray Source Development for a New Hydrodynamics Research Facility at AWE Aldermaston. *IEEE Trans. Plasma Sci.* **30** (1): 239.

- Hawkins, D. 1961 (written 1946). "Los Alamos Scientific Laboratory—Manhattan District History, Project Y, the Los Alamos Project." Los Alamos Scientific Laboratory report LAMS-2532 (Vol. I).
- Hoddeson, L., P. W. Henriksen, R. A. Meade, and C. Westfall. 1993. *Critical Assembly—A Technical History of Los Alamos during the Oppenheimer Years, 1943–1945*. Cambridge: Cambridge University Press.
- Kwan, T. J. T., C. M. Snell, and P. J. Christenson. 2000. Electron Beam–Target Interaction and Spot Size Stabilization in Flash X-Ray Radiography. *Phys. Plasmas* **7** (5): 2215.
- Mader, C. L., T. R. Neal, and R. D. Dick. 1980. *LASL PHERMEX Data*, Vol. I. Berkeley: University of California Press.
- Mueller, K. H. 1985. Collimation Techniques for Dense Object Flash Radiography. *Proc. Soc. of Photo-Optical Instrumentation Engineers* **491** (1) :130.
- Morris, C. L. et al. Defense Review LACP-02-89 (to be published 2003).
- Neddermeyer, S., and A. R. Sayer. 1946. "Betatron Cloud-Chamber Studies with Levitated Assemblies." Los Alamos Scientific Laboratory report LA-483.
- Röntgen, W. C. 1896. On a New Kind of Rays. *Nature* **53**: 274.
- Venable, D. 1964. PHERMEX. *Phys. Today* **17** (12):19.
- Watson, S., T. Kauppila, L. Morrison, and C. Vecere. 1997. The Pulsed High-Energy Radiographic Machine Emitting X-Rays (PHERMEX) Flash Radiographic Camera. In *SPIE 22nd International Congress on High-Speed Photography and Photonics*. (International Congress on High Speed Photography and Photonics, Santa Fe, New Mexico, October 27–November 1, 1996). Edited by D. L. Paisley, and A. M. Frank. Vol. 2869, p. 920. Bellingham, Washington: SPIE.

*For further information, contact  
Gregory Cunningham (505) 667-2562  
(cunning@lanl.gov).*



# The DARHT Camera

*Scott A. Watson*

For almost 40 years, scientists in the weapons program at Los Alamos have x-rayed, or radiographed, implosions (hydrotests) using the giant PHERMEX (for pulsed high-energy radiographic machine emitting x-rays), which generated a single, brief flash of x-rays that were then recorded on film. Early on, they recognized that the design community really wanted an x-ray movie to better understand the implosion process. The value was obvious: One picture returns position; two pictures, velocity; three pictures, acceleration, and so on. Furthermore, because a movie records multiple images of a single hydrotest, the desired information could be gathered at a reduced cost. The limitation was the x-ray film.

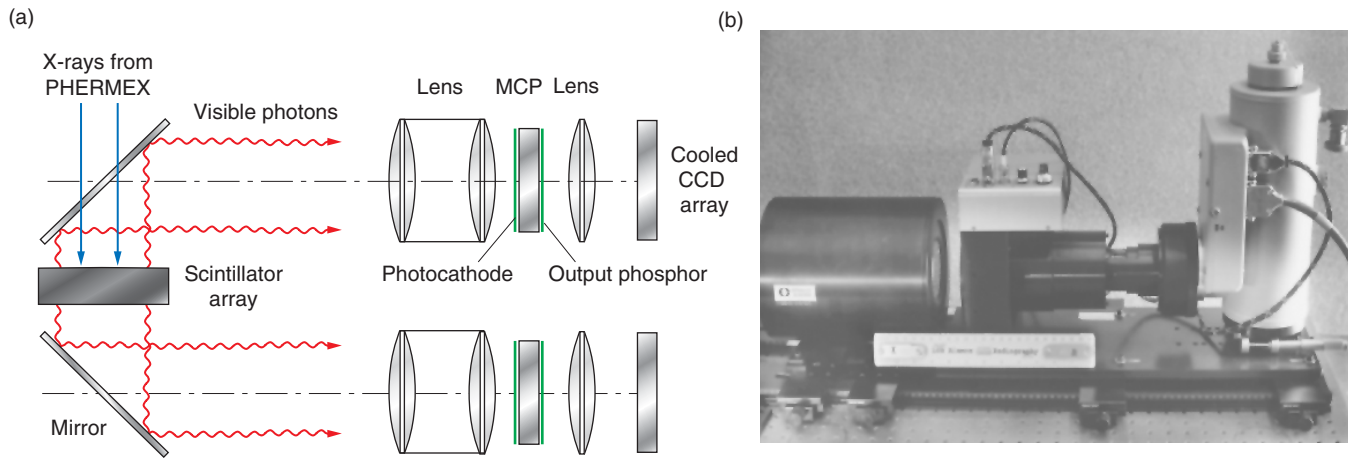
Film has been used to record x-ray images since the discovery of the x-ray, but despite over a century of development, x-ray film still suffers

from certain drawbacks. It is relatively transparent to x-ray photons—especially those at higher energies—that often pass right through without imprinting any information. In addition, film is essentially an analog recording medium with limited sensitivity; that is, it must be exposed to a minimum amount of light before an image can be recorded. Normally, for a movie, separate images are recorded on separate pieces of film. Because x-rays cannot be focused or reflected like visible light, no conventional technology existed to perform this task. Simply put, film cannot be advanced fast enough to capture the extremely rapid explosions.

Interestingly, a filmless system was proposed during the design phase of PHERMEX by Doug Venable and Ralph Stevens: “The PHERMEX detection system will consist of a mosaic of scintillation detectors that

will view pulses of . . . radiation through systems of interest . . .” (Stevens 1959). The scintillator would absorb the x-rays and convert them to visible light, which could record a limited number of channels electronically. Berlyn Brixner and the late Fred Doremire then expanded on the original concept with proposals for a high-speed electronic camera that had the potential of returning multiple radiographs for each experiment. Unfortunately, these ideas were ahead of their time; it took another 30 years for technology to catch up with this initial vision.

First used in 1996, the PHERMEX x-ray camera (Watson et al. 1995) takes just two pictures—hardly a movie. Still, it was a solid-state, all-electronic system with no film, which demonstrated higher sensitivity and absorbed more x-rays (that is, it had higher “quantum efficiency”) than

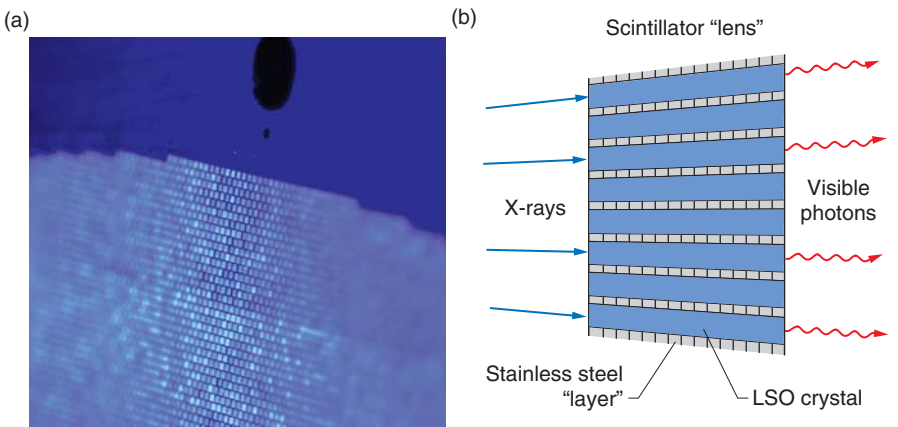


**Figure 1. PHERMEX Two-Frame Camera System**

(a) X-rays coming from the target are converted into visible photons by the scintillator. Photons emerging from the front of the scintillator follow one optical path and create one radiograph, while those emerging from the back create the second radiograph. The microchannel plate (MCP) in each pathway is the crucial electronic “shutter.” The MCP photocathode converts the photons into electrons (which are then converted back into photons by the output phosphor). By changing the voltage on the MCP, we can rapidly stop the flow of electrons and thus prevent any light from reaching the cooled CCD detector. Appropriate timing of the two MCP voltages allows us to take consecutive radiographs. (b) This photo is of the camera system. (c) The two radiographs of H-1970, a VIPER shape-charge munition, were taken 17  $\mu$ s (left) and 21  $\mu$ s (right) after detonation. These are the first Los Alamos radiographs showing an explosive event at different times.

film (see Figure 1). The large image format allowed us to see the entire imploding pit, and the increased sensitivity allowed us to see through dense materials for the first time. This revolutionary system changed forever the way we think about hydrotesting and, indeed, stockpile stewardship.

As modern hydrotesting facilities such as the Dual-Axis Radiographic Hydrotest (DARHT) come online, x-ray camera technology continues to advance significantly with highly optimized components. In particular, the “scintillator” has evolved into a large mosaic of inlaid crystals—much akin to Zuni jewelry, but with up to 350,000 pieces. Long (more than 40 millimeters) square rods of very dense (greater than 7 grams per cubic centimeter) scintillator crystals are used to facilitate the x-ray absorption process. Exotic manmade crystals such as  $\text{Lu}_2\text{SiO}_5:\text{Ce}$  (LSO) are also used because they exhibit a



**Figure 2. The DARHT Scintillator Lens**

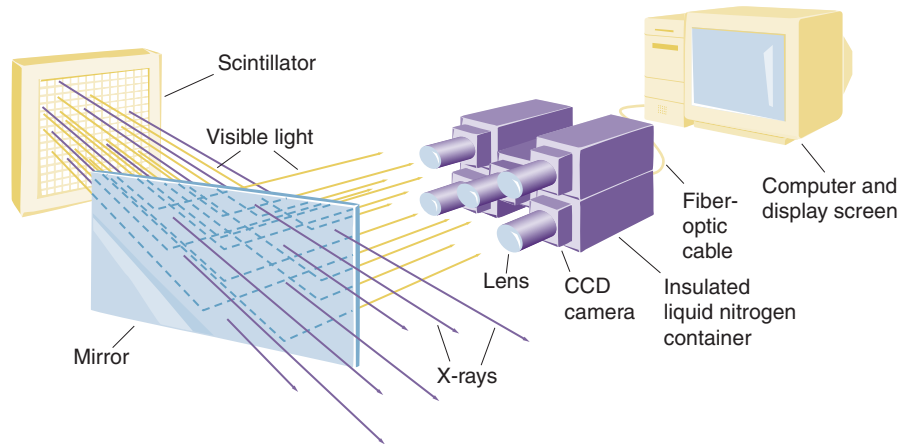
(a) The LSO inlaid scintillator shown here has more than 135,000 focused pixel elements. The blue color is a result of the natural emission spectrum of LSO, which peaks around 420 nm. (b) This schematic shows how the pixels are held in place to form the mosaic. The pixel pitch is 1.1 mm (1.0 mm LSO and 0.1 mm stainless steel).

rapid (50 nanoseconds) phosphorescent decay between x-ray flashes so that light from one image does not corrupt its neighbors in the movie sequence. These crystals are then

assembled into the mosaic by means of stack lamination constructed from hundreds of layers of photochemically etched stainless steel (see Figure 2). Because this special inlay technique

**Figure 3. The DARHT First-Axis Camera**

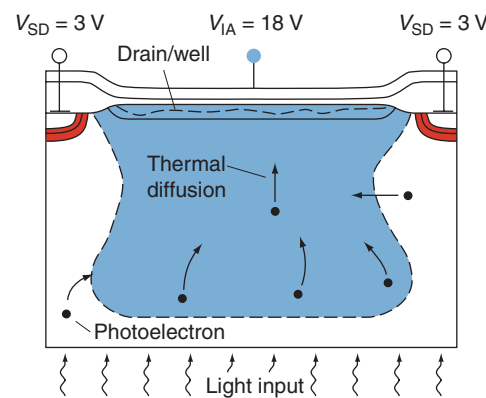
The camera consists of the scintillator, lens, and five optical lens/CCD systems for capturing the scintillator light. The multiple cameras, with overlapping fields of view, allow us to image the entire scintillator with less than 1% geometric distortion.



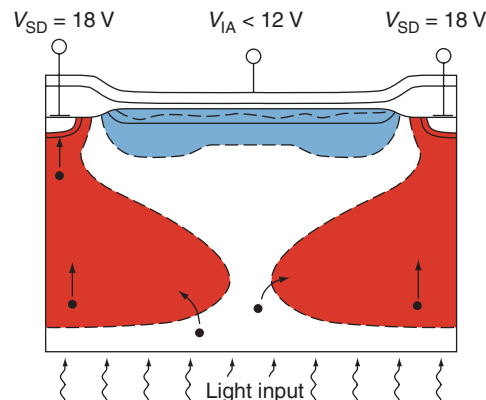
**Figure 4. Multistage CCD Pixel**

A CCD pixel can be thought of as a bathtub, complete with a faucet (DARHT) and drain, that collects the photoelectrons produced when light strikes the surface of the silicon pixel. (a) Thermal diffusion guides the photoelectrons to a “drain” region, where a local electric field captures the photoelectrons in a potential well that is ultimately connected to the readout electronics. The number of photoelectrons produced is proportional to the number of photons striking the pixel. (b) Reversing the bias on the electrodes prevents the photoelectrons from reaching the collection drain. Thus, we can shutter the pixel and control the light signal collected from that drain. (c) For the DARHT second-axis camera, each pixel is actually a superpixel with four separate drains and four storage wells. Each drain region has its own electrodes, which allow us to open a “hole” in the bathtub over any selected well region. To capture the first frame (i), drain A is opened whereas the other drains are closed. All the photoelectrons generated in the entire superpixel region are collected by well A. Thus, the device exhibits a 100% fill factor, giving increased sensitivity. After the first image is stored, we close drain A and open drain B to collect charge in region B for the second image (ii). This procedure is continued until all four frames are collected. The charge from each region is then read out slowly (to minimize noise in the charge amplifier), bucket-brigade fashion from pixel to pixel as in a conventional CCD.

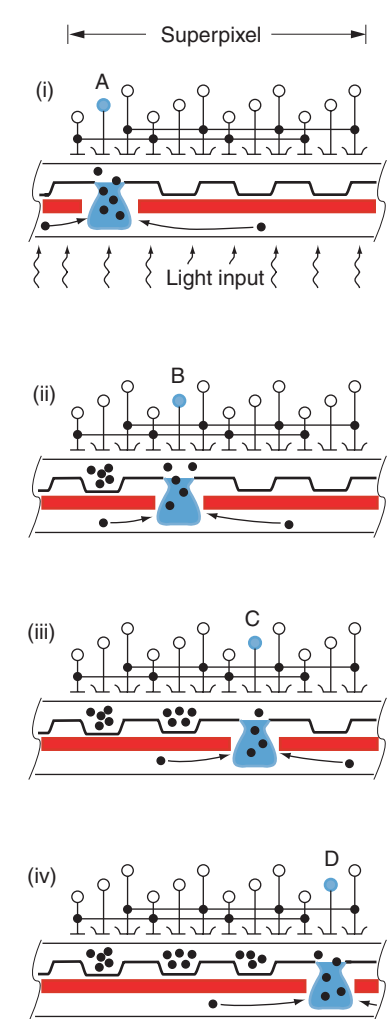
(a) Shutter Open

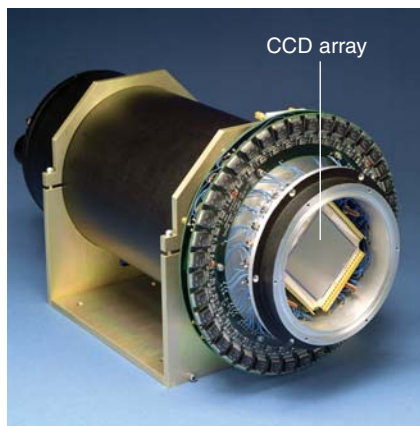


(b) Shutter Closed



(c) Four-Frame Capture





**Figure 5. Cooled CCD Array System**

This photo shows a CCD array of 512 by 512 pixels and its electronics collar, mounted on a liquid nitrogen Dewar (black cylinder).

allows each rod to point directly to the x-ray source, the scintillator exhibits no parallax blur despite the long pixels used to construct it.

Converting the x-rays into a more useful visible light signal is only one challenge. In photography, the required sensitivity normally increases with higher frame rates, but unfortunately, the available sensitivity normally decreases with higher frame rates, and the net difference is made up with bright movie lights. In our case, the movie light is DARHT, which cannot be made much brighter, so we must construct an extremely sensitive detector.

To construct the detector, we employ a number of tricks. We use a custom f1.0 lens to collect as much of the scintillator light as possible and focus that light on the largest, most sensitive optical recording devices available, namely, astronomy-grade charge-coupled devices (CCDs), which are much like those on the Hubble Space Telescope. Even this combination is not sensitive enough, so we must use multiple cameras in a mosaic, as Figure 3 shows, and cool the CCDs with liquid nitrogen to reduce electronic noise to the level of a few electrons. At this point, we

have a remarkable camera system, which is easily 100 times more sensitive than film and 40 times more efficient at absorbing x-rays. This system is now routinely used on the DARHT first axis (Watson et al. 2000).

To obtain multiple images, we employ a unique CCD architecture jointly developed by Los Alamos and Massachusetts Institute of Technology Lincoln Laboratories specifically for the DARHT second axis (Reich et al. 2003). This chip architecture retains the large format, low noise, and high sensitivity of astronomy-grade CCDs but also records four images at a rate of two million frames per second. Because there is insufficient time to transfer data off the chip at this high frame rate, the information for each frame must be stored locally on each pixel and then slowly read off when the explosive experiment is over (see Figures 4 and 5).

The next-generation camera (Watson et al. 2003) will employ a technology in which the scintillator light is collected by an avalanche photodiode, amplified, and then pipelined into a dedicated high-speed digitizer for every pixel. Although this approach requires a larger, more complex electronics package, the enhanced performance should be astounding. Whereas the PHERMEX camera can take two radiographs at 500 kilohertz and the DARHT camera can take four radiographs at 2 megahertz, the next-generation camera will take thousands of pictures at 20 megahertz. We hope that the advanced camera will generate useful results for the weapons community in a timely manner. ■

## Acknowledgments

This work was the result of an extensive decade-long collaboration. The author thanks the Radiographic Detector Team from the Hydrodynamics Group (Debra

Archuleta, Steve Balzer, Chris Gossein, Mark Hoverson, Henry Olivas, Mike Ulibarri, Carl Vecere, and Chuck Vecere), Massachusetts Institute of Technology Lincoln Laboratories, Princeton Instruments, Bicon, Tecomet, and Spindler Hoyer for significant contributions.

## Further Reading

- Reich, R. K., D. D. Rathman, D. M. O'Mara, D. J. Young, A. H. Loomis, E. J. Kohler et al. 2003. High-Speed, Electronically Shuttered Solid-State Imager Technology. *Rev. Sci. Instrum.* **74** (3): 2027.
- Stevens, R. R. 1959. "An Investigation of the Statistics Inherent in the Detection of Small Numbers of X-Ray Quanta in the PHERMEX System." Los Alamos National Laboratory memorandum GMX-11-TM-141.
- Watson, S. A., T. J. Kauppila, K. H. Mueller, and R. C. Haight. 1995. "Multiframe, High-Energy, Radiographic Cameras for Submicrosecond Imaging." Los Alamos National Laboratory document LA-UR-95-3570.
- Watson, S. A., C. A. Ekdahl Jr., S. J. Balzer, H. A. Bender, and A. Daiz. 2003. "Reliable, Low-Current, Continuous Cavity Imaging at DARHT." Los Alamos National Laboratory document LA-UR-03-0908.
- Watson, S. A., J. M. Gonzales, C. Gossein, M. Hoverson, and M. Ulibarri. 2000. "Quantum Efficiency, Noise Power Spectrum, Linearity and Sensitivity of the DARHT g-Ray Camera." Los Alamos National Laboratory document LA-UR-00-0653.

*For further information, contact Scott Watson (505) 665-6233 (scottw@lanl.gov).*

# High-Explosives Performance

## *Understanding the effects of a finite-length reaction zone*

John B. Dzil

with Tariq D. Aslam, Rudolph Henninger, and James J. Quirk

**H**igh explosives—explosives with very high energy density—are used to drive the implosion of the primary in a nuclear weapon. That circumstance demands precision in the action of the high explosive. To predict with high accuracy the course of energy release under various conditions is therefore an important problem that we face in certifying the safety, reliability, and performance of nuclear weapons in the stockpile. Here we survey our progress on the problem of explosives performance: predicting the outcome of intentional detonation of high explosives in complex three-dimensional (3-D) geometries. The problems of safety (accidental initiation) and reliability (reproducible response to a prescribed stimulus) are also under investigation but will be only briefly mentioned here.

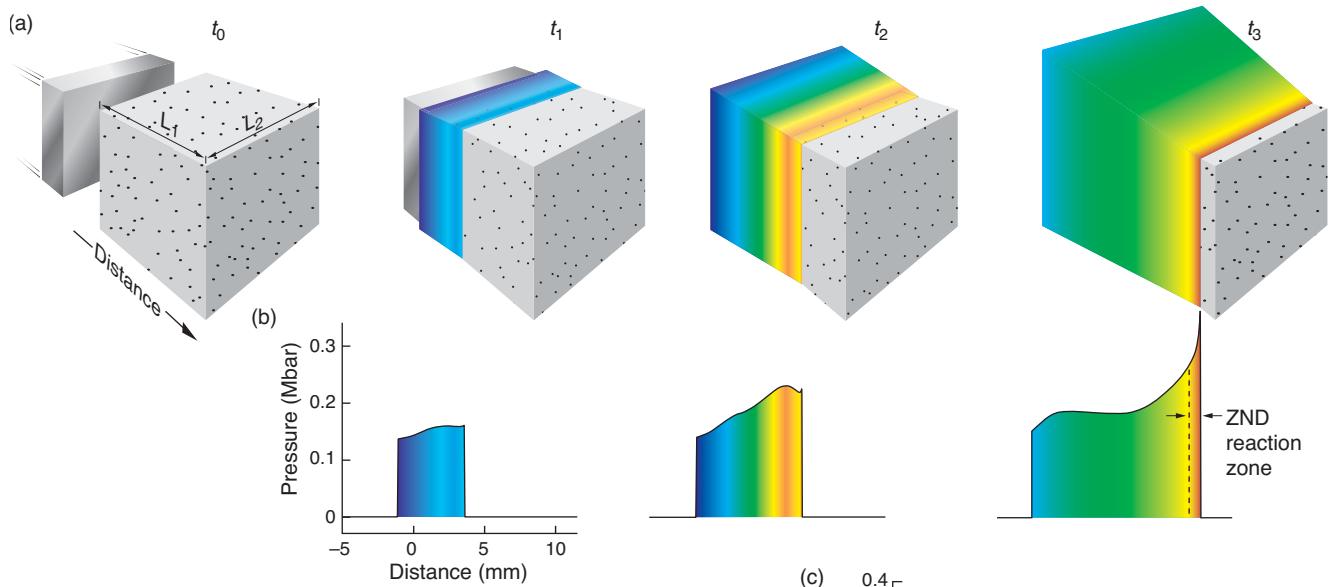
Explosives belong to the class of combustibles known as energetic materials, which means that they contain both fuel and oxidizer premixed on a molecular level. Such materials can support a whole range of combus-

tion, including ordinary combustion such as that in a match head. Ordinary combustion is a coupled physico-chemical process in which the interface separating fresh from burnt energetic material travels as a wave through the sample. Exothermic chemical reactions begin on the surface of the match head and burn the outer layer of material. The heat released is transferred through thermal conduction to an adjacent unreacted layer until that second layer ignites, and this layer-by-layer process continues until the entire sample is consumed. The speed of the combustion wave is relatively low, depending on both the rate of energy transport from one layer to the next and the rate of the local exothermic chemical reactions in each layer.

Explosives, in contrast, support very high speed combustion known as detonation. Like an ordinary combustion wave, a detonation wave derives its energy from the chemical reactions in the material, but the energy transport occurs not by thermal conduction but rather by a high-speed compres-

sion, or shock, wave. The high-pressure detonation wave streaks through the material at supersonic speeds, turning the material into high-pressure, high-temperature gaseous products that can do mechanical work at an awesome rate. Figure 1 shows the initiation of a detonation wave from shock compression through the formation of a self-sustaining Zeldovich–von Neumann–Döring (ZND) detonation reaction zone behind the shock. The power delivered by an explosive

*The pressure plot on this opening page shows a steady-state detonation wave propagating through a cylindrical explosive (gray) confined by a low-density inert material (yellow). Red is the highest pressure; purple, the lowest. The reaction starts along the shock (red curve) and ends along the sonic surface (white curve). A large pressure drop at the edge of the explosive leads to a significant lengthening of the chemical reaction zone near the edges of the detonating explosive, a reduction in the speed of the detonation wave, and the development of a curved detonation shock front.*



### Figure 1. Initiation and Propagation of a ZND Detonation Wave

(a) A schematic 1-D (planar) experiment is shown at different times. In the experiment, the impact of a plate thrown on one face of a cube of explosive ( $t = t_0$ ) produces a planar shock wave ( $t = t_1$ ) that gradually accelerates ( $t = t_2$ ) to a steady-state detonation ( $t = t_3$ ) as the shock sweeps through the explosive and causes chemical energy to be released to the flow at a finite rate. (b) The corresponding pressure-vs-distance snapshots show the evolution of an essentially inert shock wave at  $t = t_1$  growing into a classical 1-D ZND detonation structure at  $t = t_3$ , namely, a shock, or pressure, discontinuity at the ZND point followed by decreasing pressure through the reaction zone, ending at the CJ point, the pressure predicted by the simple CJ model (see text). (c) Pressure-vs-time plots for material particles originally at the shock front locations in (b) show the particle pressure (or velocity) histories in the form measured in actual experiments (see Figures 5, 6, and 7). Only at the location of the right-most particle has a ZND detonation fully formed. Note: The point of maximum acceleration of the shock, called the point of detonation formation, coincides with the shape change in the pressure profile and the first appearance of a choked flow condition (sonic condition). Refer to Figure 3.

depends on its energy density and its detonation wave speed. Solid high explosives, like those used in nuclear weapons, have a detonation speed of about 8000 meters per second (m/s), or three times the speed of sound in the explosive, a high liberated energy density of about 5 megajoules per kilogram (MJ/kg), and an initial material density of about 2000 kilograms per cubic meter ( $\text{kg}/\text{m}^3$ ). The product of these three quantities yields the enormous power density of 80,000,000 MJ/m<sup>2</sup>/s or  $8 \times 10^9$  watts per centimeter squared ( $\text{W}/\text{cm}^2$ ). By comparison, a detonation with a surface area of 100 centimeters squared operates at a power level equal to the

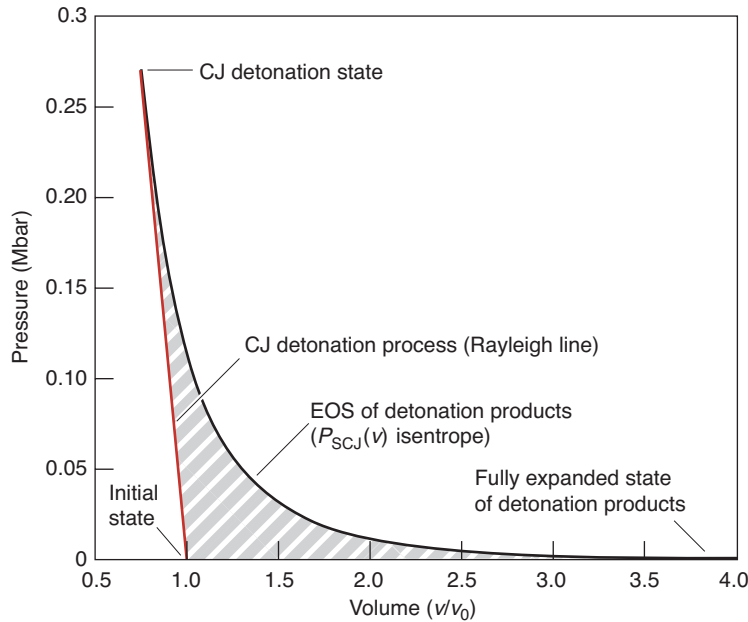
total electric generating capacity of the United States! This very rapid rate of energy liberation is what makes solid explosives unique and useful.

The legacy weapons codes have long used the simple Chapman-Jouguet (CJ) model to compute the performance of high explosives. In this classical, one-dimensional (1-D) model of detonation, it is assumed that the chemical reaction rate is infinite (and therefore the length of the reaction zone is zero rather than finite, as in the opening figure and Figure 1). That assumption leads to the prediction that the detonation speed is constant. Moreover, the values of the detonation speed,  $D_{CJ}$ , as well as the det-

onation pressure,  $P_{CJ}$ , are independent of the initiating shock strength and depend on only certain properties of the explosive before and after passage of the detonation front, namely, the initial density of the unreacted material, the liberated energy density of the explosive, and the pressure-volume ( $P$ - $v$ ) response function of the reacted material (called the mechanical equation of state, or EOS). In this CJ limit, the explosive performance problem is reduced to providing an accurate mechanical EOS for the gaseous products of detonation,  $E_g(P, v)$ —see Figure 2.

In this article, we focus on another aspect of the performance problem:





**Figure 2. Maximum Work Obtainable from a CJ Detonation**

The  $E_g(P, v)$  mechanical equation of state (EOS) is required to model explosive performance. Most often, it is measured along a restricted curve—an isentrope,  $P_S(v)$ , or a shock Hugoniot curve,  $P_H(v)$ —in the state space defined by the thermodynamic variables  $E_g$ ,  $P$ , and  $v$ . To characterize the maximum work that a detonation can perform (shaded area above), we need determine only the principal or CJ expansion isentrope of the detonation products,  $P_{SCJ}(v)$ , given that we know  $D_{CJ}$  and  $P_{CJ}$ . The two curves shown above are the detonation Rayleigh line, shown in red (detonation process), and the detonation products expansion isentrope,  $P_{SCJ}(v)$ . The area under the isentrope (to some cut-off pressure) minus the area under the Rayleigh line (work done by the shock in compressing the explosive) is the maximum mechanical work that can be obtained from the explosive. For our high-performance, monomolecular explosives, such as HMX, this work compared with the available explosive energy can be very high (more than 90%). We perform experiments to measure this isentrope and then construct the  $E_g(P, v)$  mechanical EOS for the products of detonation, which is an essential ingredient in every model of how detonations do work on their surroundings. We are working on both better theoretical (Shaw 2002) and experimental (Hill 2002) methods for determining the  $E_g(P, v)$  EOS.

creating accurate 3-D detonation models that account for the effects of finite chemical reaction rates (and therefore a finite reaction-zone length behind the detonation front). The finite length of the reaction zone has many effects. For example, it can affect the detonation speed and therefore the power level at which a detonation engine operates on inert materials. It also places limits on the minimum size of the explosive and the minimum input pressure that will lead to detonation, especially in

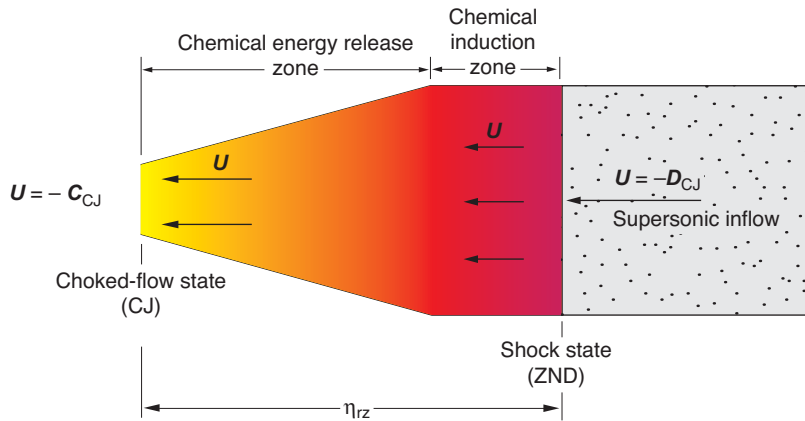
geometries that cause detonation waves to go around corners, say, near a small detonator. The models we have been developing are specifically designed for adaptation to the legacy codes and to the Advanced Simulation and Computing (ASCI) high-fidelity codes used to study weapons performance. Known as detonation shock dynamics (DSD), these are subscale (or subgrid) models that capture the physics of the reaction zone without explicitly modeling that zone and, therefore, without requiring enormous

computing time. Although they are state of the art for modeling 3-D detonation flows, our models predict detonation propagation only in homogeneous explosives under standard conditions. That is, they do not fully account for the effects that the heterogeneity of the real explosives we use today have on detonation. We therefore conclude this article with our vision for the future of detonation propagation modeling—one that accounts for that heterogeneity yet remains practical for weapons performance studies.

## The Detonation Process

How does a detonation wave reach and then maintain such enormous power levels as it sweeps through the explosive? The enormous pressures (a few hundred thousand atmospheres, or a few hundred kilobars) and temperatures (2000 to 4000 kelvins) behind the detonation front originate from the very rapid release of chemical energy. Reactions are 90 percent complete in less than a millionth of a second. As a result of this rapid release, the reaction zone is very short. But how are the pressures sustained?

As shown in Figure 3, the reaction zone is bounded by two surfaces that isolate it from the regions ahead and behind it and thereby maintain its extreme pressure. First is the shock surface, which initiates the reaction. Because it travels at supersonic speed relative to the unreacted material, it prevents any leakage of pressure ahead of the shock. Second is the sonic surface (labeled choked-flow state), which moves at the local speed of sound in the frame of the moving shock front. To explain the effect of this surface, we consider an observer riding with the shock and looking back. The observer sees an increasing amount of energy release back into the reaction zone as a func-



**Figure 3. The Finite-Length, Self-Sustaining Reaction Zone**

In many respects, the self-sustaining detonation reaction zone operates like a rocket motor. The reaction zone is bounded by the shock surface at the detonation front and the choked flow-state surface some distance behind. Those two surfaces isolate the reaction zone from the regions in front of it and behind it and thereby maintain its extreme pressure. Looking backward in the frame that moves with the detonation shock front, one observes that an increasing amount of heat added to the flow with increasing distance into the reaction zone acts like a nozzle in a rocket, accelerating the flow to sonic speeds,  $C_{CJ}$ .

tion of distance. This energy release serves to accelerate the flow away from the shock front and reduce the pressure, in much the same way as a rocket nozzle accelerates the gas ejected from a rocket and thereby propels the rocket forward. As the reaction is completed, the flow speed at the end of the reaction zone becomes equal to the local speed of sound in the frame moving with the shock,  $C_{CJ}$ . As a result, the flow becomes choked and thereby stops any further pressure decrease in the reaction zone. Collectively, these two effects are referred to as inertial confinement.

Another way to understand inertial confinement at the sonic surface is to note that the postreaction-zone flow (left of the sonic surface) in the reference frame of the shock is supersonic. Consequently, the reaction zone is essentially isolated from disturbances originating in the flow behind it. Insulated from its surroundings, detonation is self-propagating, depending only on what is happening

in the reaction zone.

### Real vs Idealized Explosives

If an explosive is to be useful in engineering applications—be they mining, nuclear weapons, or modern “smart” munitions—its chemical reaction rate must be essentially zero at the ambient state and must become extremely fast once passage of a shock wave substantially increases the pressure and temperature in the material. As mentioned above, in the classical CJ model, the chemical reaction rate after the shock front has passed is infinite, the reaction-zone length goes to zero, and the detonation wave travels through the material at a constant speed and pressure. In reality, the explosives we use in practical applications do not behave like the ideal CJ model but have finite reaction rates. This situation is indeed fortunate. If the reaction rate were infinite and the reaction zone of length zero, then subjecting even a tiny region of the explosive to a high pressure or high temperature would initiate detonation

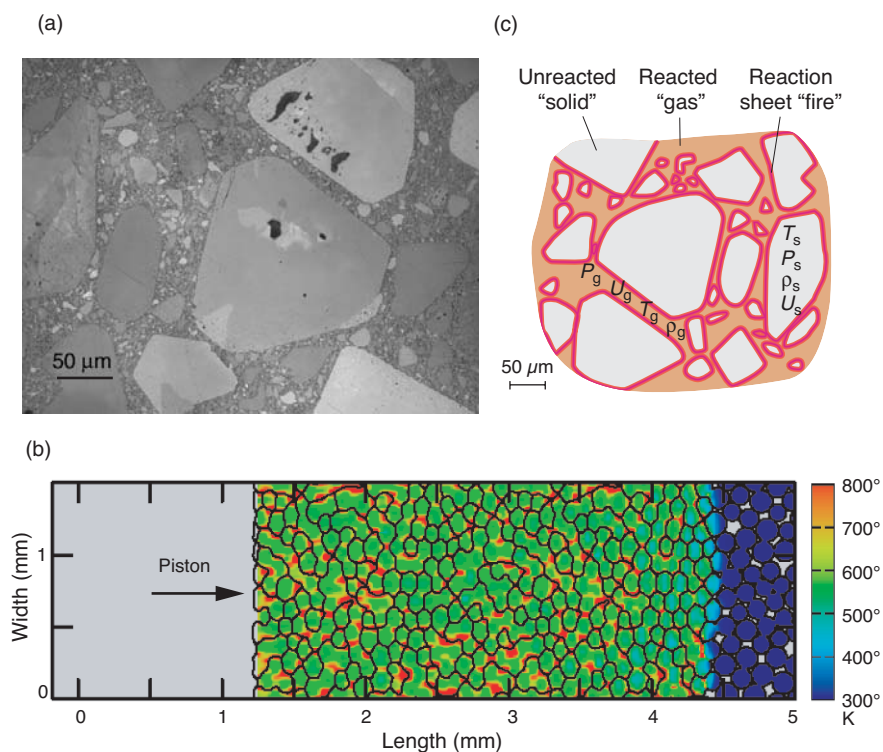
of the entire sample. The extreme sensitivity of explosives such as nitroglycerine is legendary in this regard.

Because the reaction rate and reaction-zone length of practical explosives depend significantly on pressure and temperature, a sample subjected to a weak initial shock will experience transients during initiation of detonation. If the sample is a slab of finite thickness ( $L_1$  in Figure 1) but infinite lateral extent ( $L_2 \rightarrow \infty$ ), the shock can pass through the slab in a short time compared with the duration of the transient, and no detonation occurs.

Conversely, to initiate detonation in a sample of finite thickness with finite reaction-zone length, the shock must have a finite strength. That decrease in sensitivity caused by a finite reaction-zone length is what makes practical explosives safe enough to handle.

The fact that real explosive samples have a finite lateral extent ( $L_2$  in Figure 1) also contributes to reducing sensitivity. Some of the energy released in the reaction zone leaks out of the sides of the explosive transverse to the direction of detonation propagation and thereby reduces the support for the forward motion of the shock. If that energy loss is too great, detonation dies out. Thus, the longer the reaction zone in practical explosives, the more difficult they are to detonate—or, in other words, the more insensitive (and safer) they are.

One way to control sensitivity is to control the “effective,” or global, reaction rate as opposed to the local reaction rates. Alfred Nobel used this technique to turn the liquid explosive nitroglycerine into dynamite. Nitroglycerine is an extremely sensitive explosive because its high viscosity allows it to form bubbles easily. When these bubbles collapse, they generate localized high pressures and temperatures called hotspots. The hotspots serve as initiation sites for localized, rapid reaction, leading to the establishment of



**Figure 4. Substructure of Heterogeneous High Explosives**  
 (a) The photomicrograph shows the granular substructure of PBX 9501 (Skidmore et al. 1998). (b) A numerical simulation shows the temperature distribution that develops in a heterogeneous material subjected to rapid, compressive loading (Menikoff and Kober 1999). (c) The drawing shows a detailed view of the hotspots that develop when explosives such as PBX 9501 are subjected to a shock wave. We are not yet able to accurately model such complex, micromechanical hydrodynamic interactions.

localized detonation that spreads through the otherwise cool material and consumes it all. By adding a highly porous silica to nitroglycerine, Nobel turned the material into a paste, thereby suppressing small bubble formation and dramatically reducing its sensitivity.

At Los Alamos, we have followed the reverse path. We start from a very insensitive explosive and increase its sensitivity by using it in the form of small granules that serve as centers for initiation of chemical reaction hotspots and subsequent detonation. A typical example of these insensitive, high-mass, high-energy-density solid explosives is HMX. To detonate a single crystal of this material, a few centimeters on a side and free of most physical defects, requires an input shock pressure of hundreds of kilobars

(Campbell and Travis 1985). To increase the effective, average global reaction rate, we formulate a mixture of small, heterogeneous HMX granules and polymeric binder and then press the mixture to a density approaching that of pure, crystalline HMX. By controlling the size of the granules and the final pressed density, we can vary the sensitivity of the explosive. The granular HMX explosive PBX 9501 requires only tens of kilobars of pressure to initiate detonation within a fraction of a centimeter.

Despite our control over the manufacturing and thus the reproducibility of explosive detonation, we cannot predict the effective reaction rates in such heterogeneous HMX explosives from first principles. One reason is that we have been unable to measure the chemical route by which the solid

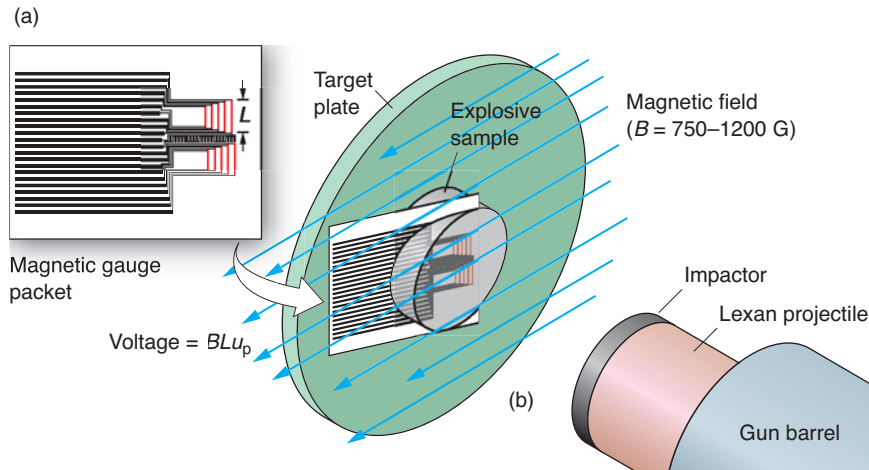
explosives decompose to gaseous products under the extreme conditions of detonation (about 0.5 megabar in pressure at temperatures of 3000 kelvins). Only recently did we acquire appropriate techniques to address those questions. In particular, we can now generate and characterize planar shocks using a combination of ultrafast lasers and interferometers. In the future, we hope to use ultrafast laser spectroscopy to observe, in real time, the chemistry behind those laser-generated shocks (McGrane et al. 2003).

We also have little understanding of how the fine-scale substructures and hotspots in the detonation reaction zone affect detonation initiation and propagation. Figure 4(a) shows a photomicrograph of the granular substructure of PBX 9501. Research on the complex, micromechanical, hydrodynamic interactions that develop when such a material is subjected to a shock wave is still in its infancy—see Figures 4(b) and 4(c).

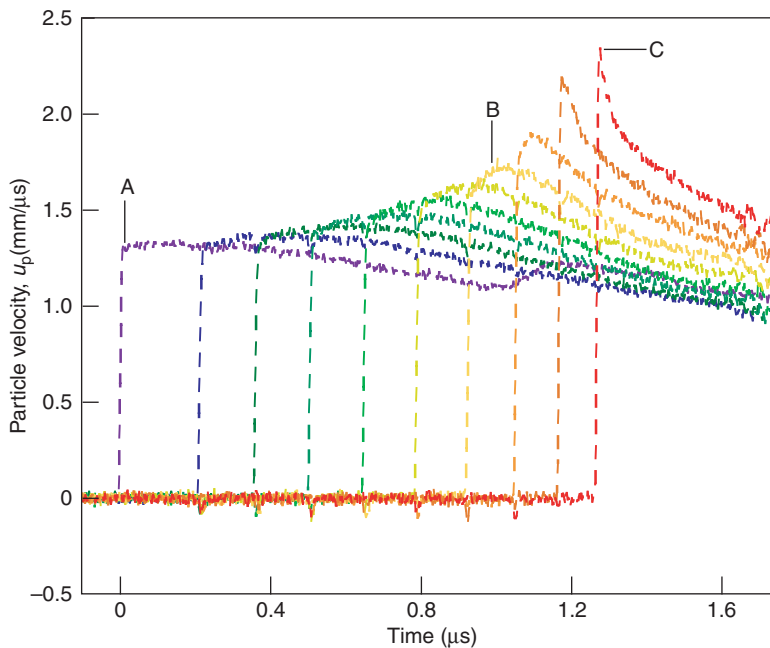
## Measuring Reaction-Zone Effects

Because the length of the effective reaction zone affects the sensitivity to initiation as well as the detonation speed and extinction rates, we would like to predict its size. Since we cannot predict the reaction scale *ab initio*, we have taken a more phenomenological approach. That is, we have performed macroscale continuum experiments to measure the effects of the reaction-zone length, and we have developed continuum theories and models that, when forced to match those measurements, allow us to infer the global reaction-zone lengths and reaction rates.

The experiments are done on samples whose dimensions run from a few to many centimeters. Some experiments subject the explosive sample to



**Figure 5. Magnetic Gauge Measurement of 1-D Detonation Flows**  
 (a) A nested set of 10 magnetic gauges made of thin conducting wires is embedded obliquely relative to the faces of an explosive sample, and the sample is placed in a magnetic field of strength  $B$ . (b) The impact of a planar projectile initiates detonation. When the active elements in the gauge package, of length  $L$  and shown in red, are moved by the explosive flow, they cut the magnetic flux lines, thereby producing a voltage directly proportional to the velocity of the flow,  $u_p$ , at each gauge location. Thus, the 10 active elements track the history of 10 different particles in the explosive. The gauge package has a thickness of  $60\ \mu\text{m}$  and is capable of a time resolution of  $20\ \text{ns}$  (Sheffield et al. 1999).



**Figure 6. Magnetic Gauge Particle Histories for 1-D Detonation Flow**  
 Shown here are the results from a magnetic gauge experiment on the insensitive high explosive PBX 9502. The input pressure was  $0.135\ \text{Mbar}$ . The experimental traces follow the transformation of a planar shock wave into a detonation. (Point A is the input state, point B indicates an interior velocity maximum, and point C is the ZND point. The shape change in the particle velocity profile (from an interior velocity maximum to a maximum at the shock) coincides with the first appearance of a choked flow condition, or sonic condition (Sheffield et al. 1998).

1-D detonation hydrodynamic flows (planar impacts in simple explosive geometries), whereas others generate and measure fully 3-D flows (resulting mostly from complex explosive geometries). Although the reaction-zone length can be quite short—about  $0.01\ \text{millimeter}$  for some of the sensitive explosives—its effects can be detected because detonation hydrodynamics tends to amplify any changes in initial or boundary conditions. (This property can be seen in Figure 1, where the transients occur over a distance of many reaction-zone lengths and later in Figure 10, where the overall displacement of the multidimensional detonation shock is measured in a number of reaction-zone lengths.) Still, the experiments must be capable of nanosecond time resolution in order to characterize the hydrodynamic response of these explosives.

In the high-resolution 1-D experiment shown in Figure 5, a nested set of 10 magnetic velocity gauges made of thin conducting wires is embedded obliquely, relative to the faces of an explosive sample, and the sample is placed in a magnetic field (Sheffield et al. 1999). A planar projectile impacts one of its faces as shown, and the quiescent sample begins to react and ultimately detonates. The active elements in the gauge package, shown in red, maintain their shape as they move with the explosive flow in the direction of the detonation front. As they move, they cut through the magnetic flux lines, producing a voltage directly proportional to the velocity of the flow at each gauge location. Thus, the 10 active elements track the history of 10 different particles in the explosive. The gauge package has a thickness of  $60\ \mu\text{m}$  and is capable of a time resolution of  $20\ \text{nanoseconds}$ . Note that, because the multiple-gauge package is mounted obliquely to the principal flow direction, the gauges farther upstream are not perturbed by those

downstream.

Figure 6 shows 10 particle-velocity histories, one from each magnetic gauge, for the insensitive high explosive PBX 9502 subjected to a planar impact with an initial pressure of 0.135 megabar (Sheffield et al. 1998). The series of particle histories (from left to right) reflects the transformation of a planar shock wave supported by the energy from the projectile impact into detonation supported by the energy release in the explosive. The gauges nearest the impact surface (leftmost trace) record the progress of what is essentially an inert shock wave passing through the explosive, whereas the later gauges show what, at least at first glance, resembles a classical ZND detonation structure (a shock followed by decreasing particle velocity through the reaction zone, as in Figure 1). The shock speed, also recorded in these experiments, shows an initial constant-velocity shock that then accelerates rapidly to a new, higher speed (approaching the detonation speed). The point of maximum acceleration, called the point of detonation formation, coincides with the shape change in the particle-velocity profile and the first appearance of the condition of choked flow (sonic condition).

### Modeling the Detonation Reaction Zone

To infer more specific information on the global heat-release rate and the detonation reaction-zone length from these and other hydrodynamic measurements, we must model the detonation process. We first discuss the standard modeling paradigm. By comparing its predictions with experiment, we show that it can model 1-D flows fairly well but has serious shortcomings when applied to 3-D flows. Finally, we show how we have altered the standard paradigm to create the

DSD model that not only solves some of those shortcomings but also is computationally efficient and suitable for use in the ASCI codes.

In the standard models, a detonating explosive is described as a continuous medium that obeys the conservation of mass, momentum, and energy for an Euler fluid:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \quad (1)$$

$$\frac{\partial \rho \mathbf{u}}{\partial t} + \nabla \cdot (\rho \mathbf{u} \mathbf{u} + \mathbf{I}P) = 0, \quad (2)$$

$$\frac{\partial \rho e}{\partial t} + \nabla \cdot [(\rho e + P)\mathbf{u}] = 0, \quad (3)$$

where  $\mathbf{I}$  is the identity matrix,  $\mathbf{u}$  is the particle velocity in the laboratory frame,  $P$  is the pressure,  $\rho = v^{-1}$  is the density,  $e = E + \mathbf{u} \cdot \mathbf{u}/2$ , and  $E$  is the specific internal energy of the reacting explosive as a function of density and pressure. The energy  $E$  as a function of pressure and specific volume,  $E(P, v)$ , is the particular constitutive law (a law determined solely by the intrinsic properties of the material) that we introduced earlier as the mechanical EOS, and it must be provided as input to the fluid equations.

Because the much used Chapman-Jouguet theory requires as input a mechanical EOS of the form,  $E_g(P_g, v_g)$ , where the subscript g denotes detonation product gas, some realizations of  $E_g(P_g, v_g)$  are available. To obtain an analogous expression for unreacted solid explosive,  $E_s(P_s, v_s)$ , one can start from a simple Mie-Gruneisen EOS and calibrate it to replicate the measured jump-off (shock state) value of the particle velocity seen with the magnetic gauges (as in Figure 6) and the measured shock velocity. To construct a mechanical EOS for the reacting mixture of solid and gas, one typically assumes pressure equilibrium between the solid and the gas,  $P = P_s = P_g$ . Then, to interpolate between the

equations of state for the gas and the solid, one assumes that the internal energy and density of the mixture are given by

$$E(P, v, \lambda) = (1 - \lambda)E_s(P, v_s) + \lambda E_g(P, v_g), \quad (4)$$

and

$$1/\rho = (1 - \lambda)v_s + \lambda v_g, \quad (5)$$

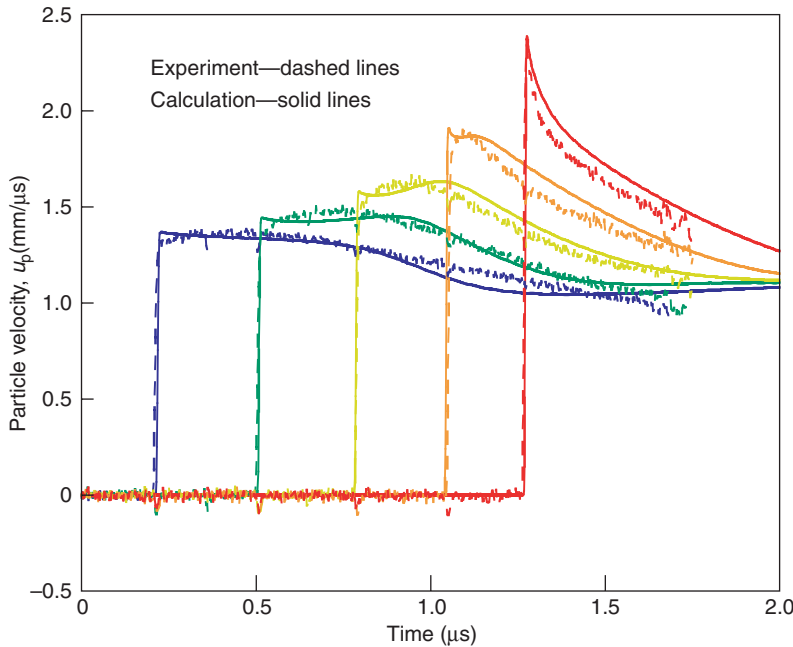
where  $\lambda$  is the mass fraction of reaction product gases.

Closure is brought to this system of equations, namely, Equations (1)–(5), with two additional assumptions. First, by extending the mechanical equations of state to include a simple temperature dependence and then assuming that the temperatures of the two phases are equal,  $T = T_s = T_g$ , one can relate  $v_s$  and  $v_g$  and thereby eliminate these intermediate variables from the problem. Second, one assumes that the rate of conversion of solid to gas in the reaction zone is given by an average, effective global heat-release rate law of the form

$$\frac{\partial \lambda}{\partial t} + \mathbf{u} \cdot \nabla \lambda = R(P, v, \lambda), \quad (6)$$

where the heat-release rate function  $R(P, v, \lambda)$  is constituted so that the gauge data in Figure 6 and other rate-dependent data are reproduced.

The Lee-Tarver Ignition and Growth model (Tarver and McGuire 2002) is an example of the standard modeling paradigm. It uses an empirical EOS for each of the components and takes appropriate account of the detonation energy in the unreacted explosive. It also uses an empirical form for the global heat-release rate function in Equation (6). The sets of constants for the equations of state of each explosive have been calibrated to a suite of hydrodynamic experiments performed on each explosive (see the box on the opposite page).



**Figure 7. Comparison of Model with Experiment for 1-D Detonation**  
Predictions of the standard modeling paradigm are compared with the measured particle histories for PBX 9502 shown in Figure 6. The calculations were done using the Lee-Tarver EOS and a recalibrated rate law. The computing mesh had a minimum of 1000 computational zones in the reaction zone. The agreement is reasonably good, but there are noticeable discrepancies.

### EOS and Rate Law for the Ignition and Growth Model

The empirical Jones-Wilkins-Lee forms are used for both the solid ( $i = s$ ) and the gas ( $i = g$ )

$$\text{EOS} \quad T = \frac{V_i}{\omega_i C_{Vi}} \{ P - A_i \exp(-R_{1i} V_i) - B_i \exp(-R_{2i} V_i) \},$$

where  $V_i = \rho_0 / \rho_i$  and  $\omega_i$ ,  $C_{Vi}$ ,  $A_i$ ,  $B_i$ ,  $R_{1i}$ , and  $R_{2i}$  are calibration parameters.

The internal energy is solved for using the thermodynamic constraint

$$\left\{ P - \rho^2 \left( \frac{\partial e}{\partial \rho} \right)_P \right\} \left\{ \left( \frac{\partial T}{\partial P} \right)_\rho + \rho^2 \left( \frac{\partial e}{\partial P} \right)_\rho \left( \frac{\partial T}{\partial \rho} \right)_P \right\} = T.$$

The heat-release rate law is given by

$$\text{Rate Law} \quad R(P, v, \lambda) = IH(\lambda_1^* - \lambda)(1 - \lambda)(1/V - 1 - a)^7 \\ + G_1 H(\lambda_2^* - \lambda)(1 - \lambda) \lambda^{0.111} P + G_2 H(\lambda - \lambda_3^*)(1 - \lambda) \lambda P^3,$$

where  $0 \leq \lambda \leq 1$  describes the progress of the global heat-release reaction ( $\lambda = 0$ , is unreacted, and  $\lambda = 1$  is fully reacted),  $H(\lambda_i^* - \lambda)$  is the unit step function, and  $I$ ,  $a$ ,  $G_1$ ,  $G_2$ , and  $\lambda_i^*$  are parameters.

The authors and Ashwani Kapila of Rensselaer Polytechnic Institute (private communication—2003), used the Lee-Tarver Ignition and Growth model to predict the results of the magnetic gauge experiment for PBX 9502 shown in Figure 6. To solve Equations (1)–(6), these authors and others have developed solution algorithms and adaptive mesh refinement codes (Aslam 2003, Fedkiw et al. 1999, Quirk 1998, Henshaw and Schwendeman 2003). Here, we used a minimum of 1000 computational zones to model the reaction zone. Figure 7 compares the simulation results with the experimental data for PBX 9502. The wave profile that develops far from the initiating piston surface (that is, to the far right) clearly shows a nearly steady-state reaction zone. We see that this phenomenological model—a simple homogeneous fluid model with a global reaction rate—mimics a 1-D experiment reasonably well. However, it does not describe the complicated interaction between hydrodynamic hotspots and fundamental chemical processes, an interaction that produces the heat release rate in effective, global, heterogeneous explosives.

### Application of Standard Models to Multidimensional Flows

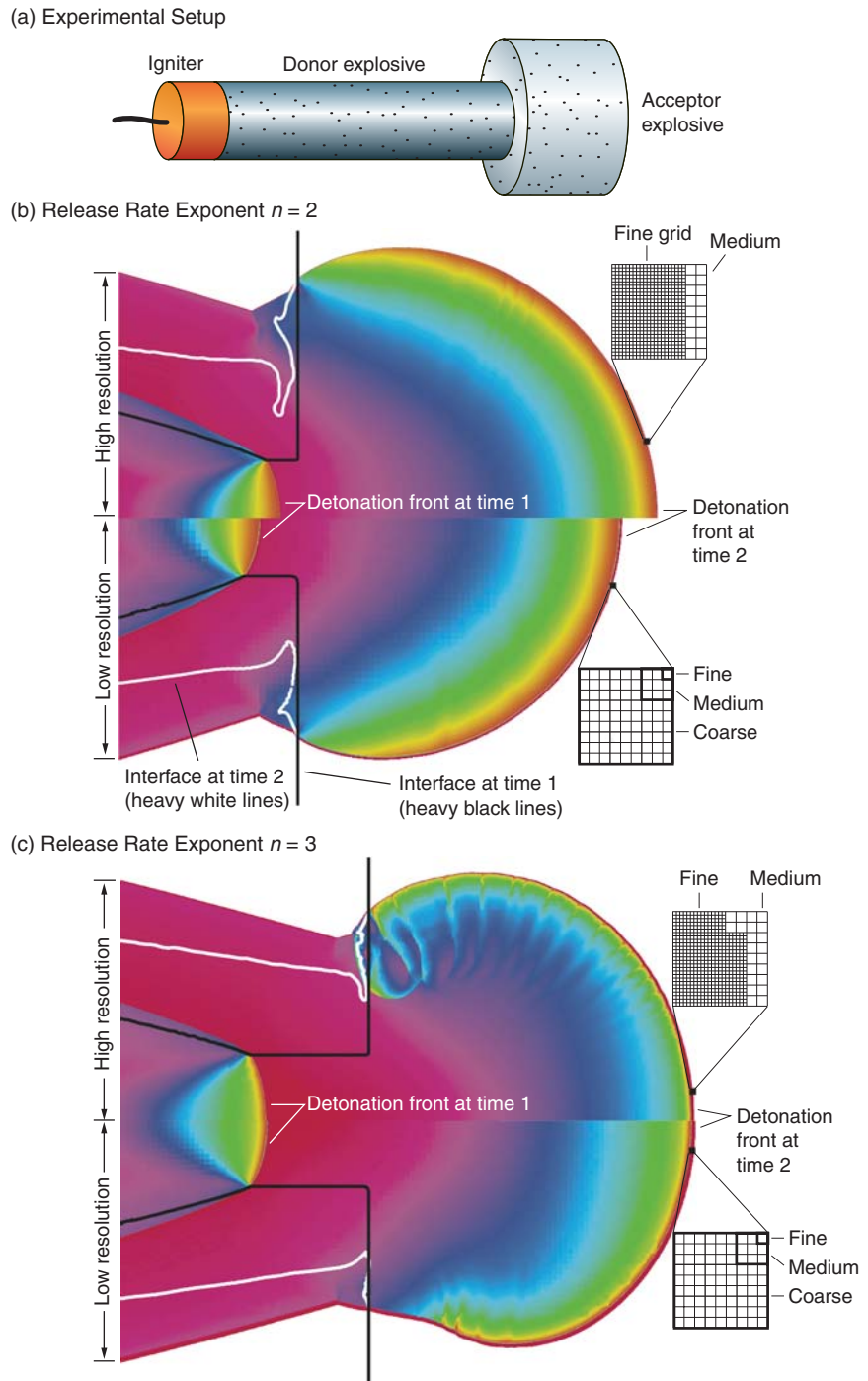
The class of models just described has been applied to problems with fully 3-D geometries, but our studies show that the solutions contain features that are unsatisfactory for use in real performance codes. As an example, we consider the propagation of a detonation wave in a stack of right-circular cylinders of explosive—see Figure 8(a). The object is to predict the progress of detonation as it diffracts from a smaller to a larger coaxial cylinder. We simulated this experiment with a model similar to that described above but with a simpler EOS and a simpler rate law

for the reaction zone. The EOS for the unreacted and reacted explosive is obtained by setting  $A_1 = B_1 = 0$  and  $\omega_s = \omega_g = 2$  and by simplifying the rate law to

$$R(P, v, \lambda) = kP^n(1 - \lambda)^\mu, \quad (7)$$

where  $k$  is a constant and  $\mu$  is set to  $\mu = 1/2$ . All constants were selected to mimic the condensed-phase explosive PBX 9502. In the experiment, the explosive is embedded in a low-density plastic. The plastic does not affect the flow in the reaction zone; rather the explosive behaves as it would if it were totally unconfined (floating in free space). We simulated the experiment using the Amrita (Quirk 1998) computational environment, which provides adaptive mesh refinement, simulation scheduling, and documentation of the results. We also used the Ghost Fluid interface-tracking algorithm (Fedkiw et al. 1999) to keep a sharp interface between the explosive and the confining inert plastic and a Lax-Friedrichs solver to update the flow.

Figures 8(b) and 8(c) are composites. Each shows two solutions for the pressure—one before and one after the detonation passes into the wider (acceptor) section of the explosive. These two figures differ in that they show solutions for two different energy-release rate functions  $R$ —Equation (7)—proportional to the square of the pressure,  $n = 2$ , and the cube of the pressure,  $n = 3$ , respectively. The top and bottom halves of each figure show results for different resolutions in the steady-state ZND reaction zone, 72 and 18 points, respectively, for the  $n = 2$  solution, and 18 and 9 points, respectively, for the  $n = 3$  solution. For both rate laws, the location of the detonation front depends significantly on numerical resolution. Also, for the  $n = 3$  rate law, the detonation is highly unstable: There are very large pres-



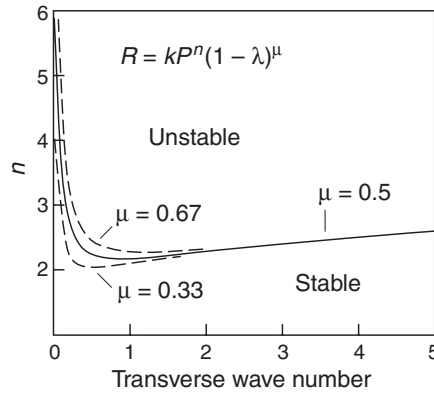
**Figure 8. Standard Modeling of Detonation in 3-D Geometries**  
 In (a) we show the setup for detonation wave propagation in a stack of right-circular cylinders, and in (b) and (c) we show the results from the standard modeling paradigm with different energy-release rate laws,  $n = 2$  and  $n = 3$  with  $\mu = 1/2$  in Equation (7). In each case, top and bottom figures display the results for different resolutions: 72 and 18 points in the reaction zone for  $n = 2$  and 18 and 9 points in the reaction zone for  $n = 3$ , respectively. In both cases, the results change markedly with increasing resolution. Also, with the more pressure-sensitive rate law,  $n = 3$ , the reaction zone shows high-frequency structure that is an artifact of this modeling paradigm for heterogeneous explosives.

sure and high-frequency structures in the acceptor explosive (wider section), and the details of the instability are very much resolution dependent.

Short et al. (2003) have analyzed the stability of the classical, steady-state ZND reaction-zone structure to small, multi-dimensional disturbances and shown that it is unstable to even small perturbations whenever  $n$  is greater than 2.1675 for the model that we have described here. Figure 9 shows the results of this stability analysis. Also, the addition of a nozzle term to Equation (3) (that term mimics the energy loss from the reaction zone because of multidimensional flow effects) leads to a further destabilization of the reaction zone to 1-D disturbances. (Note: The high-resolution  $n = 2$  simulation also shows some signs of instability.) The root of this instability can be understood with the following argument. Detonation in this homogeneous-fluid model is a balance between shock-initiated energy-releasing reactions and the acoustic transport of that energy to support the shock. When a pressure perturbation in the reaction zone affects the reaction rate much more than it does the sound speed, then small pressure fluctuations can disrupt the balance between energy liberation and transport, and instability can be the result.

### Problems with the Reaction-Zone Modeling Paradigm

Both the dependence on numerical resolution and the appearance of the high-frequency structure in the acceptor region of the explosive represent a significant problem for this modeling paradigm. In independent studies of this simple model for the  $n = 2$  case, we have shown that to predict the detonation speed in the donor section to within 10 m/s requires 50 or more points in the ZND reaction zone. (Aslam et al. 1998). This number



**Figure 9. Stability Analysis Results for the Standard Detonation Modeling Paradigm**

The vertical axis is the pressure exponent of the heat-release rate  $R$ , and the horizontal axis is the wave number of the perturbing transverse disturbance. For any state above the curve, the ZND detonation is unstable. For  $\mu = 0.5$ , a ZND detonation is unstable to two-dimensional disturbances whenever  $n > 2.1675$ .

translates into a very computationally intensive problem for typical 3-D engineering scale problems, where the reaction-zone length is very much shorter than the dimension of the explosive piece. At any instant, about  $10^{10}$  relatively small computational nodes would be needed in the reaction zone, and the computational time on a large parallel processing computer would be about 100 days. Second, the high-frequency, acoustically based transverse wave structure is an artifact of this simple homogeneous-fluid model and is not observed in our heterogeneous explosives. The substructure associated with the granular structure of real heterogeneous explosives derives from the material particles, not acoustic waves. In fact, the granularity inhibits the formation of large transverse acoustic waves. Thus, although the simple homogeneous-fluid model reproduces reasonably well the leading-order features of detonation, such as the ZND structure, it fails to

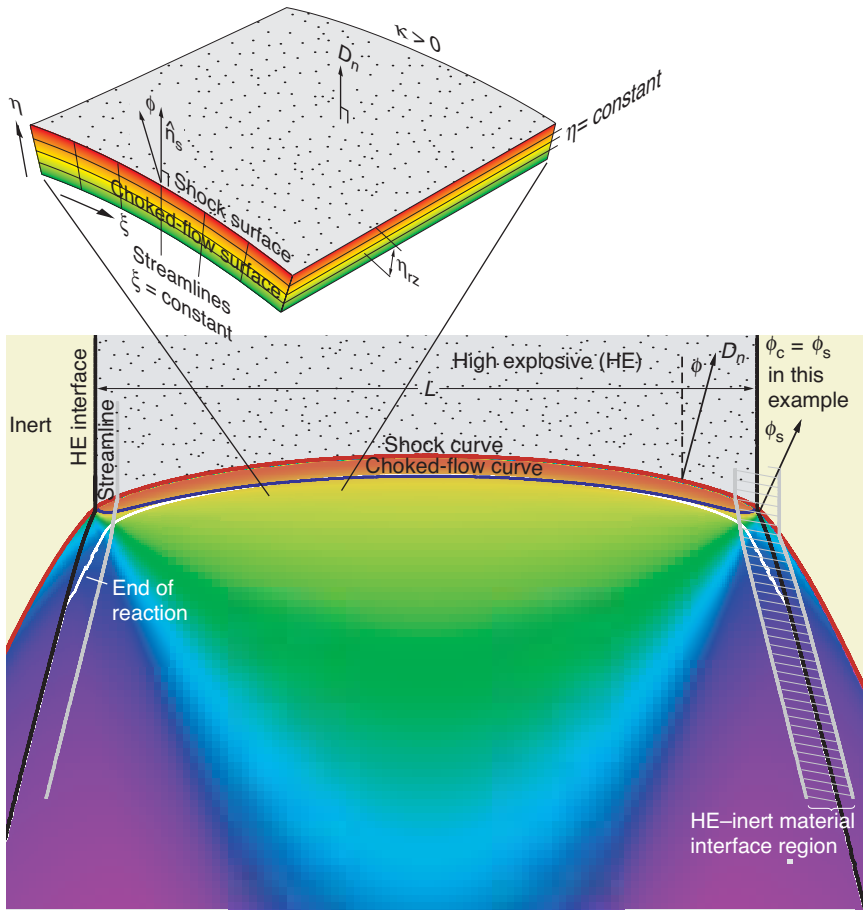
describe higher-order features of real heterogeneous explosives.

### DSD, a Subscale Model of Detonation

We have championed an approach to the performance problem that is philosophically different from the standard paradigm just described (Aslam et al. 1996). In the DSD approach, we exploit the fact that the explosive pieces of engineering interest are large compared with the reaction-zone length and substitute a subscale model for the detailed model of the reaction zone. To construct this subscale model, we consider how the detonation reaction zone is influenced by weak curvature of the shock front and derive a constraint equation relating the speed of the detonation front to the shape of that front. We then derive a boundary condition on that equation that relates edge effects to the detonation wave shape. Thus, on the scale of the explosive, the reaction zone becomes a front, a discontinuity, separating fresh from burnt explosive. In this way, DSD focuses on the two primary goals of the performance problem: accurate prediction of (1) the local detonation speed and detonation arrival times in a weapons simulation and (2) the  $P$ - $v$  (pressure-volume) work that the high-pressure detonation products can perform on the inert materials with which the explosive is in contact. As we will see, this approach also filters out the high-frequency features and vastly reduces the computational requirements by comparison with the standard modeling paradigm described above.

Figure 10 shows a detailed view of the reaction zone and shock front, with unburnt explosive above and burnt explosive below. The reaction-zone length is short compared with the explosive charge dimension,  $L$ . In place of Cartesian coordinates, we use coordinates that are attached to the shock surface (see the upper inset in Figure 10). The constant  $\eta$  is the dis-





**Figure 10. Multidimensional Reaction Zone and the DSD Shock-Attached Coordinates**

A multidimensional reaction zone in a cylindrical detonating explosive (gray) is weakly confined at its edges by a low-density inert material (yellow). As shown, the reaction zone is typically short compared with the dimension of the explosive charge,  $L$ . The upper inset depicts a segment of the 3-D reaction zone with the intrinsic shock-attached coordinates used in DSD analysis. In the DSD limit, the ratio of the reaction-zone thickness to the scale of the explosive charge is very, very small,  $\eta_{rz}/L = O(\epsilon) \ll 1$ , and a subscale front model takes the place of the detailed reaction-zone model. The crosshatched area, approximately the width of the reaction zone and straddling the explosive–inert material interface, defines the region where an analysis of the boundary region is performed. As explained in the text, that analysis leads to boundary conditions for the subscale model.

tance through the reaction zone normal to the shock surface, and  $\xi$  is the distance along the shock measured from the centerline of the explosive. Thus, curves of constant  $\eta$  are parallel to the shock, and the lines of constant  $\xi$  are in the direction of the local normal to the shock surface. Because the features of interest are on the scale of the explosive, we define a dimensionless scale,  $\epsilon = \eta_{rz}/L \ll 1$ , where  $\eta_{rz}$  is the scale of the detonation reaction-

zone length. This scale aids in the derivation of the subscale model.

DSD assumes that the detonation reaction zone departs from its 1-D (planar) steady-state ZND form by a small amount. That small departure is determined by both the size of the shock curvature  $\kappa$  measured in units of the reaction-zone length scale  $\eta_{rz}$ , ( $\eta_{rz}\kappa = O(\epsilon) \ll 1$ ) and the departure of  $D_n$ , the detonation speed in the direction of the shock normal vector,

from  $D_{CJ}$ , the detonation speed for a 1-D steady-state wave. The relevant scaled detonation speed is  $(D_n/D_{CJ} - 1) = \mathcal{D} = \epsilon \mathcal{D}^*$

To construct an asymptotic solution—a solution in the limit that  $\epsilon \ll 1$ —for the multidimensional detonation reaction-zone flow, we first introduce into the standard detonation model, Equations (1)–(3) and (6), the following slowly changing, scaled, independent variables:

$$\tilde{t} = \epsilon t, \quad \tilde{\xi} = \epsilon^{1/2} \xi, \quad \phi = \epsilon^{1/2} \tilde{\phi}, \quad (8)$$

where  $\phi$  is the shock normal angle defined in Figure 10. We then expand the solution vector  $Y = (\rho, u_\eta, P)^T$  as

$$Y = Y^{(0)} + \epsilon Y^{(1)} + \epsilon^2 Y^{(2)} + \dots, \quad (9)$$

$$u_\xi = \epsilon^{3/2} u_\xi^{3/2} + \dots, \quad (10)$$

The leading order term in the solution vector (designated with a superscript 0) represents the 1-D ZND solution, whereas the terms proportional to powers of  $\epsilon$  bring in the effects of multidimensionality and time dependence. A compatibility constraint emerges on the solution that forces a relationship between the shock curvature  $\kappa$  and various derivatives of the scaled detonation speed  $\mathcal{D}$

$$\mathcal{F}(\mathcal{D}) = \kappa + \mathcal{A}(\mathcal{D}) \frac{D\mathcal{D}}{Dt} - \mathcal{B}(\mathcal{D}) \frac{\partial^2 \mathcal{D}}{\partial \tilde{\xi}^2} + \dots, \quad (11)$$

where  $\mathcal{F}(\mathcal{D})$  is a decreasing function of  $\mathcal{D}$  with  $\mathcal{F}(0) = 0$ ,  $\mathcal{A}(\mathcal{D}) > 0$  and  $\mathcal{B}(\mathcal{D}) > 0$ . A term-by-term examination of the right-hand side of this equation shows that (1) increasing  $\kappa$  (shock curvature) slows the detonation wave, (2) the acceleration term  $\mathcal{A}(\mathcal{D})(D\mathcal{D}/Dt)$  acts like inertia and resists changes in the front speed and

shape, and (3) the dissipative term  $\mathcal{B}(\mathcal{D})(\partial^2 \mathcal{D}/\partial \xi^2)$  damps high frequencies and thus the formation of kinks on the wave front. The dispersion relation for the linearized form of Equation (11)

$$\omega = -ik^2(\mathcal{B}/2\mathcal{A}) \pm \sqrt{k^2/\mathcal{A} - k^4(\mathcal{B}/2\mathcal{A})^2} \quad (12)$$

reveals that, at low transverse frequencies (small values of the wave number  $k$ ), Equation (11) predicts dissipative, transverse waves moving along the front. On the other hand, at high transverse frequencies (large  $k$ ), Equation (11) is purely dissipative. Thus, by adopting the scaled variables of Equation (8), high-frequency features such as kinks on the front will not form.

Thus, on the scale of the explosive piece, the detonation reaction zone looks like a discontinuity separating fresh and burnt explosive. Moreover, this discontinuity occurs along a surface that propagates according to the dynamics given by Equation (11), which can be viewed as an intrinsic detonation propagation law for an explosive. It is important to recognize that the forms of the coefficient functions  $\mathcal{F}(\mathcal{D})$ ,  $\mathcal{A}(\mathcal{D})$ , and  $\mathcal{B}(\mathcal{D})$  depend on the material description of the explosive (the EOS and the global heat-release rate).

In addition to specifying a propagation law such as Equation (11), we need to prescribe a boundary condition on the front, where the front meets the edge of the explosive piece (see the right portion of Figure 10). Just as we constructed a subscale model to mimic the effects of the reaction zone on the front motion, we construct a subscale model to mimic the effect of confinement by adjacent inert materials on the detonation speed. Roughly speaking, the more compliant the inert materials, the

greater the deflection of explosive streamlines and shock angle  $\phi$  and the greater the pressure drop. A boundary layer analysis performed within a distance of one reaction-zone length on either side of the explosive–inert material interface reveals (see cross-hatched region) that this interaction establishes a unique shock-edge angle,  $\phi_c$ , which is a function of the explosive and inert material pair considered. That angle serves as a boundary condition for Equation (11). The weaker the confinement, the greater the value of  $\phi_c$ , up to the point where the lateral expansion of the detonation products becomes choked (the development of a sonic state behind the shock, as shown in Figure 10), halting any further drop in pressure at the shock. This phenomenon occurs at a value of  $\phi$  called the sonic angle,  $\phi_s$ , which depends solely on the properties of the explosive and is about  $45^\circ$  for our explosives (Aslam and Bdzil 2002, Bdzil 1981).

### DSD Calibration and Propagation of Detonation Front

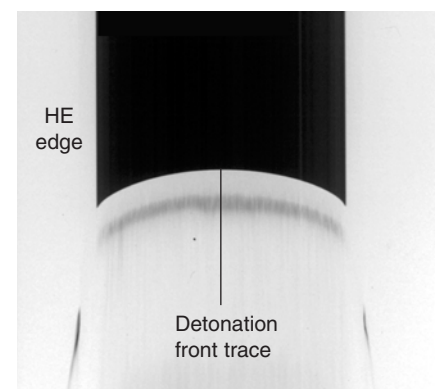
To validate the DSD approach, we used it to compute the detonation front shapes for the rate law of Equation (7) and compared the DSD results with numerical results from the standard paradigm, Equations (1)–(3), (6), and (7). The good agreement validates the DSD approach, at least for  $n < 2.1675$ , for which the standard approach is fairly accurate (Aslam et al. 1998). Although we could have derived a detonation propagation law directly from a calibrated shock-initiation model, such as the Lee-Tarver Ignition and Growth model, we chose, instead, to derive Equation (11) more generically and calibrate it from experimental data on multidimensional detonation. In that way, we bypassed artifacts of the homoge-

neous-fluid model paradigm and built in features faithful to real, heterogeneous explosives but not easily modeled with the standard paradigm.

Calibration data are often obtained from measurements of the detonation speed and front curvature in explosive cylinders of various sizes (see Figure 11). For explosives such as PBX 9502, those data can be fit reasonably well with just the leading term in the propagation law of Equation (11):

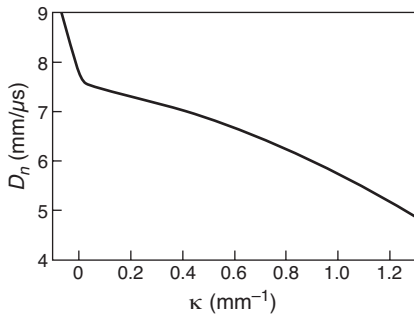
$$\kappa = \mathcal{F}(\mathcal{D}), \quad (13)$$

which specifies a simple relationship between the detonation speed and the detonation shock-front curvature. The propagation law so obtained for PBX 9502 predicts that the shock-normal detonation speed decreases substantially with increasing shock-front curvature (see Figure 12). Figure 13(a) shows a DSD prediction for the detonation front shape at initiation and two later times as the detonation propagates through an arc of PBX 9502. Figure 13(b), a top-down view, shows the DSD solution lagging behind the simple constant-velocity CJ solution. The significant differences between these two argue for the importance of



**Figure 11. Detonation Front Measurement in a Cylinder of Explosive**

This image, taken by a streak camera, shows the detonation front arrival time vs radius at the planar face of the explosive cylinder (Hill 1998).



**Figure 12. DSD Propagation Law for PBX 9502 at 25°C**

The DSD propagation law predicts that detonation speed normal to the shock  $D_n$  decreases substantially with increasing shock-front curvature  $\kappa$ .

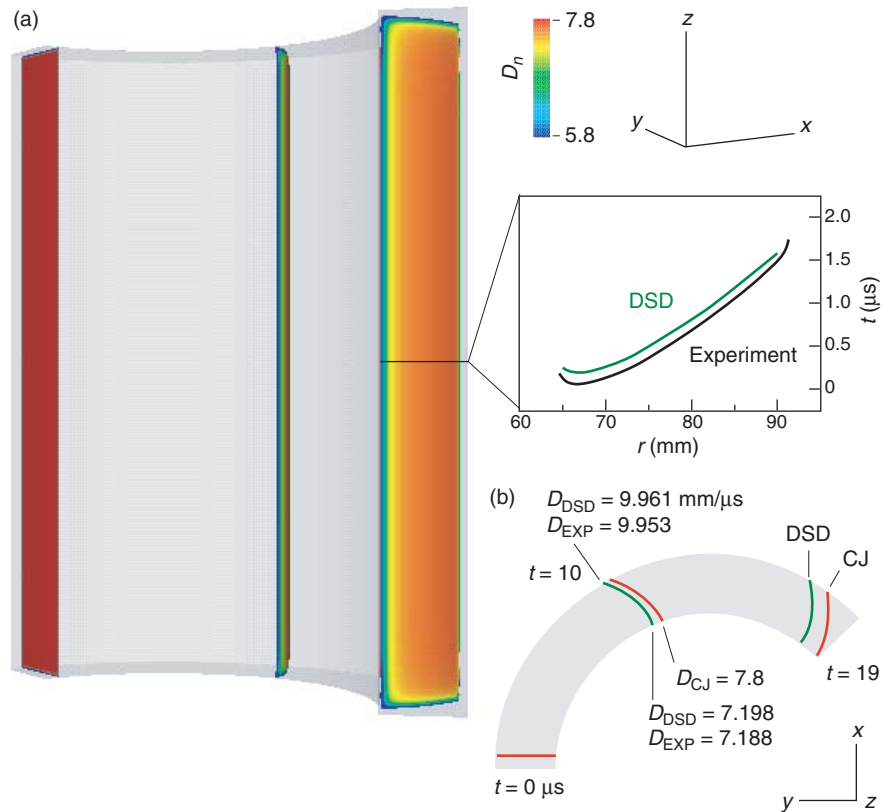
including reaction-zone effects. The DSD shapes and velocities are in very good agreement with experiment

To obtain the DSD solution shown in Figure 13, we start from the front propagation law and edge boundary conditions and use level-set methods to compute the propagation of the front. These methods work by embedding the detonation front in a level-set function,  $\Psi$ , and then evolving this function according to

$$\partial\psi/\partial t + D_n|\nabla\psi| = 0 \quad (14)$$

By our definition, the level surface  $\psi = 0$  corresponds to the detonation front initially and at any subsequent time. We compute the actual detonation front by finding the  $\psi = 0$  contour. The level-set methodology offers significant computational advantages because it enables easy handling of complex explosive topologies and detonation interactions.

This past year, we have developed a first-order accurate, 3-D computational algorithm that uses the level-set method and that runs on parallel computing platforms. Results from that method are shown in Figure 14, which displays a detonation initiated simultaneously at the ends of two symmetrical legs and propagating through a piece of PBX 9502 with



**Figure 13. DSD Detonation through an Arc of PBX 9502**

(a) The 3-D composite image shows the progress of the DSD front through an explosive arc. Each detonation front is colored by the local instantaneous normal detonation speed. The slowing of the detonation near the edges is apparent by the change in color from red to green. Shown in the inset are plots of the DSD (green) and experimental (black) times of arrival of the detonation front along the midline of the planar edge of the arc (measured in units of the cylinder radius). The resulting curves show the similarity between the DSD and measured wave-front shapes. The DSD and experimental plots are set off to display results. (b) This top-down view shows the intersection of the DSD and CJ fronts with a plane passing through the middle of the arc. The DSD detonation speed slows down by 10% relative to its initial value, whereas the CJ detonation speed is constant. Consequently, there is a growing separation of DSD and CJ fronts. The phase velocities of the DSD wave along the inner and outer surfaces of the arc agree well with the experimental values.

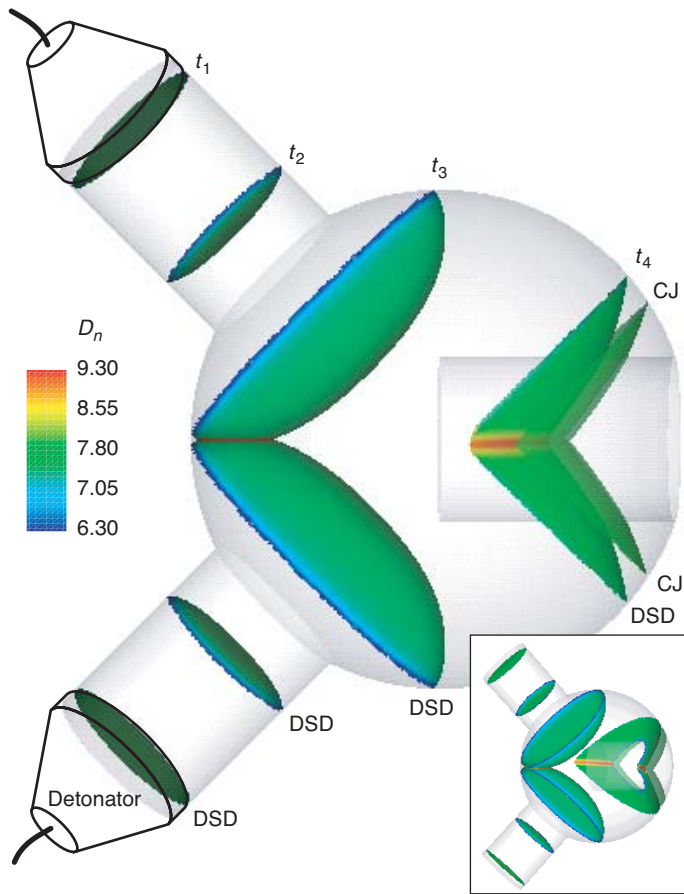
complex geometry. The detonation fronts are shown at four different times. The shape of the explosive, two protruding legs on one side and a cylindrical hole on the other, forces the detonation fronts to merge ( $t = t_3$ ) and then bifurcate ( $t = t_4$ ). Merging and bifurcation of different detonation fronts are automatically treated with the level-set-based DSD approach. For comparison, the CJ

wave is also shown (as ahead) in the last snapshot ( $t = t_4$ ).

### New Modeling Paradigms for Detonation Reaction Zones

The methods just described represent the state of the art in detonation modeling for engineering applications. In closing, we outline our vision for the future.

The modeling paradigm represented



**Figure 14. A 3-D DSD Calculation**

This example shows how DSD can handle the merging of separated detonation fronts, the acceleration of the detonation in regions where the fronts converge, and the bifurcation of the detonation wave around obstacles. Detonation starts in the two legs (left side of the figure) and progresses toward the cylindrical hole on the right. Four snapshots show the progress of the detonation waves through the sample, and the DSD and CJ calculations are compared on the fourth snapshot. The detonation front is colored with the local value of the detonation speed. The inset shows an oblique view.

by Equations (1)–(6) has serious shortcomings. Principally, such a continuum model does not include any phenomenon that might be important for heterogeneous materials, such as scattering or dispersion of acoustic waves and dissipation of energy from the reaction-zone scale to the subreaction-zone scales of the hotspots. Current models are effective for homogeneous explosives. As we better understand the details of the hotspots and reaction chemistry under detonation conditions, we will need to develop better continuum models. If subnano-scale measurements support our current view that the chemical

reactions important in detonation are extremely state sensitive, then very short scale regions where the hotspot temperatures are most extreme will have a disproportionate effect toward accelerating the reaction chemistry (Bdzil et al. 1999, Menikoff and Kober 1999). These would be subgrain scales, related more to details of the grain shape than the grain volume (Figure 4 shows how complex this substructure can be). The notion of doing numerically resolved meso-scale simulations of detonation in granular explosives and then somehow averaging those results to develop appropriate continu-

um-level engineering models seems to be many years away.

Given our lack of detailed information about the relationship between the geometry of explosive grains and the constitutive properties of our materials (including, for example, heat conductivity) under detonation conditions, a more realistic midterm goal is to develop subscale models for the reaction zone in which behaviors on the grain and subgrain scales are parameterized in terms of longer wavelength variables. We are working to develop rational asymptotic models that indicate not only how the explosives' global heat-release rate should be modeled but also how the presence of granularity and hotspots in these materials affects the basic modeling structure—that is, what modifications need to be made to Equations (1)–(6). For example, significant density variations in a material on a short-wavelength scale will appear on the long-wavelength continuum scale as dispersion terms added to the basic conservation laws—refer to Equations (1)–(3). The presence of such terms could be expected to scatter acoustic waves and inhibit the detonation modeling instabilities that we have observed in our homogeneous-explosive models. Whatever improved modeling paradigms are developed for the continuum response of heterogeneous explosives, we expect that, for the foreseeable future, models will have to be calibrated to experiments if they are to make the accurate predictions of detonation propagation necessary for weapons simulations. ■

## Further Reading

- Aslam, T. D. 2003. A Level Set Algorithm for Tracking Discontinuities in Hyperbolic Conservation Laws II: Systems of Equations (to be published in *J. Sci. Comput.* **19** (1–3), December 2003).
- Aslam, T. D., and J. B. Bdzil. 2002. “Numerical and Theoretical Investigations on Detonation-Inert Confinement Interaction.” Twelfth Symposium (Int.) on Detonation, San Diego, California, August 11–16, 2002. Los Alamos National Laboratory document LA-UR-01-4664 (to be published in *Proceedings—Twelfth International Detonation Symposium*). [Online]: <http://www.sainc.com/onr/detsymp/technicalProgram.htm>
- Aslam, T. D., J. B. Bdzil, and L. G. Hill. 2000. “Extensions to DSD Theory: Analysis of PBX 9502 Rate Stick Data.” In *Proceedings—Eleventh International Detonation Symposium*, ONR 33300-5, pp. 21–29. Arlington, Virginia: Office of Naval Research.
- Aslam, T. D., J. B. Bdzil, and D. S. Stewart. 1996. Level Set Methods Applied to Modeling Detonation Shock Dynamics. *J. Comput. Phys.* **126** (2): 390.
- Bdzil, J. B. 1981. Steady-State Two-Dimensional Detonation. *J. Fluid Mech.* **108**: 195.
- Bdzil, J. B., R. Menikoff, S. F. Son, A. K. Kapila, and D. S. Stewart. 1999. Two-Phase Modeling of Deflagration-to-Detonation Transition in Granular Materials: A Critical Examination of Modeling Issues. *Phys. Fluids* **11** (2): 378.
- Campbell, A. W., and J. R. Travis. 1985. “Shock Desensitization of PBX-9404 and Composition B-3.” In *Proceedings—Eighth International Detonation Symposium*, J. M. Short, Ed., pp. 1057–1068. Naval Surface Weapons Center, NSWC MP 86-194.
- Fedkiw, R. P., T. Aslam, B. Merriman, and S. Osher. 1999. A Non-Oscillatory Eulerian Approach to Interfaces in Multimaterial Flows (the Ghost Fluid Method). *J. Comput. Phys.* **152** (2): 457.
- Henshaw, W. D., and D. W. Schwendeman. 2003. “An Adaptive Numerical Scheme for High-Speed Reactive Flow on Overlapping Grids.” Lawrence Livermore National Laboratory preprint UCRL-JC-151574 (submitted to *J. Comput. Phys.*).
- Hill, L. G. 2002. “Development of the LANL Sandwich Test.” In *Shock Compression of Condensed Matter—2001: Proceedings of the Conference of the American Physical Society, Topical Group on Shock Compression of Condensed Matter*, M. D. Furnish, Y. Horie, and N. N. Thadhani, Eds. Vol. 620 (1), pp. 149–152. Secaucus, New Jersey: Springer-Verlag.
- Hill, L. G., Bdzil, J. B., and T. D. Aslam. 2000. “Front Curvature Rate Stick Measurements and Detonation Shock Dynamics Calibration for PBX9502 over a Wide Temperature Range.” In *Proceedings—Eleventh International Detonation Symposium*, ONR 33300-5, pp. 1029–1037. Arlington, Virginia: Office of Naval Research.
- McGrane, S. D., D. S. Moore, and D. J. Funk. 2003. Sub-Picosecond Shock Interferometry of Transparent Thin Films. *J. Appl. Phys.* **93** (9): 5063.
- Menikoff, R., and E. M. Kober. 1999. “Compaction Waves in Granular HMX,” Los Alamos National Laboratory report LA-13546-MS.
- Quirk, J. J. 1998. AMRITA: “A Computational Facility (for CFD Modeling).” In 29th Computational Fluid Dynamics, von Karman Institute Lecture Series. ISSN 0377-8312.
- Shaw, M. S. 2002. “Direct Simulation of Detonation Products Equation of State by a Composite Monte Carlo Method,” Twelfth Symposium (Int.) on Detonation, San Diego, California, August 11–16, 2002 (to be published in *Proceedings—Twelfth International Detonation Symposium*). [Online]: <http://www.sainc.com/onr/detsymp/technicalProgram.htm>
- Sheffield, S. A., R. L. Gustavsen, and R. R. Alcon. 2000. “In Situ Magnetic Gauging Technique Used at LANL—Method and Shock Information Obtained.” In *Shock Compression of Condensed Matter—1999, AIP Conference Proceedings*, M. D. Furnish, L. C. Chhabildas, and R. S. Hixson, Eds., Vol. 505, p. 1043. Secaucus, New Jersey: Springer-Verlag.
- Sheffield, S. A., R. L. Gustavsen, L. G. Hill, and R. R. Alcon. 2000. “Electromagnetic Gauge Measurements of Shock Initiating PBX9501 and PBX9502 Explosives.” In *Proceedings—Eleventh International Detonation Symposium*, ONR 33300-5, pp. 451–458. Arlington, Virginia: Office of Naval Research.
- Short, M., J. B. Bdzil, G. J. Sharpe, and I. Angelova. Detonation Stability for a Condensed Phase Reaction Model with a Variable Reaction Order (submitted to *Combust. Theory Modell.*).
- Skidmore, C. B., D. S. Phillips, and N. B. Crane. 1997. Microscopical Examination of Plastic-Bonded Explosives. *Microscope* **45** (4): 127.
- Tarver, C., and E. McGuire. 2002. “Reactive Flow Modeling of the Interaction of TATB Detonation Waves with Inert Materials.” Twelfth Symposium (Int.) on Detonation, San Diego, California, August 11–16, 2002 (to be published in *Proceedings—Twelfth International Detonation Symposium*). [Online]: <http://www.sainc.com/onr/detsymp/technicalProgram.htm>
- Watt, S. D., and G. J. Sharpe. One-Dimensional Linear Stability of Curved Detonations (submitted to *Proc. R. Soc. Lond., Ser. A*).

For further information, contact  
Scott Watson (505) 667-8054  
([jbb@lanl.gov](mailto:jbb@lanl.gov)).

# Damage Evolution in Ductile Metals

## *Developing a quantitative and predictive understanding*

*Anna K. Zurek, W. Richards Thissell, Carl P. Trujillo, Davis L. Tonks,  
Benjamin L. Henrie, and Rhonald K. Keinigs*

If one pulls hard enough on a bar of soft, ductile metal such as aluminum, it will stretch, and if one continues to pull, the bar will eventually break. What happens internally to cause it to break? To begin with, the pulling puts the material into a state of tension, which, if high enough, will cause tiny voids to form. This process, known as nucleation, typically begins near sites of defects such as impurities that are introduced during the original processing of the material. Following nucleation, the voids begin to expand, and, if close enough together, they coalesce to form microscopic cracks. In regions having a high density of voids and microcracks, this progression culminates with the development of a complete surface failure; that is, the bar breaks. That scenario is the currently accepted model of damage evolution in most ductile metals.

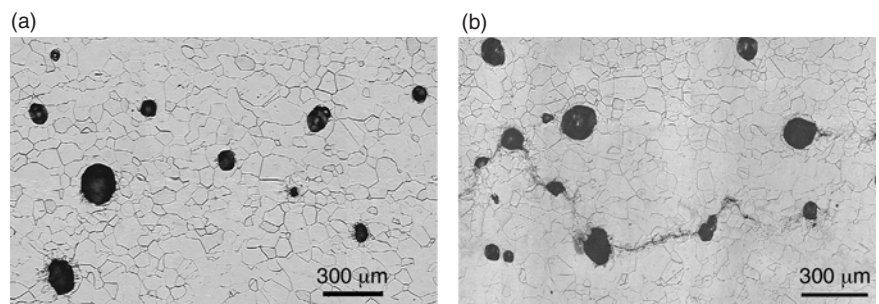
Metals that are subjected to shock waves can also fail via this pathway. In shocked materials, a state of high tension can be produced as pressure waves reflect off free surfaces and interact with each other. Shock-induced damage, or spall, as it is known among material physicists, occurs in metals shocked by lasers and in tank armor hit by conventional munitions. Because even the plutonium in a nuclear weapon can spall, this process is an important area of research for science-based stockpile stewardship. At Los Alamos collaboration between experimentalists and modelers is beginning to paint a detailed picture of the events leading up to spall. In this article we discuss recent results from gas-gun shock spall experi-

ments specially designed to investigate the dynamics of ductile damage and failure.

Rather sophisticated models of damage evolution that incorporate many of the steps involved in metals spallation are being developed at Los Alamos. One of the authors is developing a new micromechanical model that includes void growth, void coalescence, and crack formation (Tonks et al. 2002). When validated, the model will replace simpler damage models currently employed in advanced simulation and computing codes, tensile plasticity codes, and others. To aid the validation process and provide direction for further improvements in the model, we performed a number of well-controlled gas-gun experiments on the evolution of spall in tantalum and copper targets. Ideally, one would like to have enough control to arrest the damage evolution at different stages of development. In our gas-gun experiments, we made the shock pressures large enough to initiate

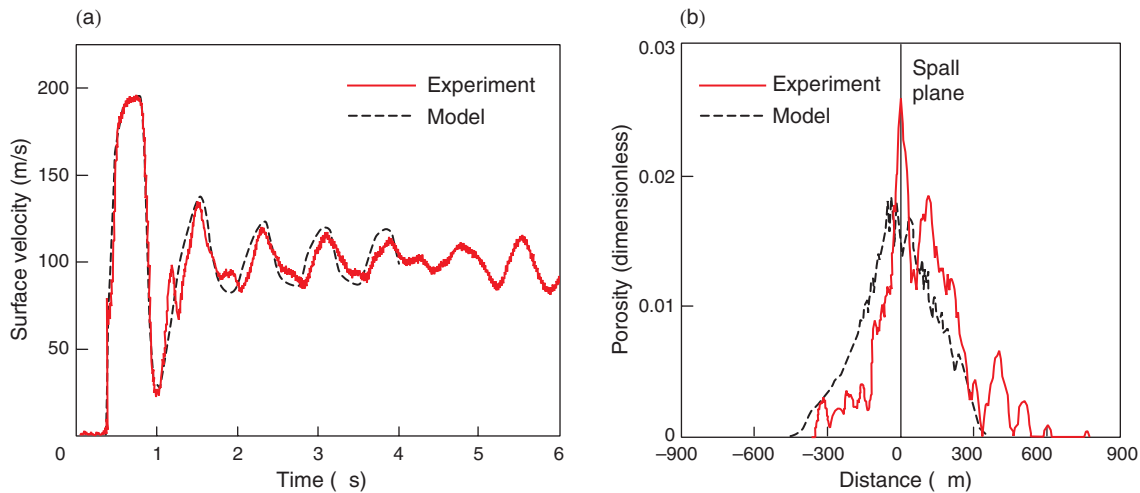
the damage evolution sequence, but not so large as to result in fracturing the samples. The resulting damage is called incipient spall. We also varied the shock loading to investigate the effects on damage from changes in peak pressure and shock duration. The targets were recovered and microscopically examined to determine the degree and type of damage produced under each loading condition.

Figure 1 shows optical micrographs of cross sections through the damaged region of two tantalum samples. Both samples were shocked to the same peak stress, but the duration of the shocks differed by a factor of two. Both samples show damage in the form of spherical voids, but the sample subjected to the longer period of shock loading developed discontinuities in the microstructure of the metal—see Figure 1(b). These “linkages” are attributed to strain localizations that presumably had time to develop during the longer loading period. Such areas



**Figure 1. Optical Micrographs of Incipient Spall in Tantalum**

The optical micrographs show the microstructure of damaged regions of tantalum samples after spall tests at a shock pressure of 5.6 GPa (56 kb). The samples were subjected to different shock durations: (a) 1.1  $\mu$ s, resulting in a final porosity of 4.1% and (b) 2.2  $\mu$ s, resulting in a final porosity of 11.6%. The sample in (b), which was subjected to a longer period of shock loading, shows a line of damage connecting two voids, a form of damage not included in most models of spall.



**Figure 2. Comparison of Model Predictions and Experimental Results for Tantalum Spall**

In the gas-gun experiment, a quartz flyer plate, 1.5 mm thick, hit a tantalum target of the same thickness at 448 m/s. The impact produced in the target a shock pressure of 5.6 GPa for a duration of 0.4  $\mu$ s. (a) The plots show close agreement between model predictions and experimental results for the velocity history of the target's back surface (VISAR trace). (b) The plots show the resulting porosity distribution in the region of the spall plane, with maximum porosity of 0.025. The standard error between the measured and modeled porosity was calculated as 0.05.

of high deformation between voids, although not accounted for in most models, might be the precursor to the coalescence of voids into microcracks. However, this hypothesis requires further investigation.

At present, we can measure void sizes and distributions, volumetric void-number-density distributions, clustering, near-neighbor distances, strain localization distances, and final porosity in recovered targets. Those data are provided to the modelers to test the accuracy of their predictions. On each spall test, we used velocity interferometry (VISAR, or velocity interferometer system for any reflector) to measure the back free-surface velocity of the shocked target as a function of time, providing yet another piece of constraining data for the model predictions. (See Figure 3 in the Hixson article on page 117 for a discussion of this measurement technique.) Figure 2 shows an example in which the model very accurately predicted the free-surface velocity history as well as the incipient damage of tantalum. In that case, postshot metallurgical inspection revealed no evidence of strain localization in the sample. However, when

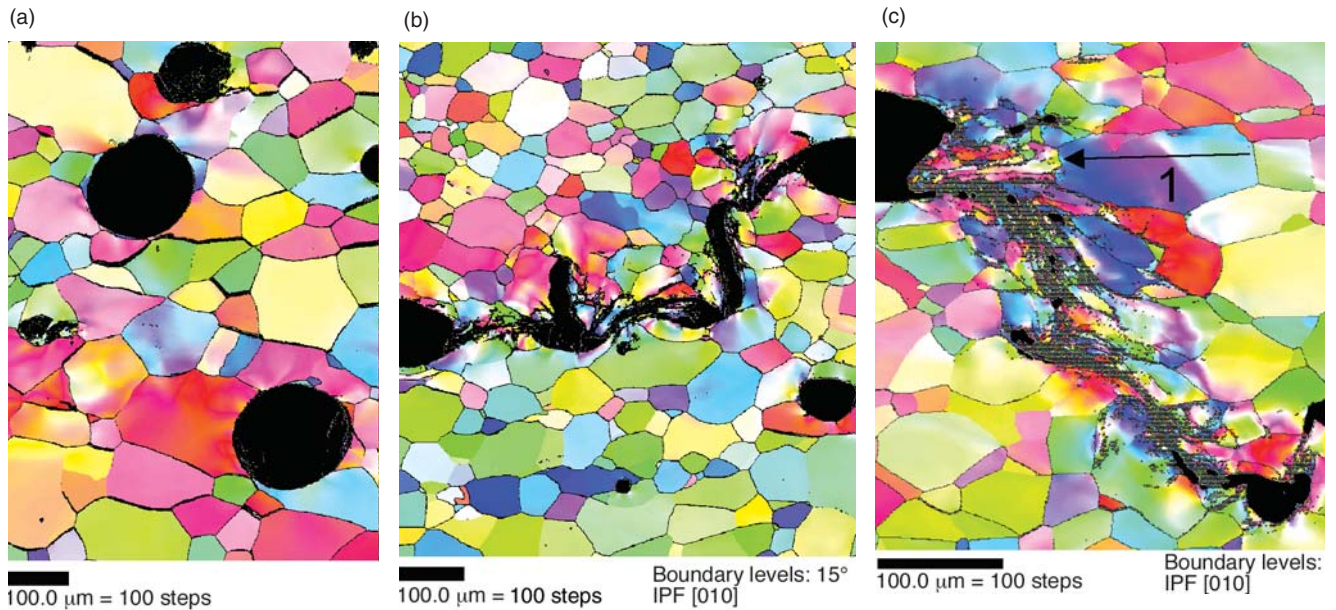
shock loading produced more-extensive damage in the sample, including coalescence and strain localization, the model was significantly less accurate in predicting the results. In particular, the model overpredicted the amount of porosity developed in the sample. We surmise that the energy from the shock that went into making the extra voids in the model calculation was in reality partitioned into creating linkages between voids.

To gain additional quantitative information pertaining to strain localization and to guide further improvements to our damage models, we are developing the technique of automated electron backscatter diffraction (EBSD) (Adams et al. 1993). In regions where the strain in a material has localized, the dislocation density increases and causes degradation in the electron backscatter patterns. From these degraded patterns, one can extract strain localization information, such as information on their distributions, lengths, and widths.

Figure 3 shows examples of the details that can be analyzed by EBSD. False coloring is used to identify regions of specific crystallographic

orientation within the individual grains of the sample tested. The highlighted grain boundaries have a misorientation (relative to the bulk) no larger than 15°. Regions where the misorientation is so high that the technique does not resolve the details appear as gray pixels. We equate those regions of a high deformation with regions of strain localization. Since the information from EBSD is in a digital form, we can clearly differentiate the portion of the energy consumed by the void formation from that consumed by the linkages, or the strain instabilities.

We are making significant progress in our quest to understand the phenomenon of ductile damage evolution and failure. More spall experiments are planned, as are experiments to investigate spall in shocked plutonium. To appreciate the complexity of this problem, consider that the first studies of material failure have been attributed to Leonardo da Vinci. Nearly five hundred years later we are still at the beginning; however, with better models, new diagnostic techniques, and well-controlled experiments, the future looks promising. ■



**Figure 3. EBSD Images of Incipient Spall in Tantalum**  
 (a) The EBSD micrograph of the sample in Figure 1(a) confirms the absence of any significant strain localization between individual voids. (b) A fragment of the damaged sample in Figure 1(b), which was exposed to the higher stresses, exhibits continuous and tortuous (black) features between two voids. We equate those features with strain localization. (c) In a magnified view of (b), one large grain near the region of strain localization (overlaid with an arrow) displays a particularly high level of misorientation as represented by the continuous change in color. (d) The graph plots the misorientation angle along the arrow measured from right to left in step sizes of 1  $\mu\text{m}$ . The blue curve shows orientation changes from the origin of the arrow to points along the arrow, and the red curve shows point-to-point changes in orientation. (e) The color key correlates color with crystallographic orientation in the individual grains.

## Further Reading

- Adams, B. L., S. I. Wright, and K. Kunze. 1993. Orientation Imaging—The Emergence of a New Microscopy. *Metall. Trans. A* **24** (4): 819.
- Tonks, D. L., A. K. Zurek, W. R. Thissell, J. E. Vorthman, and R. S. Hixson. 2002. “The Tonks Ductile Damage Model,” Los Alamos National Laboratory document LA-UR-03-0809.

For further information, contact  
 Anna Zurek (505) 667-4040  
 (nesia@lanl.gov).



# Shock Compression Techniques for Developing Multiphase Equations of State

*Robert S. Hixson, George T. Gray, and Dennis B. Hayes*

**A**n airplane accidentally hits a mountainside, a bird gets sucked into the turbine blades of a jet engine, a meteorite strikes a satellite or a planet, explosives blast a cavity in the earth for a building foundation, a hammer strikes metal to forge a new part, a blast wave from a nuclear explosion strikes objects in its path—in both man-made and natural settings, shock waves and impacts produce strong impulsive loading or sudden increases in external stress. In each case, the scientist or engineer would like to predict the response of the material to that dynamic loading. How much does the density increase? Does the material heat up? Does it melt?

The key to answering some of these questions lies in knowing those intrinsic properties of materials characterized by their equations of state. Like the familiar equation of state (EOS) of an ideal gas,  $PV = nRT$ , the equation of state of a material specifies a definite relationship between three thermodynamic variables, pressure  $P$ , temperature  $T$ , and volume  $V$  (or density  $\rho = \text{mass}/V$ ). Thus, it is not possible to adjust the three variables independently. Rather, if the density and temperature of a material are fixed, for example, then its pressure (as well as energy, entropy, and all other thermodynamic quantities) is a unique value determined by its EOS.

Los Alamos scientists have a long tradition of using dynamic loading techniques to develop the equations of state that describe solids and liquids at extremely high temperatures and pressures. During the Manhattan Project, Hans Bethe, Geoffrey I. Taylor, Cyril S. Smith, and others developed seminal theories of material response to shock wave compression. After World War II, experimentalists Stanley Minshall, John M. Walsh, Robert G. McQueen, and others developed plate impact techniques to make much more precise measurements of equations of state. Weapon designers use those equations of state, as well as others developed more recently, to improve the fidelity of their large-scale computer simulations of nuclear weapon designs.

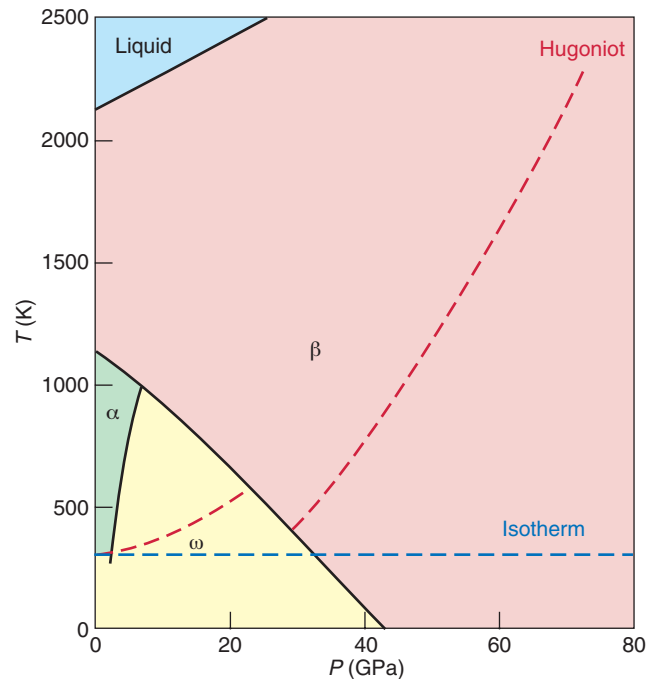
Today the goal is to perform high-fidelity simulations that predict nuclear weapon performance and safety under a wide variety of scenarios. To achieve the required level of confidence, weapon designers need equations of state that faithfully account for the complex behavior of plutonium, uranium, and many other metals when they are dynamically compressed.

Under dynamic loading, metals can change not only from solid to liquid and liquid to vapor but also from one solid crystal structure to another. Here

we describe how we are using new shock-compression techniques involving the preheating of materials to map out the boundaries between solid–solid and liquid–solid phases. We also outline how we use that information to construct sophisticated, semiempirical multiphase equations of state from which we can predict responses of materials in complex geometries, responses that have not been or cannot be directly measured experimentally. These equations of state help with our own interpretation of experiments and contribute to the development of other, more comprehensive equations of state for use in weapon design codes.

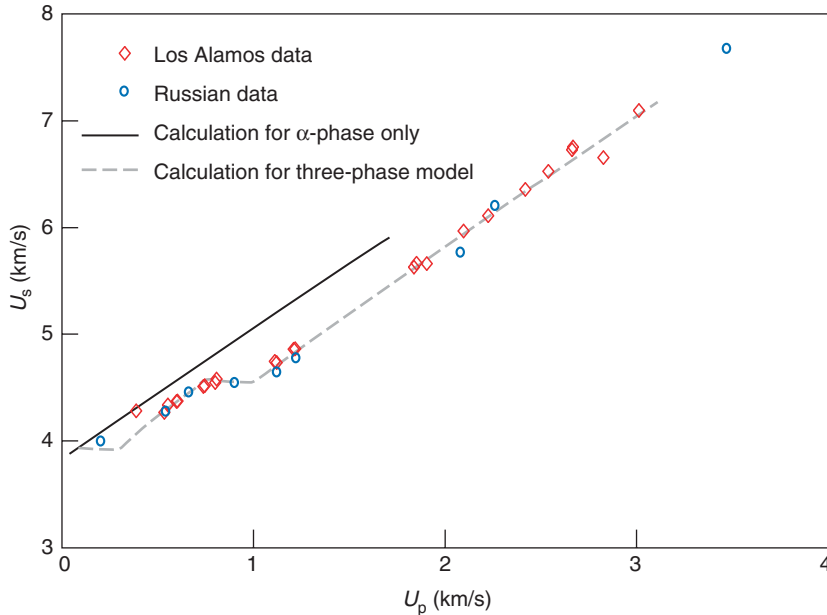
## Zirconium Phase Diagram

To illustrate the development of an EOS, we will consider our work on the EOS of zirconium, a heavy element between titanium and hafnium in the group 4B metals. Figure 1, the phase diagram of zirconium, shows current best estimates for the phase boundaries of zirconium in the pressure–temperature ( $P$ – $T$ ) plane. At a constant pressure of 1 atmosphere, zirconium, which is a hexagonal, close-packed (hcp) structure at room temperature, will change to a less-dense, body-centered-cubic (bcc) structure when heated above 1136



**Figure 1. Zirconium Phase Diagram**

This diagram by Carl Greeff of Los Alamos shows current best estimates for the phase boundaries of zirconium in the pressure–temperature ( $P$ – $T$ ) plane. The estimates are based on shock compression work and analyses from the literature. Boundaries separate the three solid phases— $\alpha$  (hcp),  $\omega$  (hex), and  $\beta$  (bcc)—and the liquid phase. Each  $P$ – $T$  point along a phase boundary defines a state in which mixtures of the two phases can coexist in equilibrium. The locus of  $P$ – $T$  states in which a liquid and a solid phase can coexist is called the coexistence curve, the melt curve, or the phase boundary. The locus of states in which two solid phases can coexist is called a coexistence curve or phase boundary. Two coexistence curves intersect at a “triple point,” and at that temperature and pressure, three phases can coexist. Also shown is the principal Hugoniot, a locus of end states reached by shock compression starting from room temperature and pressure.



**Figure 2. The Zirconium Shock Hugoniot**

The shock Hugoniot for zirconium, a plot of shock velocity ( $U_s$ ) vs particle velocity ( $U_p$ ), was determined from Los Alamos and Russian shock-compression measurements and from calculations (dashed curves) using the multiphase model developed by Greeff. The cusps in the data indicate that successively higher-impact stresses bring the metal to final states with different crystalline phases: the  $\alpha$  (hcp),  $\omega$  (hex), or  $\beta$  (bcc) phase. This shock Hugoniot was used to calculate the  $P$ - $T$  Hugoniot displayed in Figure 1.

kelvins. At still higher temperatures, it melts. Similarly, if the metal is kept at room temperature, the hcp phase of zirconium will transform to a new, higher-density hexagonal phase when the pressure is increased. (Note that for some metals, the high-pressure phase is less dense than the low-pressure phase.)

### Using Shock Waves to Develop Equations of State

Also shown in Figure 1 is the “principal” Hugoniot<sup>1</sup> for zirconium. This dashed curve is the locus of end states that can be reached through shock wave compression. The Hugoniot rises steeply in the  $P$ - $T$

<sup>1</sup> This curve is named after the nineteenth century French scientist.

plane, whereas an isotherm, a locus of states reached through static compression (slow increase in pressure) with temperature held constant, is by definition, flat in the  $P$ - $T$  plane. Both the Hugoniot and the isotherm are useful in developing an EOS for a material, but they must be combined with other thermodynamic information, such as the material’s heat capacity. In general, phase diagrams and properties of the pure phases can be experimentally determined with either static or dynamic techniques. Results may differ because of the differences in the experimental time scale. Sometimes, those differences are explained by time-dependent equilibration, but other differences reflect the fact that a material responds differently to high strain rates than it does to low strain rates.

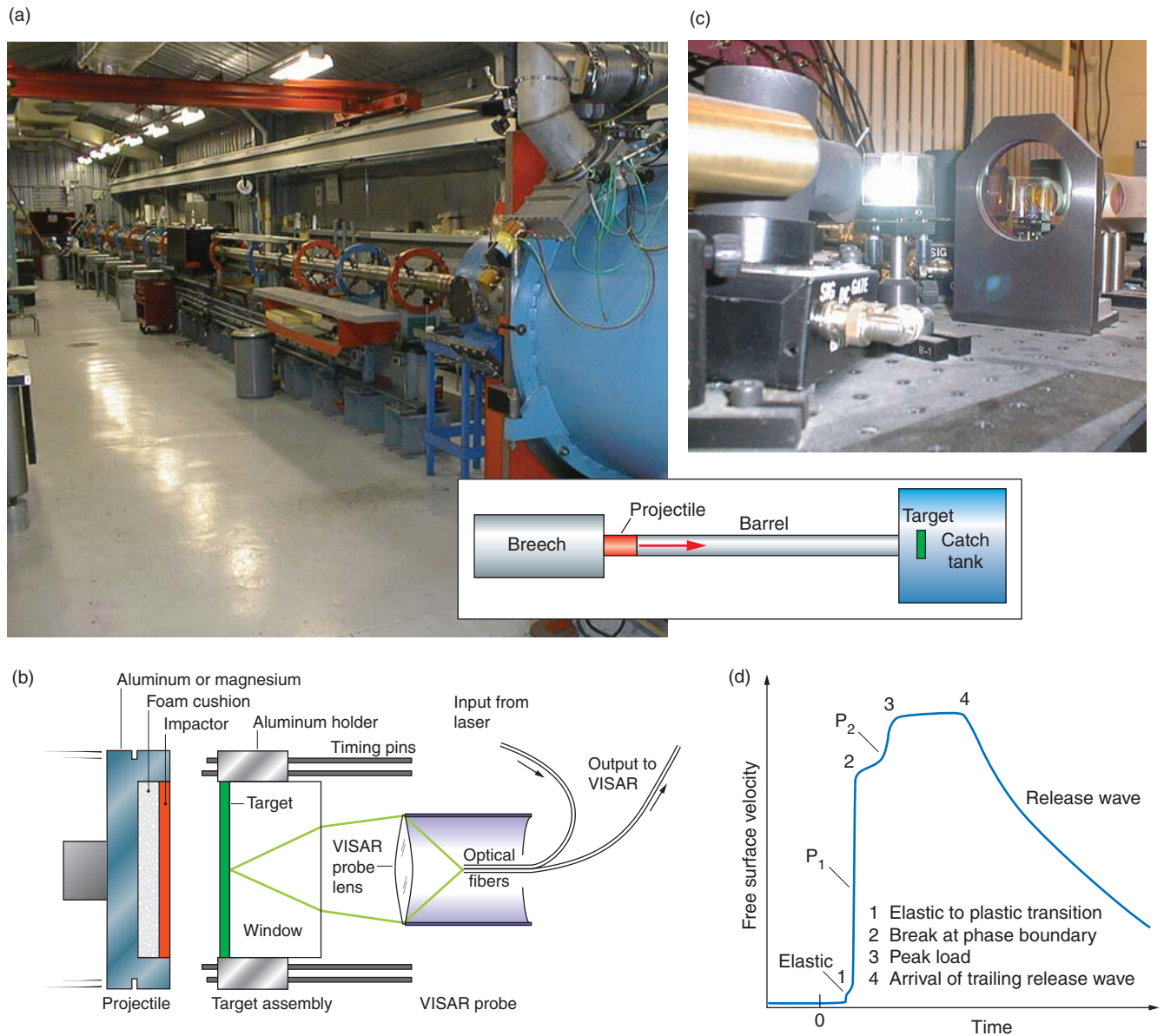
For the Hugoniot shown in

Figure 1, the temperatures were not measured directly but were calculated from the EOS that we constructed using the Los Alamos and Russian data displayed in Figure 2. This representation of the shock Hugoniot is a plot of shock velocity ( $U_s$ ) versus particle velocity ( $U_p$ ). Each cusp, or sudden change in slope, signals the location of a solid–solid phase boundary. The cusp forms because the shock pressure is sufficient to induce a phase change, and the resulting density change causes the shock velocity to change and the single shock to split into two shocks, one following the other.

Figure 3 describes how we use time-resolved laser-interferometric techniques to locate the position of that cusp, or solid–solid phase change, in the  $P$ - $T$  plane. Work is under way to use different initial temperatures of the material and thereby map out all the points on the phase boundaries in Figure 1.

The gas gun facility where we perform the experiments is shown in Figure 3(a). The gas gun accelerates a projectile carrying a thin impactor toward a flat, very thin zirconium target held in place by its edges—refer to Figure 3(b). At the back of the target is a laser probe connected by fiber-optic cables to a VISAR<sup>2</sup> (velocity interferometer system for any reflector)—refer to Figure 3(c). By determining the frequency of the laser light reflected from the center of the target’s back surface, the VISAR determines the velocity history of that surface after impact. As explained in Figure 3(d), the velocity history of the back surface directly reflects the structure of the shock waves that have propagated through the sample and, in turn, the effects of

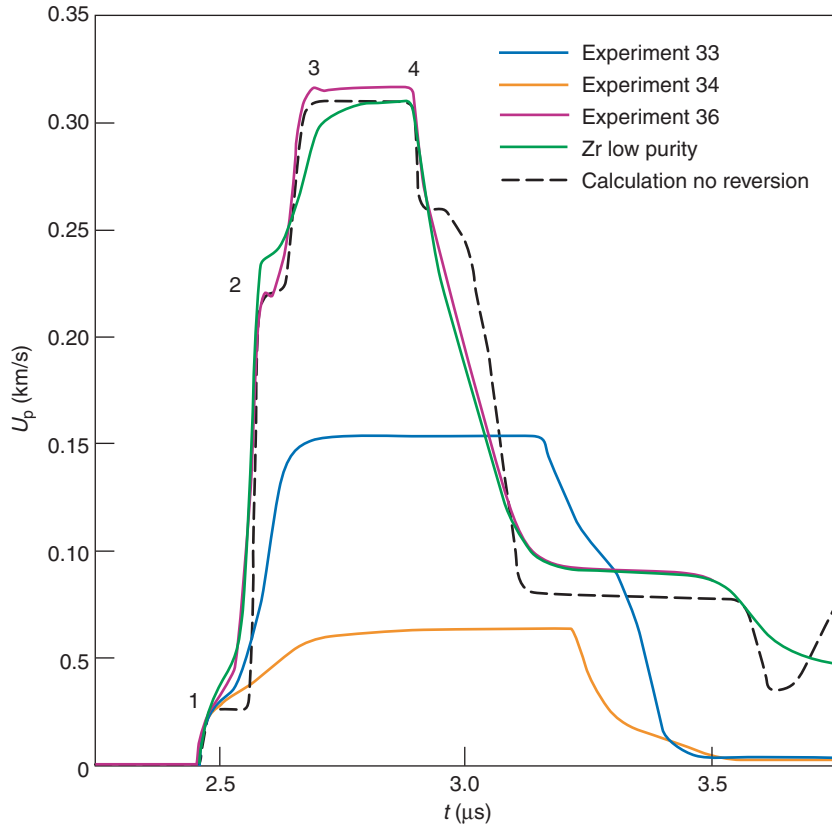
<sup>2</sup> The VISAR was invented around 1970 by Lynn Baker at Sandia National Laboratories. It is the tool of choice worldwide for shockwave work.



**Figure 3. VISAR Measurements of Shock Wave Structure**

(a) Both photo and schematic (inset) show a gas gun facility at Ancho Canyon, where we perform shock compression experiments. Typically, a projectile is accelerated by compressed gas, or sometimes by gunpowder, down a long barrel (10 to 40 ft long) and impacts a stationary target at speeds of 0.1 to 8 km/s. (b) In our equation-of-state experiments, a flat, thin impactor carried by the projectile strikes a flat, very thin zirconium sample and sends a compression wave through that target. A laser probe at the back of the target focuses laser light from one optical fiber onto a spot at the center of the zirconium plate. The light reflected from the moving surface is focused onto a second optical fiber that leads back to a laser-velocity interferometer, called a VISAR, located in the recording room (c). The VISAR determines the reflected light's frequency shift as a function of time and thus the velocity history of the back surface of the zirconium sample. The wafer is so thin and

the aspect ratio so large (say, 40 to 1) that measurements take less than  $2 \mu\text{s}$  following impact, before any edge effects could travel to the target center and affect the measurements. This technique determines the time-resolved velocity history of a moving surface with the times accurate to  $\pm 1 \text{ ns}$  and the velocities accurate to  $\pm 1\%$ . (d) The shape of the surface velocity history reflects the propagation of the shock wave through a multiphase material. Between 0 and 1, an elastic wave reaches the target's back surface; between 1 and 2, a plastic (deforming) wave called P1 increases the pressure to the point 2 where the material begins to change phase. Between 2 and 3, a second plastic wave (P2) increases the pressure slowly as more of the material changes phase until the peak load is reached at 3. Finally, at 4, a trailing release (rarefaction) wave, initiated by reflection of the shock wave from the impactor's back surface, arrives at the target's back surface.



**Figure 4. Measured Wave Profiles at Different Impact Stresses**

Under high impact stresses (strong shock compression), the (particle velocity) wave profiles for high-purity zirconium (red) and low-purity zirconium (green) display features labeled 1–4, corresponding to the features described in Figure 3(d). In particular, the cusp at point 2 indicates the stress (or pressure) at which a solid–solid phase transformation begins in the material, producing a second plastic wave P2. Note that low-purity zirconium changes phase at a higher pressure than high-purity zirconium, and its phase transformation is more sluggish—the pressure rises more slowly as the phase transformation proceeds between points 2 and 3. The impact stresses reached in Experiments 33 and 34 (also done on high-purity zirconium) were lower than the pressures at which phase change begins, so the wave profiles are smooth, indicating no phase change. These profiles were measured by Paulo Rigg of Los Alamos.

shock compression on the material. In particular, if shock compression causes the material to change to a denser phase (for example, to cross the  $\alpha$ – $\omega$  boundary or the  $\omega$ – $\beta$  boundary in Figure 1), the shock wave is unstable and breaks into two shock waves, or deforming/plastic waves, called P1 and P2. The first deforming wave P1 has a pressure corresponding to the stress at which the principal Hugoniot

intersects the phase boundary in Figure 1. It brings the material to point 2 in Figure 3(d). The second, slower plastic wave P2 brings the material to its ultimate loading stress shown as point 3 in Figure 3(d). The detailed shape of the transmitted compression wave, therefore, contains information about the location of one point on the phase boundary. The small initial elastic wave (between

point 0 and point 1) complicates interpretation of this record and is discussed below.

By changing the initial temperature of the sample, we can shift the starting point of the Hugoniot curve so that it will cross the phase boundaries at a different longitudinal stress in the material and thereby map out all the points on the phase boundaries. Such techniques for preheated shock compression are currently being developed at Los Alamos.

### Multiphase Equations of State

**Development.** After locating the boundaries between different phases of a material, we use that information to help develop sophisticated, thermodynamically consistent equations of state that take into account some or all the possible structures of the material. We then use these multiphase equations of state to do high-fidelity computer simulations of experiments that involve dynamic loading.

We model the EOS for each pure phase with, for example, a semiempirical, analytical form for the Helmholtz free energy (HFE) and determine the parameters in the model from shock wave experiments, isothermal compression data, and any other available data. To get a thermodynamically complete HFE in the mixed-phase region, we must also specify the entropy and energy at one reference point. We know that two phases have the same Gibbs free energy (GFE) along their coexistence curve. Therefore, by requiring that the calculated coexistence curves (that is, the points at which the calculated GFE for the two phases are equal) match the measured phase boundaries, we can determine uniquely all the pure-phase reference energies and entropies.

**Application.** To implement a computational technique (say, a hydrodynamic calculation involving shock loading) that accounts for the behavior of mixtures of phases, we must make some additional assumptions. One-dimensional shock wave experiments measure longitudinal stress, whereas the thermodynamic properties depend upon pressure. For solids, longitudinal stress and pressure are different. In most cases, it is satisfactory to treat strength effects as being elastic/perfectly plastic, which is our usual assumption. We further assume that crystallites of the individual phases in a mixture are small enough that pressure and temperature are locally equilibrated, although the mixture need not otherwise be in thermodynamic equilibrium. We also assume that the shock response is rapid enough that all processes are adiabatic although not isentropic. Problem closure is achieved if rules are specified for the rates of transformation between phases. Because our goal is to interpret experimental results, we use semiempirical rules. In our most elementary models, we assume that the transformation rate between two phases is proportional to the calculated GFE difference between the two phases and inversely proportional to a characteristic time for that particular transformation. Our experiments show that characteristic times for a forward phase transformation (from a low- to a high-pressure phase) are not always the same as those for the reverse transformation. We have shown this numerical approach to be quite robust because it easily handles very complex, nonequilibrium mixtures of many phases during computations of wave propagation in a phase-changing material.

Systems with complex geometries often defy direct measurement of the details of the response to impulsive loading so that often the only alternative is to develop computational mod-

els. Many materials in systems of interest undergo multiple, and sometimes nonequilibrium phase changes, when they are shocked. The equations of state and locations of phase boundaries of these materials are measured in simple experiments and the behavior is captured in a multiphase EOS. Then modeling provides the necessary bridge between the world of physics and application.

Transmitted wave profiles that are measured in these experiments contain much more information on material behavior than just wave speeds and locations of phase boundaries. Most materials display a variety of nonequilibrium effects that are prominent in shock and release experiments. Some phase changes are sluggish, and this feature is reflected in the rather broad rise time in Figure 4. (See the broad rise time between points 2 and 3 compared with the sharp rise time between points 1 and 2.) The forward transformation sometimes slows abruptly after only partial completion, clearly a nonequilibrium phenomenon, and this affects the speed of the wave between points 2 and 3 in Figure 4. There may be big differences in the speeds of the forward and reverse transformations, which will affect both the peak particle velocity and the detailed structure of the release wave. Space limitations do not allow us to go into these aspects in any detail. But it is important to recognize that these experiments provide unique data on phase change kinetics and nonequilibrium effects that are invaluable for generation or validation of fundamental theories of the phase change processes at high strain rates. ■

### Acknowledgments

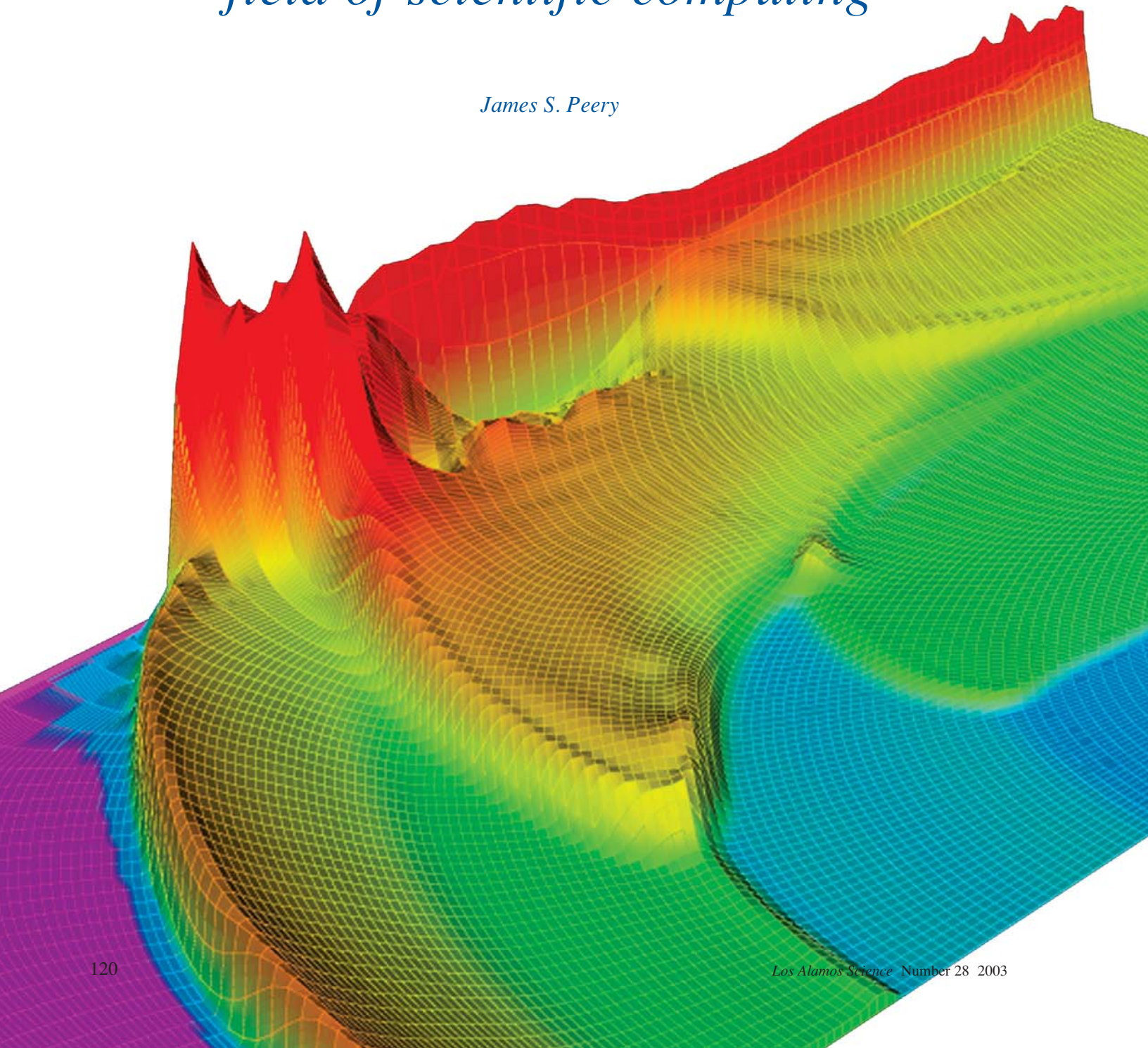
We gratefully acknowledge useful suggestions from John Vorthman, Carl Greeff, Paulo Rigg, and Bill Anderson.

*For further information, contact Robert Hixson (505) 667-1964 (hixson@lanl.gov).*

# ASCI

*Big engineering in the  
field of scientific computing*

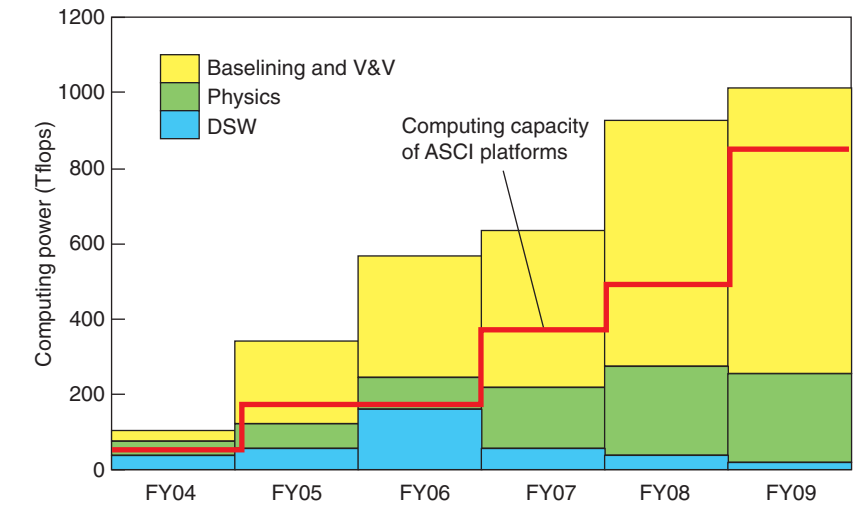
*James S. Peery*



Imagine that, during routine inspection, we open up a weapons system and find a significant change that we did not expect. Or imagine the numerous possible orientations and conditions that a weapon might assume in a fire. These scenarios represent real possibilities. Will the weapon work as intended? Will the weapon be safe under almost preposterous conditions? These are the questions that the Los Alamos Advanced Simulation and Computing program will help to answer by providing the weapons designers with high-fidelity simulation capability on the world's most powerful computers.

The Advanced Simulation and Computing program evolved from the merging of the Accelerated Strategic Computing Initiative (ASCI), begun in 1996, and the ongoing stockpile computing program known as the Advanced Simulation and Computing Campaign. Continuing to use the acronym ASCI, this effort is perhaps the largest and most encompassing computational development program in the world. Its core mission is to provide simulation tools, including both the hardware and the software application codes, that enable the weapons designers to assess and certify the safety, performance, and reliability of the enduring nuclear weapons stockpile. As such, ASCI is a pillar of the Science-Based Stockpile Stewardship (SBSS) program. The success of ASCI, however, will have an even larger significance, by demonstrating that large-scale computational science can create potent tools to address many scientific challenges.

In a *Popular Science* report summarizing 15 years of big engineering, nine major construction projects were cited among which were the Toronto SkyDome, the Eurotunnel, and the Petronas Towers. These are multibillion dollar, multiyear projects involving multidisciplinary teams. ASCI is the first scientific software project to



**Figure 1. Computing Needs and the Integrated Delivered ASCI Platform Capacity**

The anticipated computing needs for the Los Alamos and Livermore weapons program are divided into three categories: (1) direct stockpile work, or DSW (stockpile simulations, which assess stockpile issues, are part of refurbishment and annual certification); (2) physics (studies to increase understanding of weapons simulations); and (3) baselining and validation and verification, or V&V (creating 2-D and 3-D validated weapons models). The orange line represents the maximum capacity currently planned for the ASCI platforms. This integrated, delivered ASCI platform capacity is seen to fall short of the anticipated needs.

have a similar level of investment and a similar multi-institutional, multidisciplinary approach. Although the three labs involved, Los Alamos, Sandia, and Lawrence Livermore National Laboratories, develop their application codes independently, they work jointly on issues of computer science and hardware, on the testing of the application codes, and on developing visualization tools at scales never before attempted.

Although the ASCI program is well known for buying the world's most powerful computers, less than one-sixth of the total budget is spent on hardware. The major fraction of our effort goes into software development—simulation codes that faithfully model the end-to-end performance of a nuclear weapon. These multiphysics codes, validated through comparison with experiment and archival nuclear-weapons test data, represent the enduring product of the program.

Nuclear weapons are complex systems. During performance, materials change from solids to hot, dense plasmas, and physical processes operate on many different length and time scales. In order to produce predictive simulations of weapons performance, the codes must be built from accurate models of these physical processes and material behaviors validated through comparison with experiment. Further, the algorithms that represent these models must be both robust and computationally efficient, and they must be verified on simple problems by comparisons between numerical results and known solutions. Validation and verification are necessary to demonstrate the accuracy of these codes but are not sufficient to ensure their utility. The codes prove their usefulness when designers are able to set up problems rapidly, produce results in a reasonable time, and see the results in a form that can be



easily and quickly interpreted. To satisfy these additional requirements, ASCI is making significant investments in developing visualization and other enabling tools and in production support for hardware and software.

At present, ASCI has responsibility for providing the computing resources (that is, cycles) for both the near-term needs of stockpile stewardship and the long-term development and application of high-fidelity simulation capability. More detailed physical models, coupled with higher resolution and three-dimensional (3-D) rather than two-dimensional (2-D) simulations, are projected to greatly increase the need for computing capacity. Figure 1 compares the anticipated computing needs for SBSS with the integrated delivered capacity based on the current ASCI computing-platform procurement schedule. Needless to say, the ASCI platforms alone will not provide the computer cycles required to meet the various demands of SBSS.

## Beginnings of ASCI

At its inception in 1996, the ASCI program was conceived as an effort to accelerate the development of new, more-predictive weapons simulation tools. When supported by necessary computing resources, those tools would be able to support long-term stewardship of the stockpile in the absence of nuclear testing. To understand the magnitude of this undertaking, one needs to look at stockpile computing before ASCI. In the 1980s, coarsely resolved 2-D calculations might run for thousands of hours on the world's most powerful computers. Crays were the mainstays of production computing. After a decade of use, Crays had stable and well-understood vector architectures. Hundreds of those computers were in use around the world, although Los Alamos and Lawrence Livermore National

Laboratories prided themselves on acquiring the first serial number of each latest model.

Before ASCI, the weapons codes (which are now referred to as "legacy" codes) were built by small development teams to support the day-to-day needs of the design community. The legacy codes matured by being applied to one-dimensional and 2-D problems whose timely solution was needed for planning and designing underground tests. Such tests had many goals, among which were certifying new designs, performing physics experiments, and confirming stockpile confidence. The heavy test schedule limited the time that could be spent on fundamental improvements; instead, legacy codes were calibrated to the underground test data with nonphysical parameters, sometimes termed knobs. This process produced useful engineering tools for interpolation, but their predictiveness for extrapolation was indeterminate. In other words, the success of that code development strategy depended on continued testing. The interaction of modeling and experiment is part of the scientific method. However, the political decision to cease nuclear testing required an immediate and urgent change of strategy—one result was ASCI.

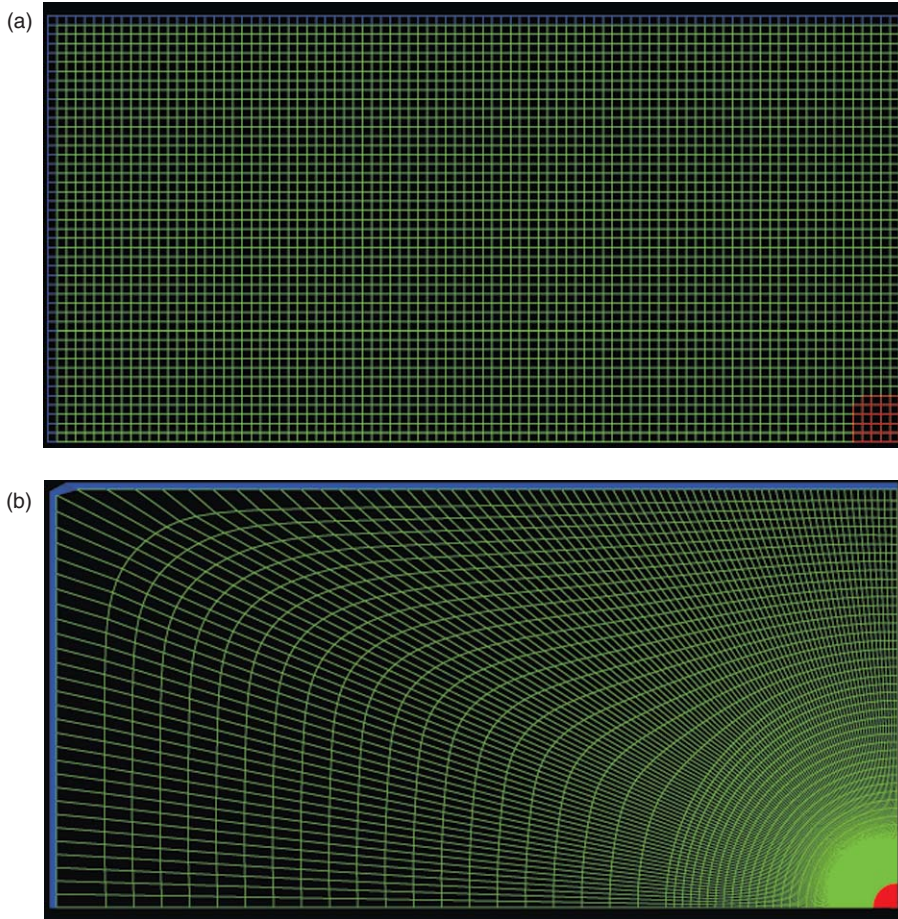
Because the legacy codes can reproduce the results of underground tests, albeit, not from first principles, they are a direct link to the past and remain important to weapons designers. However, the design community also needs codes built from better physics models to assess the effects of aging components within the weapons, newly identified safety concerns, and other stockpile issues. The architects of ASCI understood that the new, more predictive codes would require huge increases in computing capability. Indeed, the program would have to revitalize the high-performance computer industry if high-fidelity simulations of the complex physics

inside a nuclear weapon were ever to be practical. Initially, the program decided to focus on achieving long-term predictive capability at the expense of supporting short-term designer needs.

The vision was sold, and the planning began. The end goal of ASCI became the construction of new, high-fidelity, verified, and validated 3-D codes. High fidelity implies that the codes contain first-principles physics models and accurate, efficient numerical algorithms that produce converged solutions. Without fully understanding the magnitude of this vision, ASCI set out to develop 3-D codes capable of unprecedented resolution of physical processes in space and time. It was not long before the requirements were collected and the enormous complexity of the undertaking became clear. However, faced with the cessation of the underground testing and confronted by a rapidly aging weapons design community, management saw an urgent need to develop these more predictive tools and to train a new generation of designers as quickly as possible. Thus, the program grew at a rapid rate.

## Developing Codes for Massively Parallel Computers

To simulate 3-D weapons system performance with high resolution and with reasonable turnaround times, one needs computers with  $10^5$  to  $10^6$  times more power than the Cray YMPs used at the end of the underground test period. The only type of architecture capable of delivering such power is a massively parallel computer in which at least 10,000 processors can be applied simultaneously to solving a problem. ASCI generated a multiyear, multiplatform plan to achieve that goal, which should be realized in 2005. The latest ASCI platform, the Q machine at Los Alamos, is approximately  $10^4$  times faster than a Cray YMP.



**Figure 2. Eulerian and ALE Meshes for Simulating a Pulsed-Power Experiment**

The meshes in (a) and (b) represent the computational domain in 2-D cylindrical geometry for a liner experiment at the Atlas pulsed-power facility. (See Figure 6 on page 74 in the article “The New World of the Nevada Test Site” for a description of such experiments.) The stainless steel containment vessel (blue) is separated by air (green) from the liner assembly (red). The regular Cartesian mesh in (a) is suitable for Eulerian hydrodynamics calculations. The boundary-fitted curvilinear mesh in (b) is suitable for ALE hydrodynamics calculations.

Effective use of massively parallel platforms demands new algorithmic strategies. ASCI began development of both a new generation of codes employing parallel algorithms and the associated setup and visualization tools. The more predictive physics models that provide the building blocks require the solution of nonlinear partial differential equations involving multiple scales of length and time. The equations of the individual models can rarely be solved in

closed form. They must therefore be solved approximately on the computer. These approximations are based on discretization methods such as finite difference or finite elements. Discretization means that the computational domain is divided into discrete volumes, or cells, that are organized by a mesh (see Figure 2). Solution variables, such as density and temperature, are averaged over the cells. This operation effectively reduces the number of unknowns in the problem to a

level that the computer can handle. Discretization represents a tradeoff between the accuracy and required completion time of a simulation. Accuracy increases as the square of the number of cells (for example, for second-order algorithms). The required work, which is proportional to the problem time, increases as the fourth power of the number of cells in three dimensions. Therefore, the accuracy divided by the work, or the efficacy, is a strongly decreasing function of the number of cells. Computer power can be traded off for longer run times, but the run times can quickly become unacceptably long if the problem is very large. Ultimately, the accuracy of a simulation is limited by computer speed and the time one is willing to wait for an answer. And, of course, the discretized problem must fit into the available memory of the computer.

From a physics viewpoint, the models depend on experimentally measured properties, and these imply scales of length and time that must be resolved if the simulation results are to be valid. Furthermore, the individual models are coupled to each other, and their collective behavior is more complicated than the sum of their individual behaviors. This complication is ignored in the legacy codes, a simplification termed operator splitting (see the article “Massively Parallel Multiphysics Code Development” on page 128), but recent research has indicated that this simplification is a poor approximation. In other words, resolving the individual physics models is part of verifying the algorithms, but that step does not guarantee getting the right answer. Verification must be followed by validating the full multiphysics code system against experimental data.

Setting up the mesh to represent the initial geometry of a weapons system in three dimensions can be a for-

midable challenge. There are two basic frameworks for solving the hydrodynamic equations that describe the motion of materials. Known as Eulerian and Lagrangian, they utilize different types of mesh (see Figure 2). Eulerian algorithms solve the equations on a mesh that remains fixed in space while the material flows through it. Lagrangian algorithms solve the equations on a mesh that moves with the material. Each method has advantages, and therefore ASCI set forth to develop both. Redundancy, in the sense of multiple independent approaches to code development, has long been a staple of the nuclear weapons program.

In general, we use computer-aided design (CAD) software to generate the 3-D geometries of weapons systems. But CAD software was designed for manufacturing applications; consequently, the CAD setups suffer from incompleteness, overlapping parts, and unnecessary detail and are therefore ill suited for ASCI set-up applications. To overcome these deficiencies, ASCI set forth to develop 3-D meshing algorithms for Lagrangian-based codes and volume filling techniques for Eulerian-based codes. Indeed, the regularity of the Eulerian meshes has already allowed us to create effective setup tools. Meshes for Lagrangian codes need to reflect both the initial geometry and the subsequent material motion. The tendency of imperfect meshes to tangle and thus bring the simulation to a premature end is a more difficult problem to overcome, and the lack of adequate setup continues to limit the use of Lagrangian codes in three dimensions. An example of a 2-D calculation using arbitrary Lagrangian-Eulerian (ALE) techniques, a method that combines the advantages of Eulerian and Lagrangian approaches, is shown in Figure 3. In particular, a rezone and remap procedure is added to a Lagrangian algorithm.

In addition to being accurate and robust, our solution algorithms must scale on parallel architectures. In other words, at a minimum, if we increase the domain of the problem by two and the number of processors by two, we want the problem to run for the same time. Historically, scaling has been achieved in mesh-based algorithms (Eulerian and Lagrangian) by dividing the domain of the problem into small chunks or subdomains. Each chunk runs on the memory attached to one processor. When the

*Only by continual investment in fundamental science can we create the realistic models and predictive capability that will enable code developers and weapons designers to address the problems presented by the aging nuclear stockpile.*

equations call for information from more than one subdomain, the processors communicate through a communication network. The efficiency of a massively parallel computer depends on our ability to minimize this communication time. This requirement has mandated significant investments in high-speed network technology, effective domain decomposition software, and parallel algorithm development.

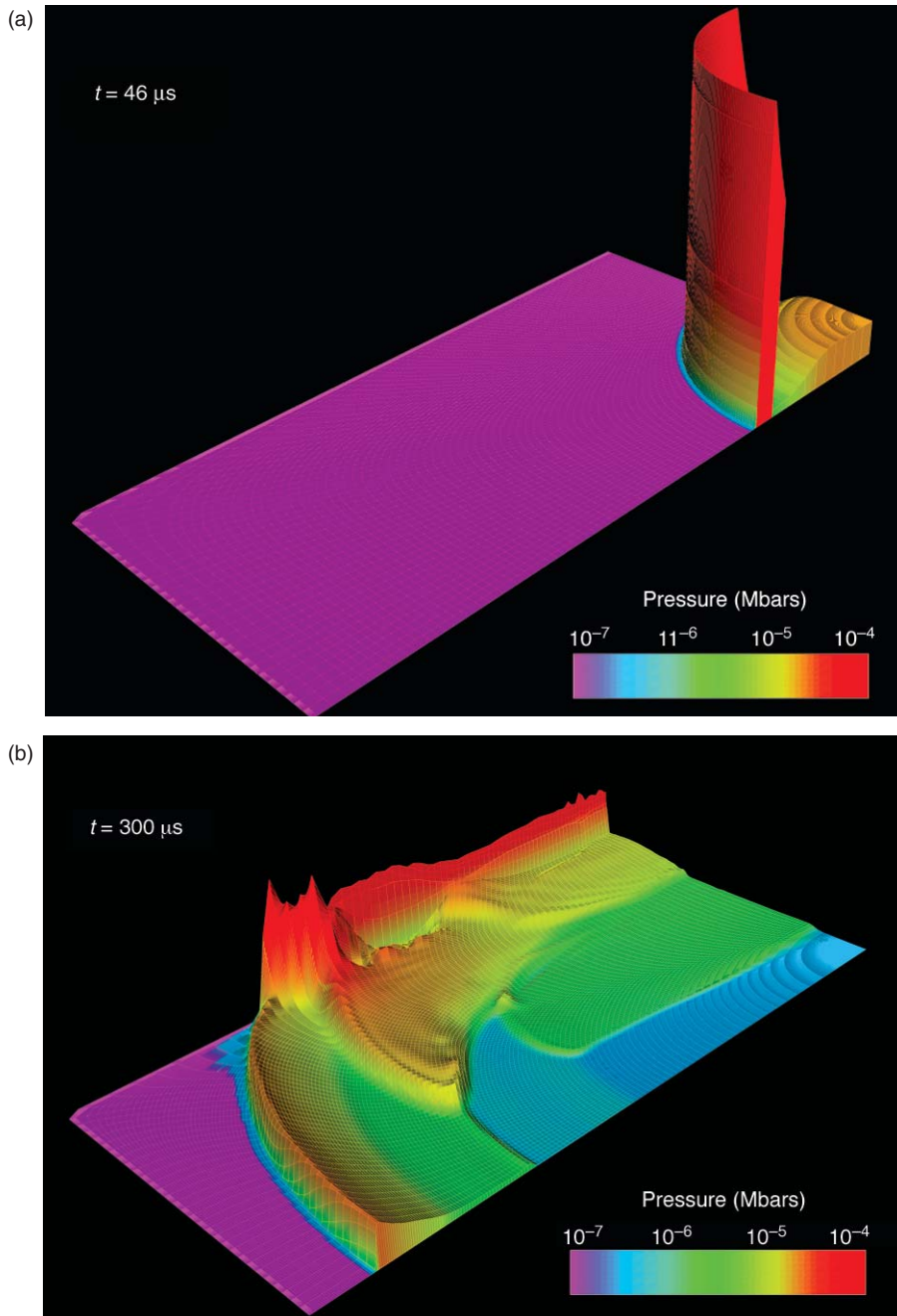
The process of discretization described above transforms the partial differential equations into a large system of algebraic equations. Although standard techniques for solving these matrices have existed for many years, the systems resulting from 3-D ASCI algorithms are too large for these matrix solution techniques and thus too expensive in time for ASCI-class computers and architectures. Therefore, major efforts were initiated at all three national laboratories to develop new techniques that reduce

the solution time. The results of these early efforts are beginning to pay off. In some areas of physics, new matrix solution techniques are not only effectively solving the large equation systems but also scaling well with problem size.

As ASCI identified new research areas such as those described above, it responded by allocating resources, forming research teams, and in many cases initiating trilaboratory collaborations. The scope of ASCI grew rapidly. Large teams of code physicists and computer scientists were assembled to write the physics codes. It is not unusual for a physics code team leader to represent a team of 20 or more developers and to interface with dozens of other teams who are producing software libraries or hardware relevant to the project. Teams of computational physicists, tasked with developing new algorithms and solution techniques, produced libraries for the physics codes. Teams of computer scientists developed tools, message-passing protocols, and encryption and system software for use on the massively parallel computers. Teams of engineers and computational scientists tackled the issues in problem setup and domain decomposition. Teams of hardware and software engineers and scientists developed tools to move vast quantities of data from disks to leading-edge, 3-D stereo display platforms for both office and custom-designed laboratories and theaters (see the article "A Vision of Hidden Worlds" on page 135). ASCI responded quickly to the technical challenges, but the ensuing growing pains are still being felt.

## ASCI Report Card

The mission of ASCI has evolved significantly since its inception. The first five years, from 1996 to 2001, were mostly directed at proof of principle. The goals were very ambitious,



**Figure 3. ALE Hydrodynamics Calculation of the Shock Structure inside Atlas Device**

These pressure plots follow a pulsed-power liner experiment in the computational domain shown in Figure 2. At  $t = 0$ , 20 megajoules of energy, the maximum deliverable from the Atlas capacitor banks, is deposited into the liner assembly. (a) At  $t = 46 \mu\text{s}$ , a sharp, spherical pressure shock is propagating outward in the air. (b) At  $t = 300 \mu\text{s}$ , before the initial shock has reached the top of the containment vessel, multiple reflections of the shock have taken place between the sidewall of the containment vessel and the expanding liner assembly. These plots demonstrate the complex wave interactions and material flows that can be simulated using an ALE code as the mesh follows the material flows and is then readjusted to avoid tangling. The sizes and shapes of the mesh cells vary as the calculation proceeds.

but the program was short on requirements. Today, the focus has changed dramatically to deployment and to support of the new user-oriented tools. The end-to-end needs of the design community are driving program priorities and new activities. While ASCI continues its development of the new capabilities, it is also applying the simulation tools to immediate stockpile concerns.

Although the program is experiencing social engineering and project management tensions because of its rapid growth, it has engendered numerous technical success stories. At Los Alamos, the Crestone project has demonstrated unprecedented capabilities and geometric resolution through the first ever 3-D full system, end-to-end simulation of nuclear weapons performance. Moreover, codes of the Crestone project are used by more than half of the secondary design community. The Shavano project has provided a significant leap forward in its ability to model complex 3-D geometries and is gaining acceptance in the primary design community. The Blanca project has just recently completed a series of safety simulations and is being merged with the Shavano project. New physics models added to both the ASCI and legacy codes will continue to increase our predictive capability and add to our understanding of nuclear weapons. The Q machine and the Blue Mountain computer are delivering cycles to the design community and to the ASCI code development teams. Last, the computing infrastructure designed and deployed by ASCI, including both hardware and software (networks, computers, visualization displays), is facilitating the use of both the legacy codes and the new ASCI codes.

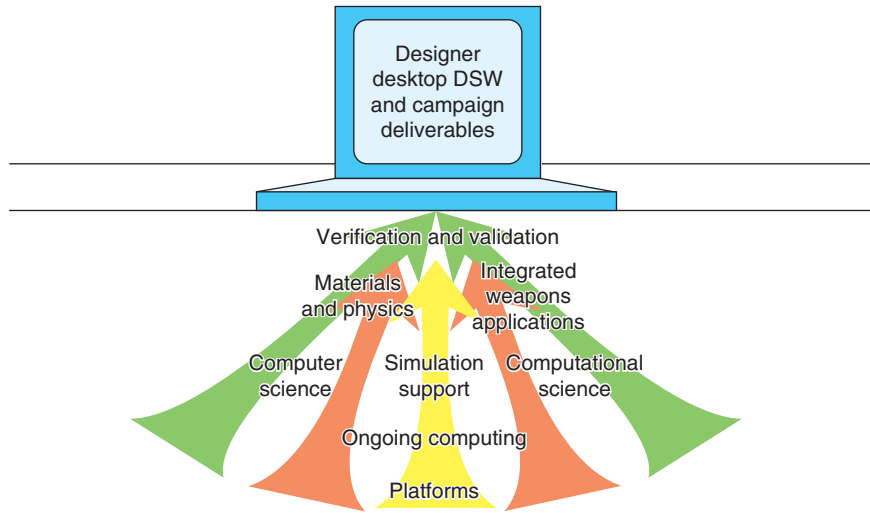
Los Alamos ASCI codes are now being used to close significant finding investigations, that is, to assess aging-weapons problems that are arising in the enduring stockpile. The ASCI

codes also support the life extension programs for individual weapons systems by providing a means to evaluate the proposed steps for extending the shelf life of our present weapons systems. Significant efforts are under way to make the transition from legacy code calculations of baseline nuclear weapons performance to ASCI code calculations of those baselines. All these activities are enabled by the continuing operation and development of the supercomputing infrastructures at the national laboratories. Research continues on new techniques for storage, visualization, networking, and all aspects of the structure required by the modern generation of computing capabilities.

ASCI's goal of maintaining a healthy high-performance computing industry has been achieved. Although some vendors have exited the high-performance computing market, many have survived, new ones have emerged, and some have reengaged (for example, Cray). In addition, other federal agencies and universities have joined the push to maintain the U.S. high-performance computing industry. When ASCI started, many doubted that a teraflop computer (capable of performing  $10^{12}$  floating point operations per second) could be built; now, through the efforts of many, ASCI has proved and enabled teraflop computing for all scientific communities.

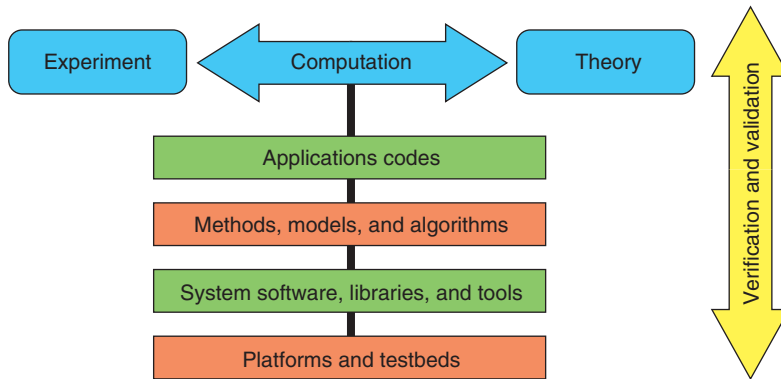
### Future of ASCI

ASCI has in place the foundations for 2- and 3-D codes based both on Eulerian and Lagrangian formulations and a computing infrastructure of more than 10 teraflops. The focus is now on integrating improved physics and engineering models into these codes and validating the codes against experimental data from both small-scale, nonnuclear integral tests and past underground nuclear tests. Once



**Figure 4. ASCI Program Elements**

Each of the eight elements of the ASCI program is shown with a line of sight to the designer's desktop. The illustration suggests that stockpile certification and design requirements will continually guide the planning and execution of ASCI program elements.



**Figure 5. Elements of Increased Predictive Capability**

The interaction between experiment and computation validates the codes and directs the course of new theoretical research.

these codes can predict the baseline performance of nuclear weapons, they will become new repositories of expert designer judgment, as well as the best scientific tools for simulating the performance of the complex weapons currently in the stockpile as those weapons age or are modified. It is widely recognized that such simulation capabilities are essential if the National Nuclear Security

Administration is to meet its statutory responsibility to assess and certify the stockpile annually. The ASCI codes will represent the ultimate integration of the theoretical and experimental efforts taking place within the stockpile stewardship program.

Inherent in the ASCI strategy is a tension between addressing the long-term simulation requirements of the weapons program and satisfying the

weapons designers' needs for short-term improvements. For the last several years, the question has been, "When will ASCI deliver on its promises to the design community?" This question started in the hallways and is now the theme in Washington. The workload associated with these competing requirements has stretched the ASCI code teams almost beyond endurance. However, a new strategy based on a more realistic assessment of current computer resources is emerging. Heroic efforts have produced prototypes of 3-D weapons-system performance simulations, but those efforts have also shown that significant increases in computer power are required before the design community can routinely run high-fidelity 3-D simulations. Thus, until suitable platforms are available, the program will focus its efforts on developing and validating a production capability for 2-D spatially resolved simulations (see Figure 3). This goal is a better match to the current parallel platforms and computing infrastructure. Efforts to validate the codes and interactions with the design community will drive the development of more predictive material and physics models. As more-capable platforms become available, we will leverage those activities toward 3-D predictive capabilities.

The new strategy is directed at satisfying the current and anticipated designer requirements. ASCI is taking conscious steps to integrate its efforts more tightly with the ongoing work of the weapons designers. For example, we are currently aligning the ASCI milestones with the work that the code users must perform in support of stockpile assessment and certification. The ASCI milestones, which are reviewed periodically by an external review committee of experts in scientific computation, will continue to ensure steady improvement in the simulation capabilities for assessing and certifying a safe, secure, and reli-

able nuclear weapons stockpile.

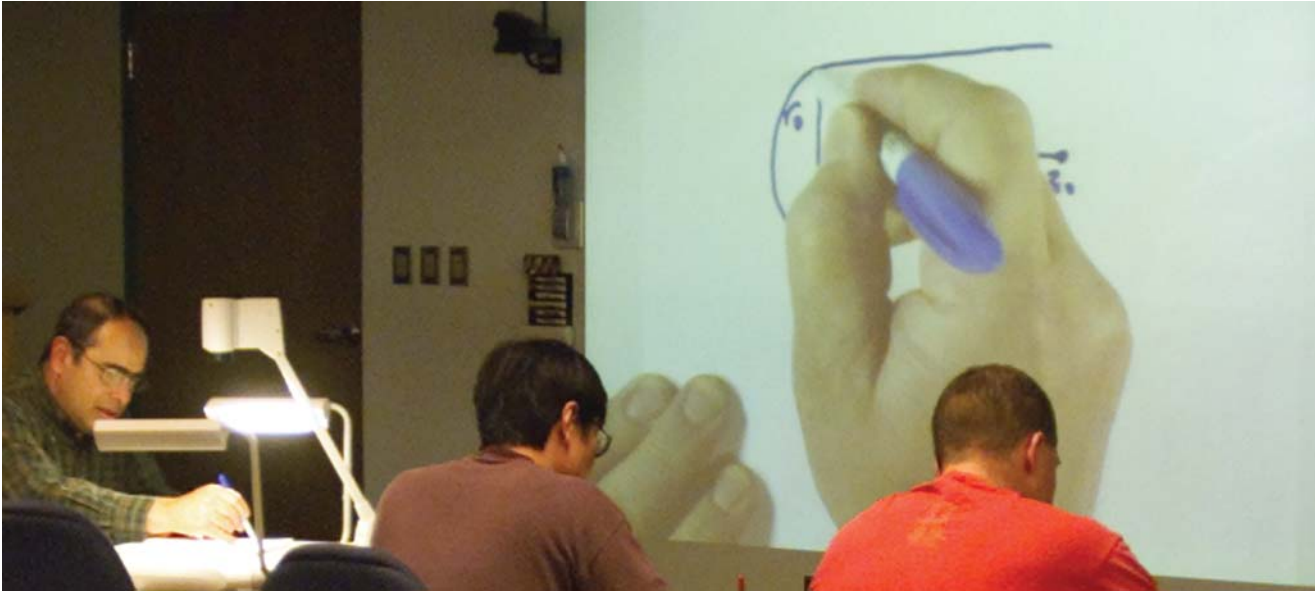
Figure 4 shows the structure for implementing the new strategy. Each of the eight elements of the ASCI program is shown with a line of sight to the designer's desktop, meaning that stockpile certification and design requirements will continually guide the planning and execution of ASCI program elements.

Despite the modification in strategy, we still plan to deliver high-fidelity full-system physics characterizations of a nuclear weapon in 2009. At that time, we will also deliver a suite of validated codes, running on supercomputer platforms acquired through open procurement. Accompanying the codes will be user-friendly environments, advanced visualization tools for analysis, and the entire support structure to tie the components together. ASCI will also deliver high-performance storage, sophisticated solvers for linear systems, and high-bandwidth networks. In support of a true trilaboratory effort, ASCI continues to push the envelope in computing across platforms located at great distances from each other and in advanced encryption techniques and other approaches to ensure secure networking.

The process of quantifying margins and uncertainties in nuclear weapons systems will continue to influence ASCI priorities. In turn, ASCI's efforts to produce high-fidelity simulations will increase the predictive science capability and thus reduce uncertainties (see the article "QMU and Nuclear Weapons Certification" on page 47). The elements required to increase our predictive capability are shown in Figure 5. Building on Laboratory basic-research activities and external collaborations, ASCI will ensure that the tools needed to support the simulation of the most complex physics devices ever modeled will be ready when needed. Only by continual investment in fundamental science can we create the realistic models and pre-

dictive capability that will enable code developers and weapons designers to address the problems presented by the aging nuclear stockpile. ■

*For further information, contact James Peery (505) 667-0940 (jspeery@lanl.gov).*



# Massively Parallel Multiphysics Code Development

*Jim E. Morel*

The Advanced Simulation and Computing (ASC) program is developing very large massively parallel multiphysics codes for reliably simulating nuclear weapons performance in the absence of nuclear testing. The task of developing multiphysics codes for the weapons program has always been a daunting one. A huge system of time-dependent, coupled nonlinear equations must be solved. These equations model many different types of physics. It is highly desirable, if not essential, that the solution process for such a system be decomposed into a series of steps, with each step consisting of the solution of equations associated with a single type of physics. Such an approach enables the code to be assembled with largely independent monophysics modules. This property is critical when one considers that essentially no one on a team is an expert in all the

types of physics modeled in the code. Traditionally, only one or two people on a weapons code team had detailed knowledge of the coupling required between all the different types of physics in the code. Most of the teams consisted of individual experts in a single type of physics that contributed to a monophysics component. Today, only the size of the teams is different. An ASC code team generally consists of subteams, rather than individuals, who are responsible for monophysics modules. The numerical technique that has traditionally been used to decompose the solution process into a sequence of essentially monophysics steps is still used in the current generation of ASC codes. It is called operator splitting.

To demonstrate this concept, we need to review some basic concepts of temporal discretization. Nonlinear systems are generally solved by using a linearization

process coupled with an iteration on the nonlinear terms. More specifically, the nonlinear equations are approximated with linear equations. After each solution of the linear equations, the nonlinear terms are updated. The process is then repeated until the nonlinear solution is converged. To understand operator splitting, we need consider only a set of linear equations. However, we must first review some basic concepts of temporal discretization. Although equations are generally discretized in all variables, we need not explicitly consider the other discretizations to illustrate the necessary points. For instance, let us consider a generic time-dependent linear system:

$$\frac{\partial f}{\partial t} = Af, \quad (1)$$

where  $f$  is the unknown,  $t$  is time, and  $A$  is a linear operator. A fully explicit time discretization is denoted as follows:

$$\frac{f^{n+1} - f(n)}{\Delta t} = Af^n, \quad (2)$$

where  $n$  is the time index,  $\Delta t = t^{n+1} - t^n$  is the time step,  $t^n$  is the initial time associated with a time step, and  $t^{n+1}$  is the final time. Solving the explicit equation is generally inexpensive because one need only apply the operator  $A$  to  $f^n$ :

$$f^{n+1} = f^n + \Delta t Af^n. \quad (3)$$

However, explicit methods are generally unstable unless a sufficiently small time step is used. This restriction is acceptable for certain types of physics (for example, for hydrodynamics calculations with strong shocks), but it may be prohibitively expensive for others (for example, for thermal radiation transport). To obtain an unconditionally stable solution technique, one must generally use a fully implicit temporal discretization:

$$\frac{f^{n+1} - f(n)}{\Delta t} = Af^{n+1}. \quad (4)$$

The solution of the implicit equation is generally much more expensive than the solution of the explicit equation because one must invert an operator and apply it to  $f^n$ :

$$f^{n+1} = (I - \Delta t A)^{-1} f^n. \quad (5)$$

Suppose that we have two coupled equations. For instance, let us consider typical equations for the electron and ion temperatures in a plasma:

$$C_{ve} \frac{\partial T_e}{\partial t} = \vec{\nabla} \cdot K_e \vec{\nabla} T_e + \alpha(T_i - T_e), \quad (6a)$$

and

$$C_{vi} \frac{\partial T_i}{\partial t} = \vec{\nabla} \cdot K_i \vec{\nabla} T_i + \alpha(T_e - T_i), \quad (6b)$$

where  $T_e$  is the electron temperature,  $T_i$  is the ion temperature,  $C_{ve}$  and  $C_{vi}$  are the electron and ion heat capacities,

respectively,  $K_e$  and  $K_i$  are the electron and ion conduction coefficients, respectively, and  $\alpha$  is the coupling coefficient for internal energy exchange between the electron and ion fields. Modern computers can easily solve this system using a fully implicit temporal discretization:

$$\frac{C_{ve}}{\Delta t} (T_e^{n+1} - T_e^n) = \vec{\nabla} \cdot K_e \vec{\nabla} T_e^{n+1} + \alpha(T_i^{n+1} - T_e^{n+1}), \quad (7a)$$

and

$$\frac{C_{vi}}{\Delta t} (T_i^{n+1} - T_i^n) = \vec{\nabla} \cdot K_i \vec{\nabla} T_i^{n+1} + \alpha(T_e^{n+1} - T_i^{n+1}), \quad (7b)$$

However, this solution was not always easy to obtain. Operator splitting was once routinely used to reduce the solution of Equations (6a) and (6b) to a sequence of simpler solutions. In particular, a conduction calculation was first performed for the electrons,

$$\frac{C_{ve}}{\Delta t} (T_e^{n+\frac{1}{2}} - T_e^n) = \vec{\nabla} \cdot K_e \vec{\nabla} T_e^{n+\frac{1}{2}}, \quad (8)$$

followed by a conduction calculation for the ions,

$$\frac{C_{vi}}{\Delta t} (T_i^{n+\frac{1}{2}} - T_i^n) = \vec{\nabla} \cdot K_i \vec{\nabla} T_i^{n+\frac{1}{2}}, \quad (9)$$

followed by a local calculation of the coupling between the unknowns,

$$\begin{aligned} \frac{C_{ve}}{\Delta t} (T_e^{n+1} - T_e^{n+\frac{1}{2}}) \\ = \alpha(T_i^{n+1} - T_e^{n+1}), \end{aligned} \quad (10a)$$

and

$$\begin{aligned} \frac{C_{vi}}{\Delta t} (T_i^{n+1} - T_i^{n+\frac{1}{2}}) \\ = \alpha(T_e^{n+1} - T_i^{n+1}). \end{aligned} \quad (10b)$$

We refer to Equation (10b) as a local

calculation because this particular type of coupling between temperatures leads to discrete equations that are independent in each spatial cell, as opposed to Equations (8) and (9), which involve coupling between adjacent cells. Compared with equations containing spatial coupling, local equations are generally very easy to solve and highly amenable to parallelization. If we add Equations (8) through (10b), we obtain a set of difference equations that are ‘‘semi-implicit’’ in that all the operators are applied to unknowns at advanced times:

$$\begin{aligned} \frac{C_{ve}}{\Delta t} (T_e^{n+1} - T_e^n) = \vec{\nabla} \cdot K_e \vec{\nabla} T_e^{n+\frac{1}{2}} \\ + \alpha(T_i^{n+1} - T_e^{n+1}), \end{aligned} \quad (11a)$$

and

$$\begin{aligned} \frac{C_{vi}}{\Delta t} (T_i^{n+1} - T_i^n) = \vec{\nabla} \cdot K_i \vec{\nabla} T_i^{n+\frac{1}{2}} \\ + \alpha(T_e^{n+1} - T_i^{n+1}). \end{aligned} \quad (11b)$$

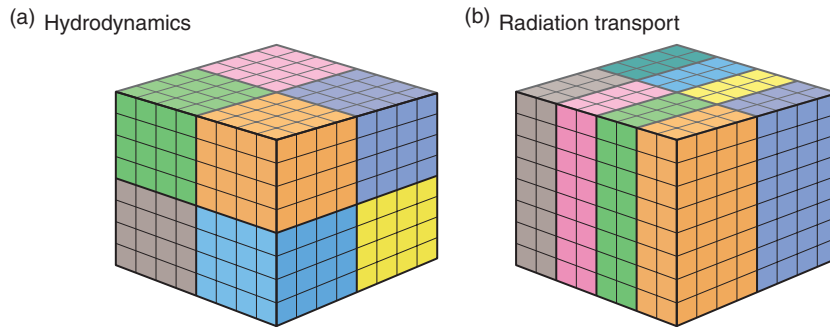
Because each solution step is fully implicit, this entire process is unconditionally stable. In general, the split solution is nearly as accurate as a fully implicit solution as long as the contributions from two or more steps are not nearly equal and opposite. If this is the case, time steps must be taken that can be extremely small relative to those required with a fully implicit discretization. Also, difficulties may be encountered in certain asymptotic limits. In recent years, an alternative to operator splitting has emerged that, in principle, can be used to solve large multiphysics systems of equations in a fully implicit manner. This technique is called the Newton-Krylov method. I will not discuss this method in detail here, but suffice it to say that it is not sufficiently mature to be used in an ASCI code project. However, it is very promising and may become the solution method of choice in the long term. In the short term, ASCI projects will continue to rely on operator splitting, and ASCI



researchers will attempt to better understand the deficiencies of operator splitting and eliminate them.

Given the previous example, it is not difficult to see that operator splitting can also be used to separate coupled multiphysics calculations into a set of multiphysics solution steps together with local coupling steps. For instance, in a radiation-hydrodynamics calculation, one might first perform a hydrodynamics calculation, followed by a radiation transport calculation, followed by a local calculation of the coupling between the hydrodynamics and transport unknowns. This approach enables the development of essentially independent hydrodynamics and radiation transport software modules, together with a relatively simple module for coupling them. However, the implication here is not that the hydrodynamics and radiation transport teams can proceed completely independently of one another and then do the coupling after their respective modules are finished. Considerable planning and coordination are required before the software is written to ensure that the respective numerical treatments are compatible. For instance, if the material temperatures are assumed to be located at cell centers in the hydrodynamics equations, it is much easier to couple the modules and probably more accurate overall if the same assumption is made for the radiation transport.

The planning and coordination that must be achieved to ensure compatibility between physics modules are much more complicated with massively parallel computers than they were with serial and vector computers. The reason is that on multiprocessor distributed-memory computers, different physics modules often require different data partitionings on the processors for optimal performance. Distributed-memory machines store data on each processor. Data partitioning is simply the mapping of data to the processors on which they will be stored. In many instances, each datum



**Figure 1. Spatial Domain Decomposition for Hydrodynamics vs Radiation Transport**

**A 3-D rectangular mesh is divided into eight computational domains (denoted by different colors), and each is assigned to a different processor. The optimal division for a hydrodynamics calculation (a) is quite different from that for radiation transport (b).**

can be uniquely associated with a single spatial grid cell. In such instances, all the data can be partitioned simply by partitioning the spatial grid itself, that is, by mapping each spatial cell (and hence the data associated with that cell) to a processor. Some of the information about a spatial grid partitioning is easily visualized. In particular, we can easily see what cells are mapped to the same processor by first assigning a unique color to each processor and then assigning a processor color to each cell in accordance with the cell-to-processor mapping. This information is generally referred to as the spatial domain decomposition. For instance, a typical domain decomposition for a hydrodynamics calculation on a three-dimensional (3-D) rectangular mesh is shown in Figure 1(a), and a typical domain decomposition for a radiation transport calculation on a 3-D rectangular mesh is shown in Figure 1(b). The partitionings are quite different. Thus, data that are shared by the hydrodynamics and transport calculations must be repartitioned during every time step at some point between the hydrodynamics and transport calculations. This requirement clearly complicates the coupling of physics modules. At one time, it was thought that such repartitioning would be prohibitively expensive. However, experience with ASCI codes indicates that repartitioning is not a problem as long as it occurs only between the execution of

modules that do a significant amount of computational work. This is certainly the case for hydrodynamics and radiation transport modules.

Another area in which massively parallel computing has significantly complicated multiphysics code development is the process of programming itself. On massively parallel computers, one must be concerned with moving data between processors while computing. This requirement adds another layer of complexity to the programming process that was not present with scalar and vector computers. A physicist working on an ASCI code team today requires much more computer science and advanced programming knowledge than a physicist working on a traditional serial or vector weapons code. This feature can be a problem for new hires coming onto code teams because they can require considerable training before being able to contribute effectively. Although there is a formal education program for training new designers in the weapons program, there is no formal education program to train new software developers. Efforts have been made to develop software frameworks that allow individuals to write parallel programs without a high level of computer science knowledge, but such approaches have not yet been effective for the large multiphysics programs written within the ASCI projects.

Finally, the ASCI program has been



### Teaching Radiation Transport

Radiation transport, a subject rarely taught at universities, is very important to the development of the ASCI multiphysics codes for nuclear weapons. To help train young people in this field, the author initiated a graduate-level class in numerical methods for radiation transport. The class, offered for credit by the Chemical and Nuclear Engineering Department of the University of New Mexico, is taught at Los Alamos and is received simultaneously at three remote sites through a new technology, Access Grid Web-based teleconferencing. The Access Grid node at Los Alamos, one of more than 300 nodes worldwide, is run by the Advanced Computing Laboratory as part of its effort in long-distance communication. The photo shows a class in progress. On the wall are projected the classrooms at the three remote sites—University of New Mexico and Sandia National Laboratories at Albuquerque and Livermore. The inset shows node operator Cindy Sievers.



asked to deliver new code capabilities in a time frame much shorter than that associated with traditional weapons code development projects. The assumption was made that this goal would be possible because each project team would consist of several tens of individuals. However, this increase in team size was coupled with our traditional code development processes. These processes, which worked well for small teams, have failed to scale with large teams. No ASCI project team has yet found a way to efficiently utilize all its team members. This is really a management problem rather than a technical problem, but it is as difficult and as important as any technical problem faced in ASCI. Furthermore, this prob-

lem is clearly exacerbated by the fact that ASCI project teams often have no time to investigate new development processes because they are struggling to make milestones. However, they may be struggling to make milestones because they do not have adequate processes. The latest ASCI strategy at the Laboratory calls for the investigation of new code-development software environments and associated code-development processes. We hope to leverage some of the work done in this regard by other high-performance computing programs funded by the Department of Energy, such as the Scientific Discovery through Advanced Computing (SIDAC) Program. ■

*For further information, contact  
Jim Morel (505) 667-6091  
(jim@lanl.gov).*

# High-Resolution Methods for Hydrodynamics

William J. Rider

The decade of the 1980s saw a revolution in computational fluid dynamics that was driven by a new breed of algorithms known as high-resolution methods. The Advanced Simulation and Computing program of the Department of Energy provides the mechanism for making these methods available to weapons designers. When combined with advances in computing hardware, these methods will result in unprecedented computational fidelity within the weapons program.

## Traditional Methods

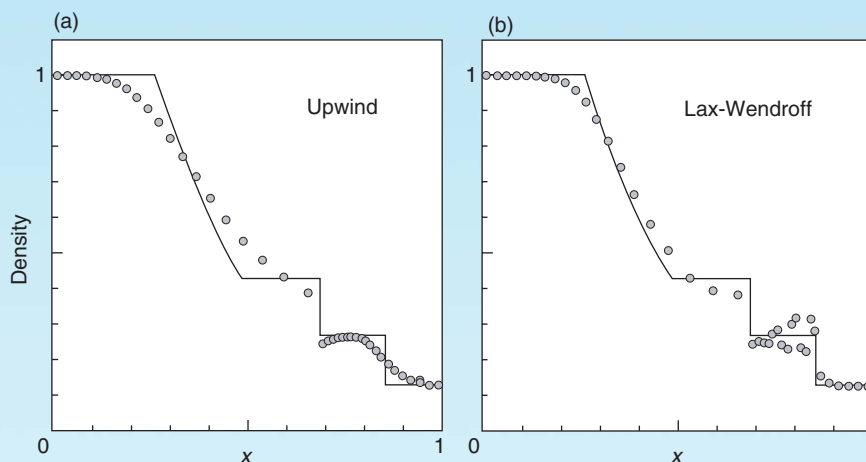
We begin a discussion of traditional numerical hydrodynamics by considering the following prototypical equation:

$$\frac{\partial u}{\partial t} + \frac{\partial au}{\partial x} = 0, \quad (1)$$

where  $u$  is a function of  $t$  (time) and  $x$  (space), and  $a$  is a positive constant. This equation describes the advection, or transport, of the quantity  $u$  along the  $x$ -axis with velocity  $a$ . The flux associated with this advection process is the quantity  $au$ . The quantity  $u$  is con-

served because the integral of  $u$  over the entire  $x$ -axis remains constant in time. As a result, Equation (1) is called a conservation equation. To solve Equation (1) numerically, we first defined a mesh consisting of discrete points in time and space. Let  $n$  denote the temporal index, and  $j$  denote the spatial index. The time step,  $\Delta t$ , is equal

to  $t^{n+1} - t^n$ , and the spatial cell width,  $h$ , is equal to  $x_{j+1} - x_j$  for all  $j$ . Once the mesh has been defined, a discretization scheme is used to evolve the function  $u$  in time. More specifically, given all values of  $u$  on the spatial mesh at time  $n$ , a discretization scheme is used to compute all values of  $u$  on the mesh at time  $n + 1$ . We first consider two tra-



**Figure 1. Simulated and Exact Density Profiles for a Shock Tube Problem**

The shock tube problem begins with a membrane between two quiescent gases at different pressures and densities. When the membrane is broken, a complex wave interaction is initiated. The solution shown here corresponds to the material density as a function of position at a specific time after the membrane has been broken. The exact solution (solid line) results from solving this so-called Riemann problem. In (a) and (b), the analytic solution is compared with the solutions calculated with the simple first-order upwind method and the second-order Lax-Wendroff method, respectively. The plots in (a) and (b) illustrate the basic tradeoff between monotonic-dissipative and oscillatory-dispersive discretization techniques.

ditional discretization schemes for solving Equation (1). The first is the simple upwind scheme,

$$u_j^{n+1} = u_j^n - \lambda(u_j^n - u_{j-1}^n), \quad (2)$$

where  $\lambda = a\Delta t/h$ , and the second is the Lax-Wendroff scheme,

$$u_j^{n+1} = u_j^n - \lambda(u_{j+1}^n - u_{j-1}^n)/2 + \lambda^2(u_{j+1}^n - 2u_j^n + u_{j-1}^n)/2. \quad (3)$$

Assuming that the solution for  $u$  is smooth, we can use a form of Taylor series analysis to determine the associated error with the discretization scheme. For instance, the error associated with the simple upwind scheme is

$$E_{UP} = h \frac{|a|}{2} \frac{\partial^2 u}{\partial x^2} + O(h^2), \quad (4)$$

where  $O(h^2)$  denotes terms proportional to  $h^k$ , where  $k$  is an integer greater than or equal to 2. The upwind scheme is said to be first-order accurate because its error is proportional to  $h$ . The error associated with the Lax-Wendroff scheme is

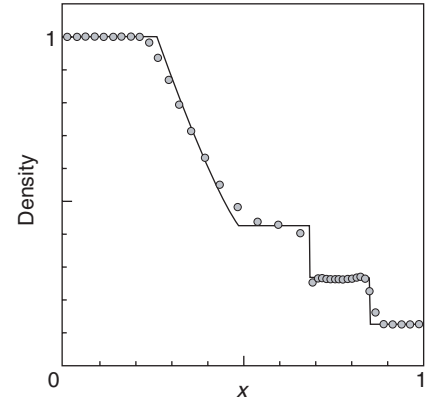
$$E_{LW} = h^2 \frac{a}{6} \frac{\partial^3 u}{\partial x^3} + O(h^3). \quad (5)$$

The Lax-Wendroff scheme is second-order accurate. If the cell width is decreased by a factor of 2, the error associated with a first-order scheme decreases by a factor of 2 ( $2^1$ ), but the error associated with a second-order method decreases by a factor of 4 ( $2^2$ ). Traditional second-order methods tend to be significantly more accurate than first-order methods, but they tend to oscillate badly when solutions are not smooth, that is, when discontinuities are present. Furthermore, they can be dispersive in that a single wave can nonphysically break up into several smaller waves. The problem with oscillations is that they can produce unphys-

ical states in the calculation, such as negative densities or pressures. First-order methods do not oscillate when shocks are present, but they tend to significantly broaden shock fronts and dissipate energy. These properties are illustrated in Figure 1, where a shock tube problem is solved with the first-order upwind and the second-order Lax-Wendroff methods. Because solutions with shocks are not smooth, the Taylor series analysis used to characterize the accuracy of the upwind and Lax-Wendroff schemes is not valid. In fact, all discretization schemes are first-order accurate for problems with shocks. Nonetheless, as is clear from Figure 1, the errors exhibited by different schemes for such problems can be far different in magnitude and character.

### High-Resolution Methods

An essential element of high-resolution discretization schemes is nonlinearity. This property follows in part from a very important theorem, originally developed by Sergei Godunov, which states that a second-order linear discretization cannot produce monotone (nonoscillatory) solutions to Equation (1). Godunov's theorem motivated researchers to investigate the addition of nonlinearities to discretization schemes, and this study resulted in a major breakthrough. High-resolution discretization schemes are generally constructed from three linear schemes: one is first order and monotone, and two are second order. These schemes are then combined in a manner that ensures second-order accuracy when the solution is smooth and both high



**Figure 2. High-Resolution Solution to the Shock Tube Problem**  
The high-resolution method results in a solution to the shock tube problem that very closely matches the analytic solution (solid line).

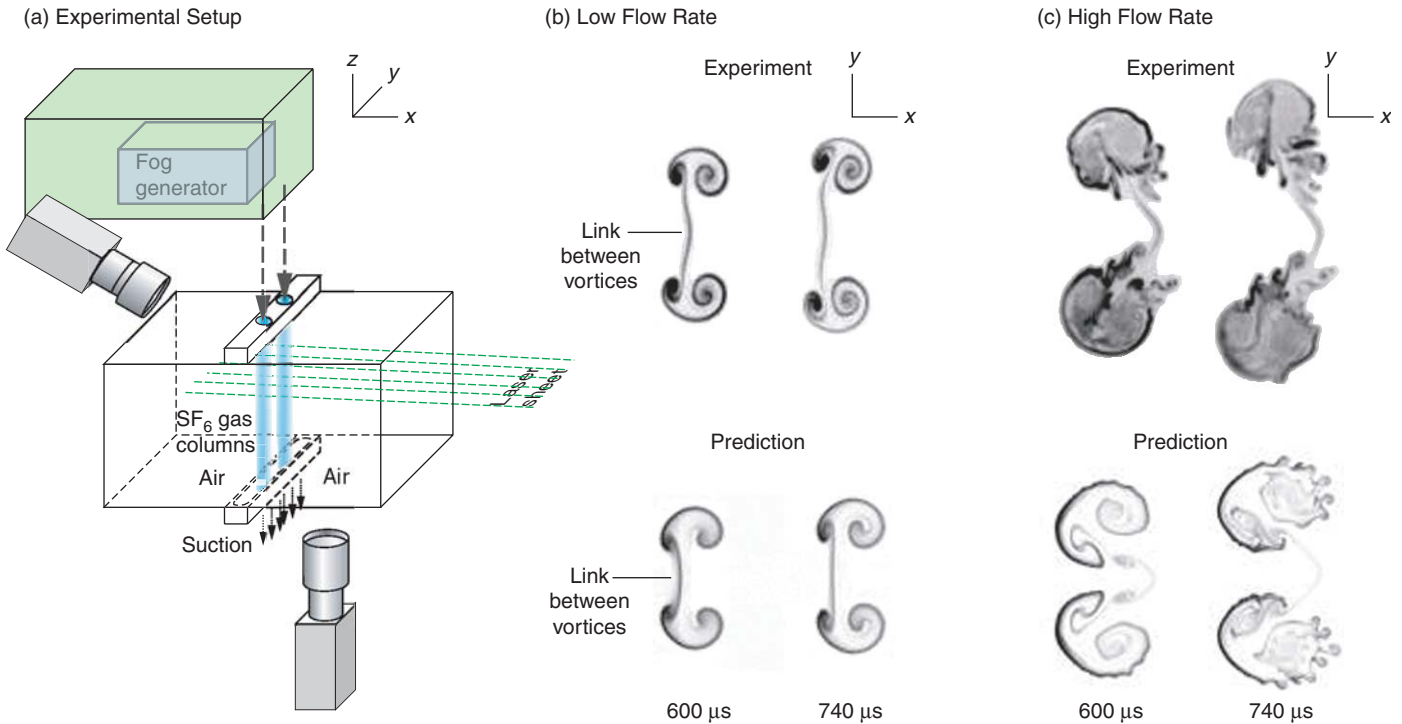
accuracy and monotonicity when the solution is not smooth. At the heart of this approach is a nonlinear limiter that effectively switches between definitions for certain terms, depending upon the local behavior of the solution. For instance, an example of a simple high-resolution method is provided in the box below—compare with Equations (2) and (3):

$$u_j^{n+1} = u_j^n - \lambda(u_j^n - u_{j-1}^n) - \lambda(1 - \lambda)(\phi_j^n - \phi_{j-1}^n)/2, \quad (6)$$

where the limiter,  $\phi$ , is defined as follows:

$$\phi_j^n = (u_j^n - u_{j-1}^n) \max \left[ 0, \min \left( 1, \left( (u_{j+1}^n - u_j^n) / (u_j^n - u_{j-1}^n) \right) \right) \right]. \quad (7)$$

This method was used to obtain the numerical solution plotted in Figure 2. The high-resolution solution shows a dramatic increase in accuracy over the two solutions plotted in Figure 1. A very important property of high-resolution methods relates to their unique ability to model turbulent behavior. Traditional methods are



**Figure 3. Simulating Richtmyer-Meshkov Experiments at Low and High Flow Rates**

Panel (a) shows an experiment to study the Richtmyer-Meshkov instability in which two columns of SF<sub>6</sub> gas (with density five times that of air) flow downward through the test section under the force of gravity and are hit by a planar shock wave with Mach number 1.2. The shock deposits vorticity along the cylinder edges, which distorts them into a “mushroom cap” shape. Images of the unstable structures are captured by laser-sheet visualization at two times after the passage of the shock. Panels (b) and (c) compare experimental and simulated results for experiments at low flow rate (and Reynolds number) and high flow rate (and Reynolds number), respectively. The simulations used the xPPM method implemented in the computer code Cuervo. Both simulations successfully model the gross features of the flow, including links between the two gas columns. (This work was conducted jointly with Christopher Tomkins, Robert Benjamin, and James Kamm of Los Alamos).

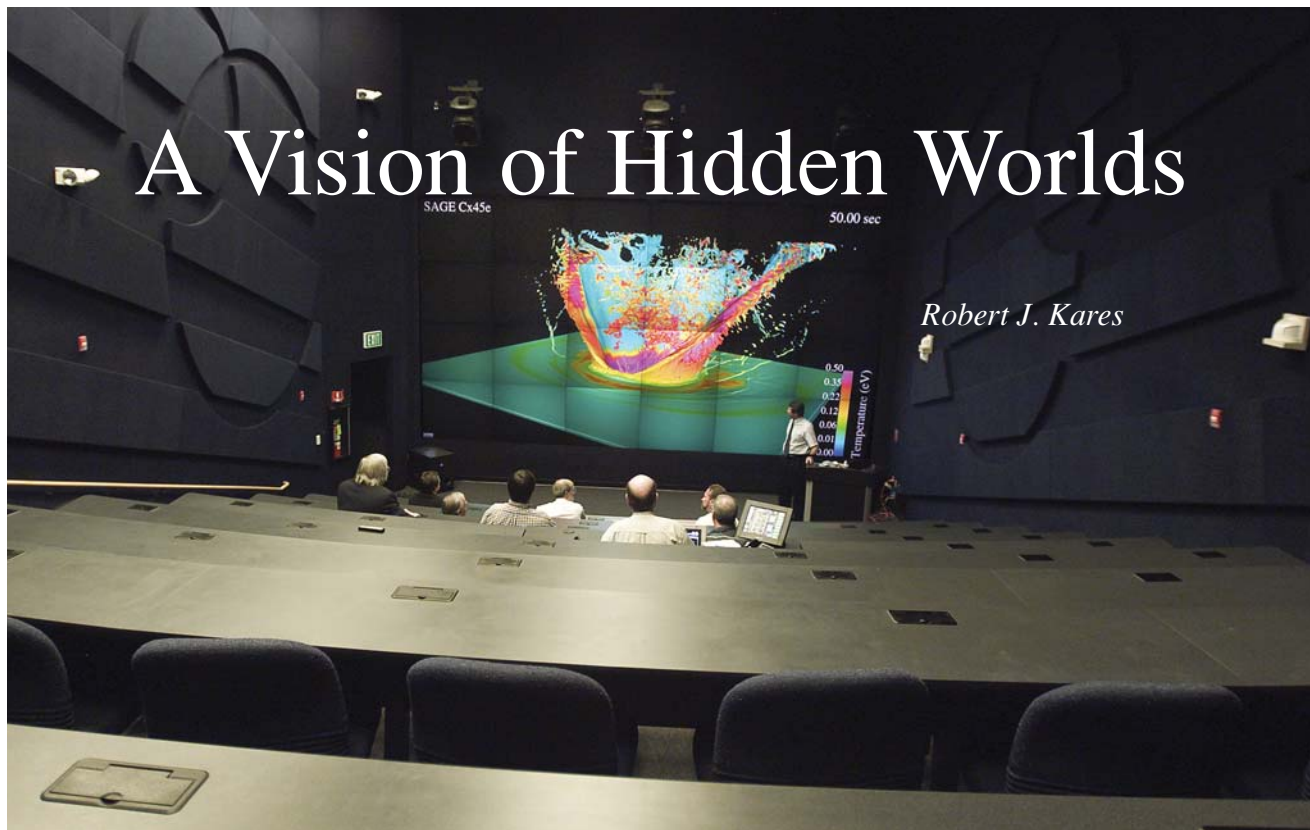
essentially unable to model such behavior. The latest high-resolution, known as second-generation, methods appear to be particularly useful for computing turbulent flows. The ultimate litmus test for computational methods and modeling is a direct comparison with experiment. A shock wave can induce turbulent mixing between two materials. This phenomenon is known as the Richtmyer-Meshkov instability, and it is important in situations arising in astrophysics, high explosives, and inertially confined fusion experiments. For the past 5 years, we have been working very closely with experimentalists to validate our methods for modeling the development of instabilities. Using one of our new second-generation methods, xPPM,

we computed the mixing of two cylinders of sulfur hexafluoride gas with a background gas (air). The simulation showed a bridge of material linking the two cylinders late in the experimental time. Although the link was heretofore unobserved, when the experimentalists succeeded in improving their ability to monitor the evolution of the instability, they discovered that the link was indeed present at both low and high flow rates (see Figure 3). The graphic on the opening page is a later result comparing first- and second-generation high-resolution results for a single gas column (calculated with Jeff Greenough of Lawrence Livermore National Laboratory). In the future, we hope to make more quantitative comparisons between experiments

and turbulent flows calculated with our second-generation methods. Finding ways to achieve this goal is a research topic in itself. ■

*For further information, contact Bill Rider (505) 665-4162 (rider@lanl.gov).*

# A Vision of Hidden Worlds



Robert J. Kares

*The latest in theaters*

No human eye will ever see the inside of a nuclear weapon as it explodes. Many important processes in the physical world will never be seen directly by any human eye because they lie too far outside the narrow confines of human existence. However, using the computational power of the Advanced Simulation and Computing (ASCI) program, we have begun to create realistic three-dimensional (3-D) simulation worlds that can be used to study such processes, to bring them into the realm of human experience (see Figure 1). Yet understanding a dynamic, 3-D simulation world whose complexity increasingly approaches that of the real physical world represents a new kind of challenge.

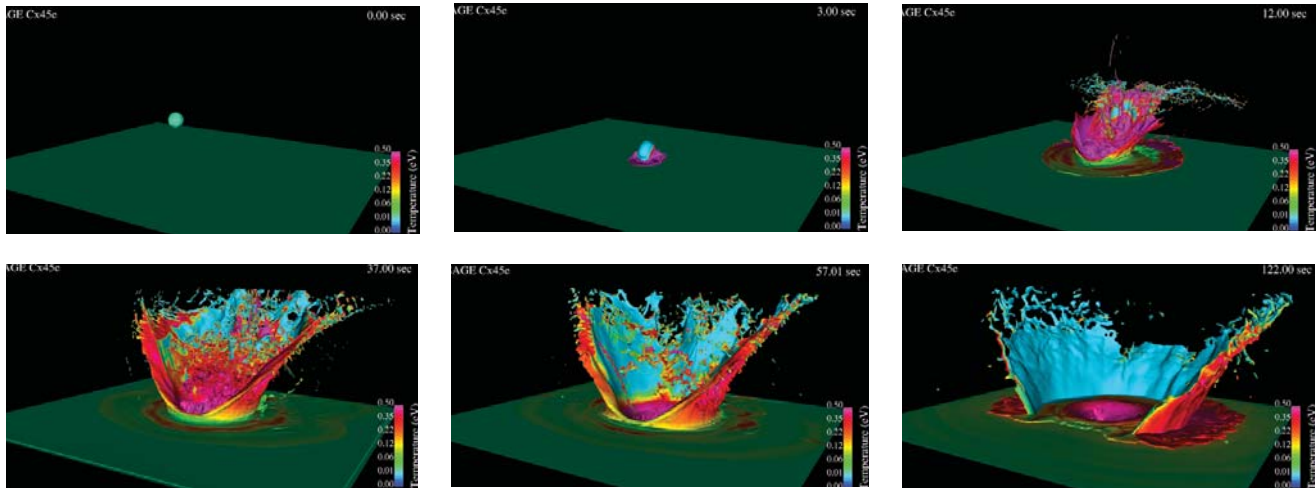
Fortunately, the same tools of perception that human beings use to understand the real world can be

applied to simulation worlds. The most important of these tools is human stereo vision. In 1998, the Laboratory began a project called the Visual Interactive Environment for Weapons Simulation (VIEWS), a part of the larger ASCI program, to build an infrastructure that exploits the power of human stereo vision to understand and study the simulation worlds we have created. With this infrastructure, we can now see with our own eyes worlds that could only be imagined by the pioneers who founded the Laboratory 60 years ago.

The visualization infrastructure created by VIEWS is called a Data Visualization Corridor (DVC). A DVC is a very high-performance, end-to-end system of hardware and software that transforms raw simulation data into realistic images and quantitative results that a human user can understand. Large, dedicated visualization server machines read

the many terabytes of raw simulation data and produce high-resolution stereo images of the results. These images are then transmitted over a high-bandwidth fiber-optic distribution system to the many viewing facilities of the corridor, which include user desktops, stereo theaters, stereo power-wall displays (see Figure 2), and immersive virtual-reality environments. VIEWS deployed its first DVC in 1998. Today, our latest generation of the DVC provides services for all the many users of the 30-teraflop Q machine in the Laboratory's new Nicholas C. Metropolis Center for Modeling and Simulation (the Metropolis Center).

The principal feature that distinguishes the Los Alamos DVC from a simple, high-end graphics workstation is its ability to manipulate and image truly enormous volumes of data, even when the information does not physi-



**Figure 1. Hidden Worlds**

Advanced visualization tools enable us to view the results of extremely large scale simulations with unprecedented realism. The images are from a 3-D simulation of the “dinosaur killer,” an asteroid or comet impact event believed to have led to a world-wide sequence of mass extinctions 65 million years ago. Dissipation of the asteroid’s 300 teratons TNT equivalent of kinetic energy produced a stupendous explosion that melted and vaporized a substantial volume of calcite, granite, and water and ejected it into the atmosphere in the form of an “ejecta blanket.” The measurable distribution of material in the blanket can be used as a diagnostic to determine the direction and angle of impact of the asteroid.

cally reside in Los Alamos. Recently, Laboratory scientists completed a single 3-D simulation on the 12-teraflop ASCI White machine at Lawrence Livermore National Laboratory in California that generated more than 21 terabytes of visualization data. (The entire Library of Congress contains only 17 terabytes of data.) The Los Alamos DVC uses a cluster of three of the largest graphics supercomputers in the world, located in Los Alamos and linked to the Livermore White machine via a high-speed, wide-area network, the DISCOM WAN, to manipulate and image this data interactively.

The video fiber distribution system that carries the images created by these graphics supercomputers to corridor displays contains a 128-way crossbar switch that allows corridor users to switch the graphics output of any of these visualization server machines to any of the viewing facilities on the corridor. The bandwidth of this fiber distribution system is sufficient to display up to 30 high-resolution, stereo color images per

second, each image using up to 16 million colors. The corridor can deliver high-resolution color movies of terascale simulation results to users at the same play rate of a normal motion picture. The difference is that corridor movies are fully 3-D stereo movies.

The Los Alamos DVC provides a variety of advanced viewing capabilities for visualization users. At the moment, more than 50 user offices are connected to the corridor. Each office is equipped with a 1920 X, 1200-pixel-resolution, 24-inch, stereo-capable video display, a



**Figure 2. A New Way to View Reality**

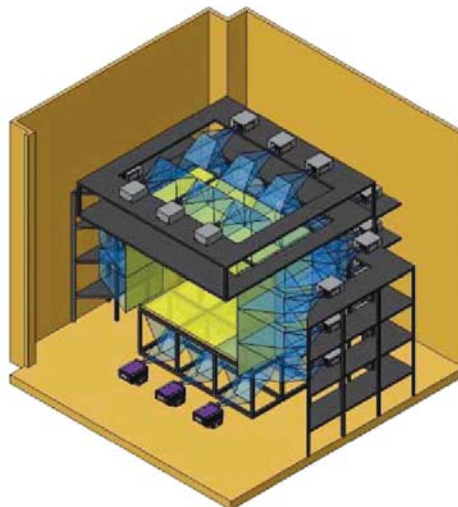
This illustration shows one of two 8-million-pixel stereo “collaboratory” power walls in the new Metropolis Center. The power wall, which is 11 ft wide by 6 ft high, combines multiple display panels and multiple projectors to create a single large display with very high resolution.

mouse, and a keyboard so that users can interactively view high-resolution stereo images in their offices. The corridor also provides two stereo viewing facilities for the collaboration of small work groups. Each of these “collaboratories” contains an 8-million-pixel stereo power-wall display, such as the one seen in Figure 2.

For larger groups, we have just completed construction of a 31-million-pixel stereo power-wall theater that will provide stadium seating and full 3-D viewing for an audience of 85 people. The power wall for this new theater is 22 feet wide by 12 feet high and uses 24 state-of-the-art Christie Digital projectors arranged in a 6-by-4 matrix to create a single huge stereo display. This theater, which was completed in March 2003, is shown in the photo introducing this article. The theater will allow a large audience full interactive 3-D viewing of terascale ASCI simulations without compromise in image scale or resolution.

The Los Alamos DVC also provides facilities for fully immersive, virtual-reality exploration of large datasets. The first such facility, the Los Alamos Reconfigurable Advanced Visualization Environment (RAVE), was deployed by the Laboratory in 2000. RAVE is a first-generation immersive environment that uses multiple surrounding panels with stereo images to create the illusion that the user is actually inside the virtual world of the simulation. Moving about in real space, one’s position is tracked, and one’s view is dynamically updated to give the illusion of looking around the hidden corners of real objects. RAVE is a relatively low resolution virtual environment using only about 6.5 million pixels for its images. In 2003, we began construction of our next-generation virtual-reality environment pictured in Figure 3.

This environment will provide a



**Figure 3. A Glimpse of the Future**

This is a sketch of the 43-million-pixel Digital Cave for the Metropolis Center. The Digital Cave is scheduled for completion late in the summer of 2003.

full 43 million pixels of stereo-image resolution. It will also provide both floor and ceiling displays, as well as side displays, so that the user will see the virtual world of the simulation in every direction, even when looking directly up or down. The Los Alamos Digital CAVE will be the most advanced visualization environment in the world and will define the state of the art in virtual reality for years to come.

IEWS has been very successful in providing the “see and understand” infrastructure so essential to the ASCI mission. Working closely with the rest of ASCI, VIEWS provides a vision of worlds that might otherwise remain forever hidden from human eyes. ■

### Further Reading

- Foley, J. D., and A. Van Dam. 1984. *Fundamentals of Interactive Computer Graphics*. Reading, MA: Addison-Wesley Publishing Company.
- Pinker, S., 1997. *How the Mind Works*. New York: W. W. Norton and Company.
- Smith, P. H., and J. van Rosendale. 1998. “Data and Visualization Corridors, Report on the 1998 DVC Workshop Series.” Center for Advanced Computing Research Technical Report CACR-164, accessible at <http://www.cacr.caltech.edu/Publications/DVC>, California Institute of Technology, Pasadena, CA.

*For further information, contact Robert Kares (505) 667-7789 (rjk@lanl.gov).*



# The LANSCE National User Facility

*Thomas Wangler and Paul W. Lisowski*

Particle accelerators have played an important role at Los Alamos ever since the Manhattan Project, when beams from two Van de Graaff accelerators and a cyclotron were used to answer experimentally important nuclear-physics questions. The data obtained from those machines helped ensure success of the “gadget” and its descendents.

Today, an 800-million-electron-volt (MeV) proton linear accelerator, or linac, forms the backbone of the Laboratory’s national user facility, the Los Alamos Neutron Science Center (LANSCE), shown in Figure 1. Protons from the linac, travelling at 84 percent the speed of light, smash into a heavy-metal target and, through a process known as spallation, produce copious numbers of neutrons. The neutrons are used in experiments that support the weapons program and advance basic research. The linac’s proton beam can simultaneously be used in a capability known as proton radiography, or pRad, to make high-speed “movies” of dynamic systems. (See the article “The Development of Flash Radiography” on page 76.)

## **A Brief History of the LANSCE Complex**

The proton linac used by LANSCE was originally built in the late 1960s to produce pi-mesons, also known as pions, for the Los Alamos Meson Physics Facility (LAMPF). The pions and their decay products (muons, electron, and muon neutrinos) were used to conduct fundamental studies in nuclear, atomic, and particle physics.

The idea for a pion factory was

first mentioned in a memo by Louis Rosen, dated May 16, 1962. As Rosen described the situation, this was a critical time in the history of what was then known as Los Alamos Scientific Laboratory. The Cold War was at its height, and the Laboratory had successfully developed our nuclear deterrent—the fission and thermonuclear weapons now in the stockpile. New challenges and new facilities in the field of nuclear science would be needed to maintain and advance the Laboratory’s world-class capabilities. The meson factory would benefit the Laboratory and the nation by providing the experimental capabilities that would support the Laboratory’s programmatic needs for the weapons program. In addition, the facility would foster opportunities in basic research, and be the catalyst for a scientific user program that would bring leading scientists and students to Los Alamos from all over the world.

However, it was a tremendous technical challenge to produce protons with enough energy to support pion production. The protons had to have more than 290 MeV of kinetic energy in the laboratory just to create the pion rest mass, but at that time, proton linac energies had not exceeded 100 MeV. After two years of design studies, a bold proposal was made in September 1964 for an 800-MeV high-current (1-milliampere average current) proton linac for medium-energy physics research. The innovative development that made this energy scale feasible was the side-coupled linac accelerating structure (see Figure 2).

The Medium-Energy Physics Division was formed in July 1965

(with Rosen as division leader) to continue with the design and development of the meson facility. Construction funds were authorized by Congress in the fiscal year 1968 budget, and physical construction began with a groundbreaking ceremony on February 15, 1968. Only four years later, on June 8, 1972, this new linac delivered its first beam. The machine builders can take great pride that LAMPF was completed on schedule and within budget (\$57 million). LAMPF also achieved all its major goals. In September 1972, the facility was dedicated to U.S. Senator Clinton P. Anderson of New Mexico, who was instrumental in turning the project into reality.

An interesting footnote to the LAMPF story is that the side-coupled linac was later widely adopted by industry. As a result, thousands of side-coupled linacs were eventually produced around the world for radiation therapy. Thus, Los Alamos research and development resulted in one of the most beneficial uses of nuclear-physics technology for mankind.

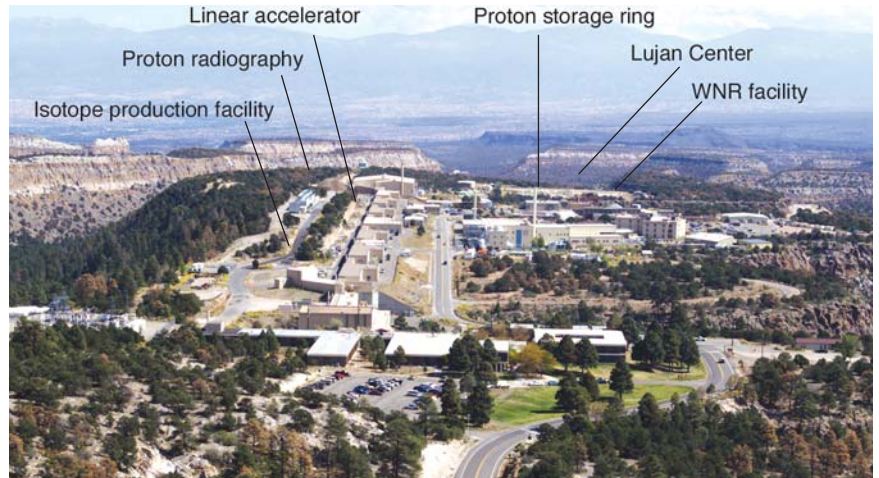
It was clear even before LAMPF was built that, with a neutron-rich, heavy-element target, rather than the light-element target that was suitable for pion production, neutrons would be produced by spallation and that such a source would compete very favorably with nuclear reactor-based neutron sources. Based on that premise, the Weapons Neutron Research (WNR) facility, consisting of a proton-beam transport system and a spallation-neutron target, was built in the LAMPF complex to support the Laboratory’s materials-science and nuclear-physics programmatic needs. The WNR facility produced neutrons

for the first time in May 1977 and gave the Laboratory an intense neutron source that could be used to study the behavior of matter at extreme temperatures and densities, as well as to study radiation effects and obtain nuclear data needed for weapons design.

The LAMPF complex expanded again in the early 1980s with the addition of a unique 30-meter-diameter Proton Storage Ring (PSR), which accumulates the proton beam injected from the linac and compresses the relatively long 625-microsecond pulses into short 125-nanosecond pulses. These short bursts of protons are then directed to the heavy-element target to produce very short bursts of spallation neutrons. The short pulses allow time-of-flight measurements for a more precise determination of the energy and wavelength of the neutrons. Construction of the PSR began in May 1982 and was completed in April 1985. The first beam was circulated in the ring on April 26, 1985.

In 1986, construction was started on a new experimental area (adjacent to the existing WNR facility) that was summarily filled with a suite of world-class instruments. The new area was named the Manuel Lujan Jr. Neutron Scattering Center (the Lujan Center) in honor of a popular congressman from New Mexico. Scientists use the low-energy neutrons at the Lujan Center to perform novel studies of materials. The old WNR facility was also rebuilt during this construction project to provide another spallation source that concentrated on producing beams of higher-energy neutrons for nuclear science. Both new sources came into operation in the early 1990s, just before the medium-energy program of pion and nuclear physics at LAMPF came to an end.

With the closeout of the nuclear physics user program and an increased national need for neutrons, the mission of the LAMPF accelerator



**Figure 1. The LANSCE Complex**

The half-mile-long 800-MeV proton linac is the backbone of the complex.

complex was changed from nuclear physics to neutron research. In October 1995, the complex was renamed as LANSCE.

### The LANSCE Complex Today

The combination of the Lujan Center and the WNR facility gives researchers access to high-intensity neutron beams covering 16 orders of magnitude in energy, and the LANSCE complex has evolved into a major international resource; a record 750 user visits by scientists from around the world occurred during the last operating period. Because so many students and professors visit each year, LANSCE is one of the Laboratory's most important "windows" into the academic community and a source of many of our brightest early-career scientists. It is estimated that LANSCE and its predecessor, LAMPF, have served as a gateway to 10 percent of the Laboratory staff.

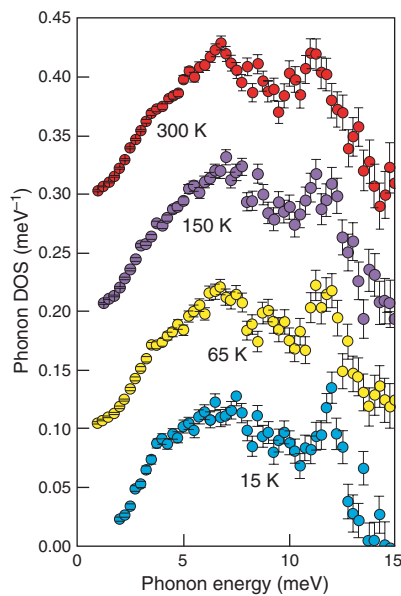
Within the past year, users have achieved many significant accomplishments in materials science, nuclear science, and technology. A few highlights are described below.



**Figure 2. From 100 to 800 MeV**  
The side-coupled linac accelerating structure (100 to 800 MeV) was invented by Los Alamos accelerator scientists during the design of LAMPF.

### Research at the WNR Facility.

The WNR facility is the only remaining broad-spectrum neutron source available to conduct the kinds of precise nuclear science measurements needed to develop predictive capability in the weapons program. As an example of this need, a problem that has been around since the early days of nuclear weapons was recently solved through accurate measurements of the plutonium ( $n,2n$ ) cross section, which is needed to quantify the production of plutonium-238 from plutonium-239. Knowledge of the plutonium-238 production rate over a range of incident neutron energies is essential for understanding past Nevada Test Site (NTS) data. This



**Figure 3. EOS Measurements**  
Shown here is the phonon density of states of a  $\delta$ -phase plutonium-aluminum alloy at different temperatures. The peaks in the data give the distribution of phonon-mode frequencies in the alloy.

definitive measurement has allowed us to significantly improve our predictive capabilities and resolve complicated physics issues in the performance of our weapons that were previously not understood.

Another use of WNR neutrons arises because electronic assemblies, particularly those used in high-altitude aircraft, are subject to neutron-induced “upsets” at the altitudes at which they operate. The neutron-beam energy spectrum from WNR mimics neutrons seen by aircraft electronics in flight, but with an intensity one million times stronger. WNR is now the standard facility for testing neutron-induced upsets in electronics. Thirteen major companies used the facility in 2002 to validate the operation of selected electronics. Similarly, Los Alamos researchers must address whether neutron-induced upsets could affect the ASCI Q machine, one of the most powerful supercomputers in the world and a pillar of the Stockpile Stewardship Program. The intense

neutron source at LANSCE was used to conduct a study of the impact of cosmic-ray-produced neutrons on the Q machine’s reliability.

#### Research at the Lujan Center.

Neutron-scattering experiments at LANSCE have provided a new method for characterizing the basic material properties of plutonium, and we now have the ability to compare plutonium parts that were produced by different manufacturing processes. The original parts were made at the Rocky Flats facility, which is now closed, and newly manufactured parts are made by a different process. The new characterization method allows us to address the question of whether the change will affect weapons performance or safety, two critical issues for the enduring nuclear-weapons stockpile.

Stockpile stewardship also requires that weapons modelers have an accurate, well-understood equation of state (EOS) of plutonium. Otherwise, it is impossible to construct a dependable model of weapons performance. Inelastic neutron scattering from a plutonium-gallium alloy allowed the first-ever determination of phonon density of states, which is an integral component in understanding the EOS (see Figure 3).

The high-pressure preferred orientation (referred to as HIPPO) diffractometer at the Lujan Center can make in situ neutron-diffraction measurements of samples that are at high temperature. Recent experiments revealed texture changes in quartzite that establish this rock as a shape-memory system. In such a system, the orientation of grains is controlled by internal stresses. Shape memory is a desirable attribute for many applications. For example, if an eyeglass frame made from a shape-memory alloy were to become bent, only modest heating would be required to return the frame to its original condition. The revelation that Earth materials have a similar attribute could

have profound implications about the plasticity of Earth’s crust.

Another research avenue involved helium, which is the decay product of the tritium used in nuclear weapons. The partial pressure of helium in a weapon’s neutron generator is a life-time-limiting factor for components of the stockpile. Working with Sandia National Laboratories, scientists at the Lujan Center investigated helium retention, providing information that has significantly influenced the design.

The Lujan Center has completed construction of a major new instrument called Asterix, which provides a polarized neutron beam for studies of magnetic materials and spin polarization. Asterix made it possible to analyze the structure of crystalline and polycrystalline materials through the use of neutron scattering while the materials are located in a magnetic field that is 100,000 times stronger than Earth’s magnetic field.

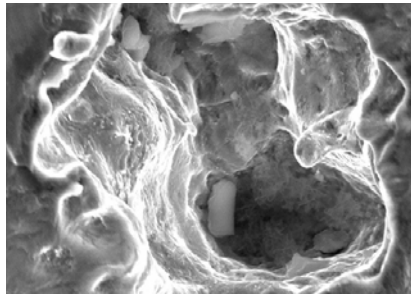
The Lujan Center is also home to the protein crystallography station, the only neutron-scattering instrument in the world designed specifically to investigate the properties of proteins. It performed the first studies of electric-field-induced structural changes in an organic single crystal (see Figure 4).

Finally, LANSCE has completed construction of a major new detector system that will enable, for the first time, measurements of the nuclear properties of radioactive targets using as little as 1 microgram of material. These measurements will affect such diverse areas as the analysis of radiochemical information from past underground nuclear tests, models of astrophysical processes, and technical issues in homeland defense.

**pRad.** Protons that are used to produce neutrons for the Lujan Center and the WNR facility can instead be used directly. Similar to the way an ordinary photon source is used to take



**Figure 4. Protein Crystallography**  
Paul Langan of the Bioscience Division points out the single-crystal  $\alpha$ -glycine sample used in the protein crystallography station.



**Figure 5. Target Investigation**  
A scanning-electron-microscope image of a large pit produced in a steel container. The pit resulted when a bubble of mercury collapsed. The bubble had been generated by an intense proton pulse during a test of SNS target technology at LANSCE.

a picture, in pRad the properties of the protons are exploited to create a radiographic “motion picture” with a frame speed of 50 billionths of a second. The pRad capability has allowed us to understand better the aging and performance of weapons systems components and to observe the properties of materials shocked by high explosives. During the 2002 run cycle, 42 dynamic pRad experiments were performed at LANSCE in support of weapons-physics research efforts at Sandia and, Lawrence Livermore National Laboratories and at the Aldermaston Weapons Establishment. A unique permanent-magnet proton “microscope” system, designed by LANSCE and fielded by

the Physics Division, has produced radiographs with up to 15-micrometer resolution, opening an entirely new realm for studying material features under extremes of pressure and speed.

**New Facilities.** There are two new areas under construction to further enhance the versatility of LANSCE. In 2003, a medical radioisotope facility, called the Isotope Production Facility (IPF), will begin operations. The IPF will allow LANSCE to provide the research community with medical radioisotopes, many of which are not otherwise available in the United States. In a separate area that reuses one of the old experimental areas from LAMPF, researchers have demonstrated the ability to produce the most intense source of ultracold neutrons (UCNs) in the world. UCNs move at speeds not much higher than those at which humans can run, and they have unique properties that allow them to be “bottled” and used for research. Work funded by the Department of Energy is under way to complete a facility for precision studies of forefront problems in physics and cosmology using UCNs.

## The Future

The LANSCE spallation neutron source of 2003 owes its existence to the accelerator work done more than 30 years ago to design and build the original LAMPF complex. Today, major progress in accelerator technology continues. Los Alamos is doing work important for the Spallation Neutron Source (SNS) under construction at Oak Ridge National Laboratory in Tennessee. The \$1.4 billion project—expected to deliver first beam in 2006—uses a particle accelerator that operates at an average power 10 times that of the Lujan Center. Mercury containers were irradiated at LANSCE to help scientists develop techniques to increase the lifetime of the SNS target (see Figure 5).

Although the future of neutron science is shifting to the SNS, LANSCE will still play an important role. With its low-repetition rate, high peak-intensity pulse from the PSR, and innovative neutron-production source, the Lujan Center will be a powerful force in neutron scattering and fundamental physics research for many years. The WNR facility will remain the only neutron source in the world that can perform basic and applied neutron science over the range needed for defense and advanced nuclear applications. In addition, the pRad facility will continue to be invaluable for defense and, possibly, for other research. ■

*For further information, contact Paul Lisowski (505) 667-5051 (lisowski@lanl.gov).*



*The Laboratory Today*

# Threat Reduction and Homeland Security

*Developing strategies and tools for protecting our nation and the world*

Los Alamos NATIONAL LABORATORY



# Six Decades of Reducing Threats and Allaying Fears

*Houston T. Hawkins*

Los Alamos National Laboratory was established amid fears expressed by many knowledgeable people that our way of life might not survive. In its 60-year history, the Laboratory has been preserved because it has successfully welded science and technology to develop weapons intended to neutralize any such fears. Following the tradition of inquiry that is so much a part of our university heritage,<sup>1</sup> we have questioned to exhaustion the principles governing those weapons, which came to embody one of the greatest paradoxes in the history of humankind: Being intrinsically both destructive and safe, they are weapons of war designed to maintain peace. Because proper maintenance of those defining features of our nation's weapons must now be ensured without actual testing, the university environment is becoming increasingly important.<sup>2</sup> But at Los Alamos, we do more than develop weapons. We also use science and technology to hammer out

plowshares that solve other intractable problems and contribute to making the world a better place.

## The Beginnings and Fear Itself

Following the infamous attack on Pearl Harbor, President Roosevelt assured the American people that they had nothing to fear but fear itself. Had Mr. Roosevelt fully understood advances that were occurring in nuclear physics, he might have added, "Of course, we should fear the real possibility that the Third Reich<sup>3</sup> might develop atomic weapons before we do." The scientists who came to this mesa top 60 years ago, many of whom

---

<sup>1</sup> "A university... is a place where inquiry is pushed forward, and discoveries verified and perfected, and rashness rendered innocuous, and error exposed, by the collision of mind with mind, and knowledge with knowledge." (John Henry Cardinal Newman, *What is a University?* From a series of lectures delivered by Cardinal Newman between 1852 and 1854.)

<sup>2</sup> University of California employees have designed all the nuclear weapons tested and stockpiled by the United States since 1943. Los Alamos scientists have designed 75 percent of our enduring nuclear weapons stockpile, and the Lab therefore retains support responsibility for those nuclear weapons.

---

<sup>3</sup> The Imperial Japanese Government also had a nuclear weapons program that, at least with respect to understanding uranium systems, was likely more advanced than the German program led by Werner Heisenberg. In fact, many of the best nuclear cross sections used at Los Alamos during the war had been derived by Japanese scientists (Kikuchi et al. 1939). The Japanese program that was based on the electromagnetic separation of uranium suffered from a lack of resources, and only a few separation machines were available. Professor Paul Karoda, who worked on the Japanese program, reported that the program suffered financially because of a *faux pas* when it was briefed to the Japanese General Staff. Because the output was expressed in "gondola cars of anthracite" instead of tons of explosive, Minister of War Tojo Hideki was not impressed and wondered aloud why someone would drop that much coal on a city. (Private communication with Professor Paul Karoda, University of Arkansas, 1979.)

had escaped from the brutality of the Third Reich, were driven by that fear. We know that the researchers who worked here believed that one morning they would read in a newspaper that an atomic weapon carried by a V-2 rocket had destroyed London and dealt a mortal blow to the hope and future of human civilization. That rational fear motivated Los Alamos researchers to accomplish the impossible. In two short years, they solved all the intervening scientific and engineering problems, developed and built atomic weapons of two different designs, and proof-tested the more complicated atomic device near Alamogordo, in New Mexico.

This success came after the Third Reich had been defeated with conventional means. However, in the Pacific, we still faced a pernicious adversary whose atrocities in Asia matched those committed by the Third Reich in Europe. In the Pacific theater, the battles had grown increasingly violent and bloody as they were coming closer to the Japanese mainland. Casualties expected in a direct invasion of Japan were estimated at millions. Faced with this possibility, President Truman ordered that the atomic bomb be dropped over Japan. Days later, an atomic device developed at Los Alamos exploded over Hiroshima and another one, over Nagasaki. Although historians might debate Truman's decision, the indisputable fact remains that his action did bring about an immediate end to the war in the Pacific and that both sides avoided the inevitable, enormous carnage that would have been caused by an invasion. (See the box on the next page for a personal story related to this period in world history.)

### **The Cold War— “Duck and Cover”**

Following the end of World War II, the free world faced a new fear. Since

its expansion began in 1600, the Muscovy Principality had taken over territory the size of the Netherlands every year for 150 consecutive years, thus making Russia larger than the rest of Europe combined by 1750. Siberia, for example, was completely conquered by the end of the seventeenth century. That conquest alone doubled the size of Russia at the time. Up until the 1950s, acquisition of territory and people continued, regardless of the label on the authoritarian government in charge in Moscow. Richard Pipes (1974) has contended that this oppressive and authoritarian system, Tsarist or Bolshevik, could not create wealth but only acquire it through conquest. After 1945, the expansion led by Soviet communism was moving into Eastern Europe at an alarming pace. Occupied countries were stripped of their treasure and industry, and the plunder was shipped to Russia. This spread of Soviet communism, which had already murdered and enslaved millions of Russia's own citizens, became increasingly frightening because the regime was already engaged in the development of nuclear and thermonuclear weapons.

As a result, we accelerated our own thermonuclear weapons development, leading to Operation Greenhouse, a series of tests conducted by Los Alamos in the Pacific in 1951. Ultimately, the strategic race in nuclear arms led to the deployment of thousands of strategic nuclear weapons by both the Soviet Union and the United States. Peace was achieved by an unlikely concept—mutual assured destruction. However, even with stability at the strategic level, many military planners thought as early as 1945 that, given Soviet capabilities, intentions, and geography, stanching Soviet advances would be a very difficult, if not impossible, task. Space for all the tanks, airfields, materiel, and troops required by such a task was simply not available. A

few planners, such as Secretary of War James Forrestal, were reportedly driven to psychological desperation, fearing that the Soviets would take over Western Europe and eventually the world.<sup>4</sup>

By February 1948, the Soviet Union had completed its network of proxy states in Eastern Europe, as communists supported by Moscow seized control in Czechoslovakia. In June 1948, the Soviets blockaded land routes from the western zones of Germany to Berlin, forcing the United States and its allies to provide supplies to Berlin by an extensive airlift. Elsewhere, the Soviets were fomenting other Marxist movements in Western Europe, Africa, South America, and Asia. In addressing this new threat, the United States initiated the rebuilding of Western Europe through the Marshall Plan and, working with its allies, established the North Atlantic Treaty Organization. Other regional pacts, such as the Central Treaty Organization and the Southeast Asia Treaty Organization, were also created to form a fence around the Soviet Union. These developments established the social, organizational, and political bases required to counter Soviet plans and strategies for acquiring more territory and proxy states.

However, the ultimate guarantor that the Soviets could not advance into Western Europe was the deployment of tactical nuclear weapons developed at Los Alamos and Lawrence Livermore National Laboratories. These weapons represented the most inexpensive and effective means for stalemating the Soviets and deterring any expansionism designs they might have harbored. Fundamentally, these

<sup>4</sup> Secretary Forrestal was the architect of much of the defense structure set up in the Truman Administration to counter the Soviet Union. He left office on March 28, 1949, and died tragically, taking his own life, less than two months later.

## The Children's Milk Fund

At the 1986 Blacksburg Conference, “nuclear winter” became “nuclear autumn” and subsequently disappeared from public debate as an issue. Since our strategic force modernization had the effect of reducing the possibility of such an effect even further, I had been accused of high-jacking nuclear winter to support our modernization efforts. In any case, I had been invited to speak at the conference. When my session was over, it was lunchtime. I picked up my lunch and walked into the lunchroom. I noticed that former Senator Albert Gore, Sr., was sitting by himself over in the corner of the room. I walked over and asked Senator Gore if I could join him for lunch, and he replied, “Sure, general, sit down.”

“Sir, I am a colonel.”

“You should be a general, and I hereby promote you to that rank for the term of our lunch. What do you do in your present assignment?”

“Sir, I work in the Office of the Secretary of Defense as the Special Assistant for Air Force Nuclear Matters. That means I staff issues pertaining to Air Force nuclear weapons.”

“I have always liked nuclear weapons. General, do you want to know why?”

“Yes sir, I would!”

“In 1940, I was a young congressman from Tennessee, serving in several committees that arranged funding for public services and works. One day, Speaker Sam Rayburn called me into his office. Albert, he said, I want you to hide a couple hundred million dollars in the federal budget. No questions asked, I left Speaker Rayburn’s office and immediately started putting two million dollars here and five million dollars there. There was a spike in the children’s milk fund, the highway program was accelerated, and more dam projects were authorized than we had water to fill them. I never stopped to ask how this money was really being used.”

“In 1945, I and several other congressmen were on a trip to the Pacific to see how the war was going. Before landing on Tinian Island, we had flown over hundreds of warships and troop transports that were stacked up awaiting the imminent invasion of the Japanese mainland. I knew that those ships held thousands of good ole Tennessee boys, and

I knew that many of those boys would never live to see the green hills of Tennessee again. I felt extremely saddened by that prospect. Upon landing, we were rushed into a large briefing room. A general officer, whose name I have forgotten, briefed us on plans for the invasion. He told us that the troops would hit the beaches first at Kyushu and then at Honshu. In defense of their homeland, the Japanese would put up intense resistance. Casualties on both sides were expected to number millions. The general told us that three large hospitals had been built on Tinian to receive the wounded. The central corridor in the largest of these hospitals was over a mile long. We sat stunned and silent. At this point, General MacArthur strode in as only he could do. MacArthur dismissed the other general. He looked at us and said, ‘Gentlemen, the war will be over before you get back to California.’ With that pronouncement, he left as suddenly as he had appeared. We went from stunned to confused. We thought, ‘how can this be?’ Millions of casualties filling up the Tinian hospitals, and the war will be over before we get home?’”

“We departed from Tinian and island-hopped east toward Hawaii. When we landed at Hickam Air Base, someone handed me a newspaper. The banner headline read, Secret Atomic Bomb Destroys Hiroshima. ‘The children’s milk fund,’ I shouted, ‘the children’s milk fund!’ By the time we left Hawaii, Nagasaki had been destroyed by a second atomic bomb. When we landed in San Francisco, California headlines gave us the news that Japan had unconditionally surrendered and the war was over. Those Tennessee boys would live to see the green hills of Tennessee again and possibly even vote for me. That, general, is why I like nuclear weapons.”

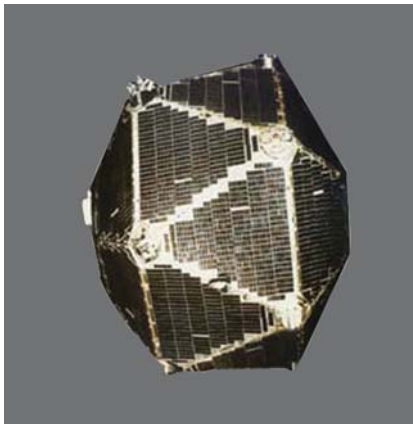
“Senator, I also like nuclear weapons. My father was on one of those ships that your party overflew before landing at Tinian.”

For me, hearing the elder Senator Gore relate this history is one of those precious moments that one never forgets. He doubted that such a secret enterprise could be accomplished in the political environment of today. He went on to say that he and his colleagues eventually hid over two billion dollars in the federal budget that he subsequently found out was used to build and operate Hanford, Oakridge, and Los Alamos. We were part of Senator Gore’s children’s milk fund, an interesting footnote in our 60-year history.



weapons confronted the Soviet juggernaut with the prospect that military aggression could result in an escalation to strategic levels and the total destruction of the Soviet Union. Once contained and its wealth-acquiring strategies thwarted, the Soviet Union began to atrophy and finally died from within. The Berlin Wall came down, and the nations of Eastern Europe and Russia joined the community of free nations. There is irony in this outcome. Because tactical nuclear weapons stood as silent sentinels of freedom, the critical role they played in establishing freedom in Eastern Europe and Russia is seldom understood or appreciated.

Of course, Los Alamos made other contributions to this outcome. The Laboratory, working mainly with Sandia National Laboratories, developed sensors launched on the Vela satellites to verify compliance with nuclear treaties such as the 1962 Limited Test Ban Treaty (refer to Figure 1 and the article “Eyes in Space” on page 152). These sensors provided data that helped build confidence that the treaties were working. They performed their mission in an outstanding manner, by detecting every atmospheric nuclear explosion within their field of view.



**Figure 1. Vela Satellite**  
Launched between 1963 and 1969, the Vela satellites verified compliance with the atmospheric test ban.



**Figure 2. The Berlin Wall**  
The Berlin Wall comes down in November 1989.

Significantly, none of those explosions was conducted by the Soviet Union. Other sensors followed, including those for the 1974 Threshold Test Ban Treaty and the 1988 Intermediate Nuclear Forces Treaty. In their own way, these verification systems helped reduce the atmosphere of fear and suspicion by increasing the opportunities for dialogue between the Soviet Union and the United States. Thereby, they helped achieve stability until the Berlin Wall came down (Figure 2).

### “Interesting Times” after the Cold War

Once the euphoria over the end of the Cold War abated, the world had to face new realities. The times have been definitely interesting, but are we also sometimes reminded of the Chinese curse that says, “May you live in interesting times”? One of those new realities was that a nuclear weapons superpower, the former Soviet Union, was in an economic meltdown. The system of balanced agendas that had been in place for decades was gone, and regional tyrants that had been kept under a

modicum of control by the old system began to act in irrational ways to establish regional hegemonies. Thus, the new age was more complex and unpredictable than the old bipolar world. Faced with these new realities, Los Alamos had to inventory its capabilities and redirect them, as appropriate, to address new threats and allay new fears. There were many strengths that we could muster. Immediately, we were asked to accelerate the use of Los Alamos safeguard systems to help manage and secure the very large inventories of excess special nuclear materials (SNM) that had been accumulating without the security nominally associated with nuclear weapons. These inventories had already been accumulating from the nuclear power industry. Plutonium, in particular, is being pulled out of spent nuclear fuel for easier long-term storage and for conversion into mixed oxide fuels. After the end of the Cold War, that inventory of SNM began to accelerate rapidly from the build-down of nuclear weapons inventories in the United States and Russia.

Within these new realities, we were also aware of the possibility of so-called “loose nukes.” Whatever the old Soviet Union’s proclivities were,

the Soviets did know how to protect their nuclear weapons. They had a very active program that combined their extensive and intrusive police powers and a robust transportation infrastructure with a disciplined, well-compensated cadre of warrant officers dedicated to the security of their weapons. However, after the collapse of the Soviet Union, the world was confronted with the possibility of nuclear weapons being sold or given to terrorists or proliferant states. This potential was particularly ominous in the early days of the collapse, when the economic situation was so dire that nuclear-armed units abandoned their weapons to forage for food. Although conditions have improved significantly since then and the Russian economy is on the upswing, the concern still remains.

It was concern over the situation in Russia that gave rise to the Nunn-Lugar-Domenici legislation.<sup>5</sup> Los Alamos scientists working under this legislation with their colleagues from other national laboratories have already accomplished some outstanding achievements in these areas, but much more remains to be done. We have had seminal successes at such places as Aktau in Kazakhstan and Novouralsk in the Russian Urals. At Aktau, on the Caspian Sea, for example, the Russians withdrew and abandoned a BN-350 nuclear reactor (see Figure 3). This site, not far removed from Iran, had enough weapons-grade plutonium in its cooling ponds for making a significant number of nuclear weapons. Now, that material is within secure boundaries under a system of positive safeguards and



**Figure 3. Kazakhstani-American Cooperation at Aktau**

**At Aktau in Kazakhstan, the Russians abandoned the BN-350 nuclear reactor, leaving unattended significant amounts of weapons-grade plutonium in its cooling ponds. This dangerous situation was averted in a cooperative Kazakhstani-American effort that secured the material under a system of safeguards and accountability. (Inset) Radiation sensor electronics built in the United States are being assembled at the reactor site. The sensor will be used by the International Atomic Energy Agency to monitor the movements of nuclear materials at the BN-350 nuclear reactor.**

accountability. Working with the Russians in Novouralsk, we have been able to blend down a very large amount of highly enriched uranium (HEU) to a level of enrichment that cannot be used for nuclear weapons. Los Alamos developed the monitoring system for verifying that the feedstock was HEU. From the start of the blend-down program until September 2002, the inventory of HEU has been reduced by 150 metric tons. That amount equates to over 6000 nuclear weapons being taken off the table and permanently removed from the grasp of potential terrorists.<sup>6</sup> In the flow of

Western history, we succeeded in this endeavor by working shoulder to shoulder with our Russian colleagues. Its magnitude can be compared with that of the Greek victory over the Persians at Plataea or Charles Martel's defeat of Islamic forces under Abd al-Rahman at Poitiers. Our accomplishment, expressed as a net reduction in military potential, is the most significant disarmament in history. However, it is seldom, if ever, mentioned in the public media. Of course, we are not yet where we need to be. Other weapons-usable material is still out there to be secured, and we cannot wait for laurels or applause. The battle for a safer world is a continuing one, and it is accomplished by dedicated scientists and engineers working under often difficult conditions to secure nuclear material kilo-

<sup>5</sup> The original Nunn-Lugar Bill (Soviet Nuclear Threat Reduction Act of 1991) was passed by Congress after the collapse of the Soviet Union to provide U.S. aid in denuclearizing and demilitarizing Soviet systems. The Nunn-Lugar-Domenici Program was authorized in the National Defense Authorization Act for fiscal year 1997.

<sup>6</sup> The Blend-down Program calls for 150 metric tons of HEU to be converted to low enriched uranium for use as reactor fuel. With that amount of HEU, 20,000 nuclear weapons could be made.

gram by kilogram to preempt possible future battles accomplished kiloton by kiloton.

### Alabaster Cities and Human Tears

Terrorism experts have suggested (Jenkins 1987) that terrorists had a social contract with society not to kill a lot of people, just enough people to seize the headlines. The events of September 11, 2001, abrogated any such social contract and introduced us to an unstable world in which foreign nonstate terrorists attacked our core defense and financial districts, causing an enormous loss of life and property.<sup>7</sup> In this new, more convoluted world, it became clear that new battles would be fought without the relative isolation we have historically enjoyed. The battles would be engaged on new killing fields—possibly in our town squares and certainly abroad, in distant mountains without names, in dusty streets seemingly without end—on our laboratory benches as we develop technologies to shape the engagement of battle, and ultimately in our minds and hearts.

Within this more unstable world, the possibility of terrorism involving weapons of mass destruction (WMD) looms larger than ever before. Los Alamos has already been called upon to provide its many well-established capabilities to address this new threat. Among these capabilities are our in-depth understanding of nuclear weapons and materials, a mature program in detecting and characterizing pathogens, our demonstrated expertise in modeling and simulating complex infrastructures and systems, and our

<sup>7</sup> The Aum Shinrikyo planned to produce sufficient sarin nerve agent to kill everyone in Japan. Arguably, the “social contract” would have been probated earlier had those plans been executed.

developments in detecting and neutralizing chemical agents. Success in preventing WMD terrorism requires developing an integrated approach to reduce the possibility that such an untoward threat against our people and facilities could occur. Whittling away at that possibility requires that we have concurrent and synergistic activities that multiply our efforts and investments. In the long run, occurrence of a WMD terrorist event can hopefully be reduced to manageable proportions. To that end, we must continue with preemption and dialogue, while not forgetting that an absolute assurance that such threats can be completely avoided would surely exhaust resources needed for addressing other pressing problems. Achievable or not, absolute assurance will certainly always be the goal, and Los Alamos science and technology will be essential in reaching that goal.

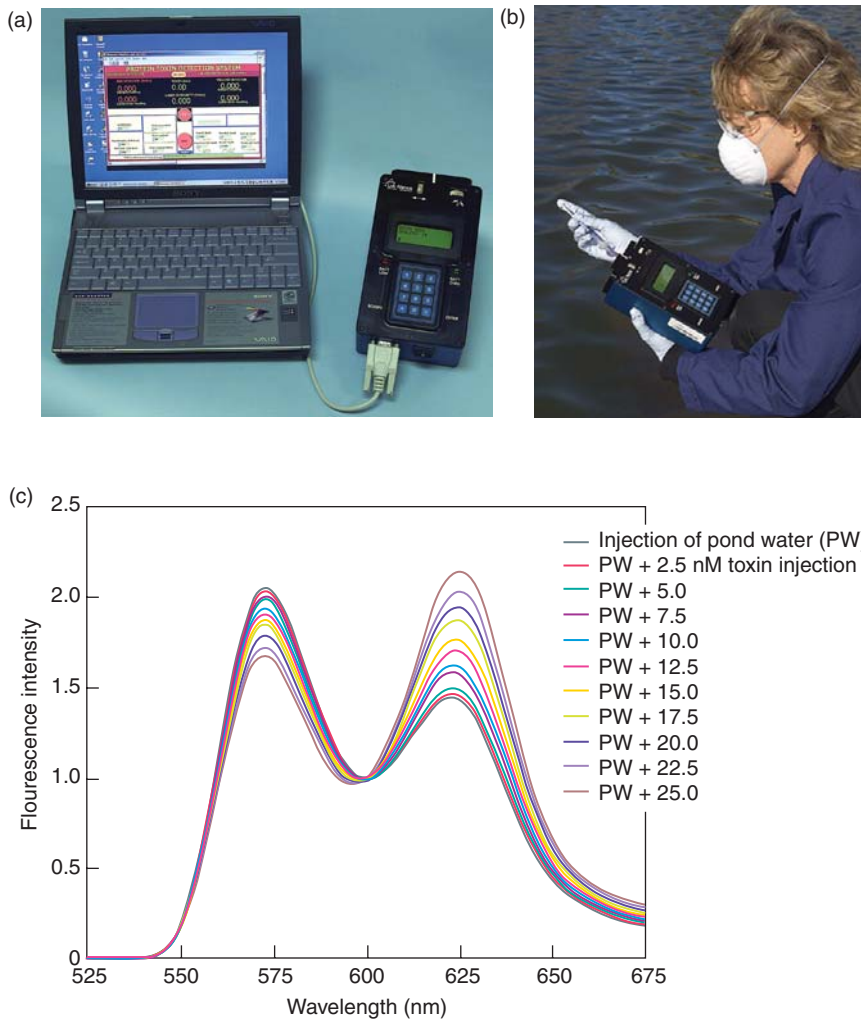
### Role and Limits of Science and Technology

Fundamentally, the war against terrorism is a war of ideas. In this war, terrorists know that they cannot defeat the United States, but they can escalate violence to the point at which they hope we destroy ourselves. We lose automatically if we conclude that being free and being secure are mutually exclusive. Certainly, we cannot defeat terrorism if we continue to siphon profit out of our economy and sacrifice our freedom and liberties upon the altar of good intentions. Responses to terrorism must be designed and executed to make our economy stronger and more efficient and our freedoms and liberties more robust and expressive. If properly applied and planned, science and technology can help our country achieve security by avoiding unnecessary intrusion into civil liberties and privacy. As the following examples show,

new technologies, many now available or under development at Los Alamos, coupled with innovative policies and appropriate implementation, can move us in the proper direction.

It is well known that repetitive manual security procedures at our airports are a significant overhead on the national economy. Adding one hour to the airport check-in procedures will rob the nation of 600 million hours of productive time each year, a figure that approximates the human productivity lost in the deaths that occurred in the collapse of the World Trade Center towers, not considering the horrendous human tragedy involved. Adding 10 minutes to the time it takes to download and transship individual cargo containers can be the difference between making a profit and incurring a loss. Such problems could be solved by an automated scanning system based on GENIE,<sup>8</sup> a genetic algorithm-based recognition technique developed at Los Alamos. (Refer to page 158 for a detailed description of how GENIE works.) GENIE can automatically process digital images fast and can ascertain potential threats at any viewing angle with very high confidence. For example, surveilling large areas for specific features with appropriate resolution requires extraordinary amounts of digital image data, but with the help of GENIE, human analysts can extract features of interest automatically, thus being able to keep up with the flood of high-quality imagery and technical data collected by satellites. In one test, researchers asked GENIE to locate every golf course of Professional Golfers’ Association caliber in the United States. Normally, this task would have required a large

<sup>8</sup> GENIE (Genetic Imagery Exploitation) received an R&D 100 Award in 2002. These awards are given by the *Research & Development Magazine* for the best 100 innovations of the year.



**Figure 4. The Reagentless Optical Biosensor (ROB)**

(a) ROB is a self-contained hand-held system that can detect pathogens and quantify the amount present. Samples are placed in a disposable sensor cartridge. Different cartridges will be designed to detect different protein toxins. Later, data can be downloaded for storage in our protein toxin database. (b) ROB was tested on pond water spiked with different amounts of cholera toxin. (c) The graph shows the fluorescence spectra measured by a fiber-optic spectrometer for different concentrations of cholera toxin.

team of photoanalysts and months of eye-straining work. GENIE finished the job in about one hour.

Similar to biological systems, GENIE's genetic-algorithm-based scanning system actually improves as it mutates, and it never becomes fatigued. Because each system mutates uniquely, terrorists will find it difficult to employ countermeasures because they will not know what spe-

cific criteria the system is using at any given time to find proscribed items. Using GENIE to scan luggage and persons will not require having another human to review results of the scan unless, of course, a proscribed item is identified. This feature is important in protecting privacy.

Science and technology applied as responsive actions to terrorism can be designed and implemented to result in

a stronger society as terrorist attacks increase. For example, investments in our public health services aimed at dealing with acts of bioterrorism, if properly planned, can help ensure that more capacity will be available to deal with natural pandemics. We are developing systems, operating at the carbon/silicon interface, that combine the antigenic recognition capabilities of natural cells with the information processing speed of modern electronic systems. These detectors will permit rapid diagnosis of pathogens in the physician's office—no endless hours of waiting for results from pathogen cultures any longer. The reagentless optical biosensor (ROB) developed at Los Alamos is one such example (refer to Figure 4). ROB uses optically tagged natural receptors embedded within an artificial cell membrane to detect medical and environmental pathogens. Figure 4 shows ROB being used in the field to detect cholera, a pathogenic protein.

While detectors might be deployed to protect against bioterrorism, they can also identify such naturally occurring pathogens as the hantavirus. The hours saved in identifying this particular virus can be the difference between surviving the infection and dying from it.

Science and technology can be used to simulate complex situations—for example, those that permit national policy makers and legislators to authorize improvements designed to protect critical infrastructures against cyberterrorists. Such simulations might, at the same time, provide a more capable and secure information architecture for businesses and private citizens. For example, Los Alamos and Sandia National Laboratories have partnered to establish the National Infrastructure Simulation and Analysis Center in order to provide improved technical planning and decision support for the analysis of critical infrastructures. Simulation approaches

developed at the center will permit effective routing of first responders, efficient allocation of resources, and effective defense options and strategies. This approach, although focused on counterterrorism, can be used to identify vulnerabilities that could grow out of natural disasters as well. The net result can be more robust and effective national infrastructures.

However, science and technology have limits on what they can accomplish. For example, they cannot deliver a solution proscribed by the laws of physics and chemistry. If we are required to assay a package passively for the presence of a radiological material, neutrons and gamma rays will behave like neutrons and gamma rays, and rates of radiological decay are fixed in nature. In like manner, detection of a lethal amount of some pathogens, such as *Yersinia pestis* or *hemorrhagic variola*, would require detection of a single microbe, a difficult task in any situation and an impossible one if the microbe were placed inside an airtight package. In addition, science and technology can present national policymakers with difficult choices. For example, detectors placed in the cargo compartment of a large airliner can, with enough integration time, locate and characterize SNM hidden in luggage. Because the detectors would probably not be able to define the configuration of that material, the national policymaker would have to decide what actions should be taken in the face of valid but inconclusive information. The consequences of making the wrong decision can be enormous. Finally, although all the examples discussed before are compelling proof of the key role of science and technology in preventing terrorist attacks, no combination of science and technology can provide absolute assurance that some clever or lucky terrorist will not succeed in carrying out a deadly attack against our citizens.

## At the Crossroads

Comedian Woody Allen once remarked, “More than at any time in history mankind faces a crossroads. One path leads to despair and utter hopelessness, the other to total extinction. Let us pray that we have the wisdom to choose correctly.” Two futures certainly lie before the free world, but unlike those referred to by Woody Allen, at least one is not as bleak. However, one possible future is indeed bleak and frightening. This is a future in which terrorism fed by radicalism and hatred has become a more significant challenge to our society and its values. It is a future in which vehicle bombs—for example, tankers or aircraft loaded with fuel—and cyberterrorism have severely damaged or destroyed one or more critical national infrastructures. This future could also include terrorism involving chemical and biological agents with attacks that are increasingly lethal and the possibility that terrorists have acquired materials for nuclear weapons and stolen nuclear weapons. In this possible future scenario, terrorism will have fundamentally changed our way of life, and the rationale for sustaining our freedoms and liberties would certainly be questioned. This future will happen if we allow fears, real or imaginary, to drive us down irrational paths, to dim our support of democratic principles, to bind our response capabilities in endless minutiae and inane agendas, and to abandon our technological and scientific strengths.

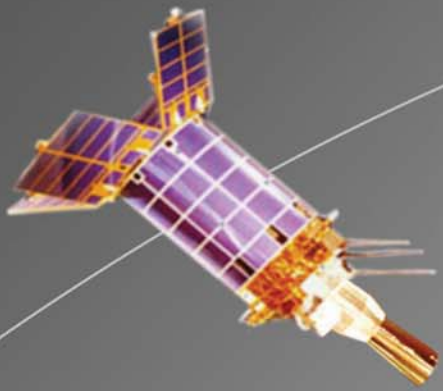
But then, as we have done over the years, we can navigate successfully amid realistic fears toward a more desirable future. In this second scenario for the future, which is within our reach and abilities, science and technology have made acts of terrorism less probable and more costly to the terrorists and have reduced the consequences of possible terrorist

acts. Science and technology have ameliorated the impact of counterterrorism measures on our basic freedoms so that we can be both free and secure. Science and technology have made the world safer with respect to terrorism and more robust and capable with respect to natural disasters and pandemics. Finally, the underlying factors and fears that made terrorism an option for achieving social change have been eliminated by the successful application of science and technology to improve dialogue, quality of life, and opportunity. This is the future in which Los Alamos skills and talents can and must play a role. That role, the reduction of threats and fears, has engaged Los Alamos throughout its 60-year history. ■

## Further Reading

- Jenkins, B. 1987. The Future Course of International Terrorism, *The Futurist* **21** (4): 8.
- Kikuchi, S., H. Aoki, and T. Wakatuki. 1939. Angular Distribution of Fast Neutrons Scattered by Atoms. *Proc. Phys. Math. Soc. Japan* **21**: 410.
- Pipes, R. 1974. *Russia under the Old Regime*. New York: Charles Scribner's Sons.

*For further information, contact Terry Hawkins (505) 665-1259 (hthawkins@lanl.gov).*



# Eyes in Space

## *Sensors for treaty verification and basic research*

*William C. Priedhorsky and contributors*

Space-based nuclear threat reduction began with the signing of the Limited Test Ban Treaty (LTBT) in 1963. The treaty prohibited nuclear tests in the atmosphere, outer space, and under water, and was a significant first step toward both slowing the nuclear arms race and curbing the environmental contamination associated with above-ground tests. But in the tense atmosphere of the Cold War, neither the United States nor the Soviet Union would trust that the other had complied with the treaty without a fool-proof method of verification.

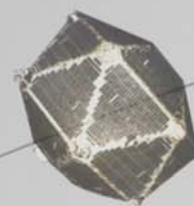
That method turned out to rely heavily on earth-orbiting satellites, each of which carried a bevy of sensors that would monitor the skies and unambiguously detect aboveground nuclear detonations. The Defense Advanced Research Projects Agency and TRW, Inc., were tasked with designing, building, and fielding these satellites called Vela for the Spanish word “velar,” meaning to watch. Los Alamos and Sandia National

Laboratories were entrusted with providing the all-important sensors.

Los Alamos was a natural choice to supply these sensitive “eyes” in space. Since the late 1950s, researchers at Los Alamos had used sounding rockets to hoist neutron, gamma-ray, and other detectors into the upper atmosphere in order to gather data from high-altitude nuclear tests. Those same instruments would be adapted for the orbital environment and the nuclear detonation (NUDET) detection mission. But numerous technical difficulties surrounded this new mission, as the sensors would be subject to a host of natural backgrounds and obfuscating signals. Would something as common as a lightning flash be confused with a nuclear event, or would something as exotic as gamma rays from a supernova<sup>1</sup> trigger the system?

<sup>1</sup>This latter question was originally posed by Stirling Colgate, now a senior fellow at Los Alamos, in a 1959 test ban summit meeting. Colgate perceptively recognized the connection between mission and basic research.

Similar to what is being done today to carry out the Laboratory’s missions, the scientists then applied their expertise to building a detection system that would behave as planned. They also initiated new research programs specifically designed to further an understanding of background sources and create sensors that could better discriminate nuclear explosions from natural signals. Furthermore, they realized from the onset that a system that was sensitive to, say, lightning could be used to study lightning. Soon, scientists were using NUDET sensors to conduct world-class research in atmospheric science, space-plasma science, and even astrophysics.



Los Alamos detectors have since journeyed over the poles of the sun, probed two comets, and flown to several planets to gather data for basic research. In turn, these highly visible space missions aid our nation's threat reduction efforts. Consider, for example, the recent discovery of water on Mars, discussed in more detail in the sidebar "Geochemical Studies of the Moon and Planets" on page 166. Scientists announced this find after analyzing the neutrons coming from the Red Planet. Because they are generated by cosmic rays bombarding the Martian surface, the neutrons have a known energy spectrum, which becomes slightly distorted if they collide with water molecules. Those spectral distortions were "seen" by the highly advanced neutron spectrometer on the orbiting Mars Odyssey satellite. While the discovery of water on Mars justly fuels the public's imagination and promotes basic research, it also reminds other nations of the United States' remarkable capabilities in neutron detection, in case any nation needs reminding.

NUDET detection for treaty verification and situational awareness remains a Los Alamos mission. The radiation detection system on the Air Force's Defense Support Program (DSP) satellites and the Global Positioning System (GPS) satellites—the same satellites that give us hand-held navigation—are being used for NUDET detection. The last DSP satellite will be launched in 2003 or 2004, after which the next generation of GPS satellites, and perhaps another system, will carry on the mission.

The end of the Cold War, however, changed the world. We needed to assess the capabilities of aspiring

nuclear states long before any bomb was detonated and to address problems of nuclear materials control and international terrorism. This broader concept of nuclear threat reduction required new sensing capabilities: new small satellites for space observations, new sensors to monitor effluent streams from factories and power plants, portable sensors for materials trafficking, and sensors that could operate in cyberspace to detect subtle patterns and connections in large masses of data. Like the NUDET systems, these advanced technologies double as research tools and have led to more discoveries of our planet, the solar system, and the cosmos.

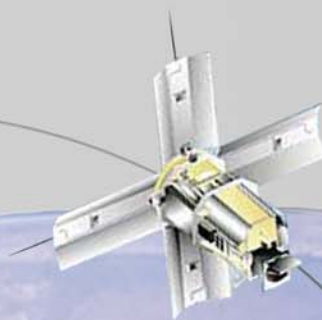
### Space-Based Nuclear Event Detection

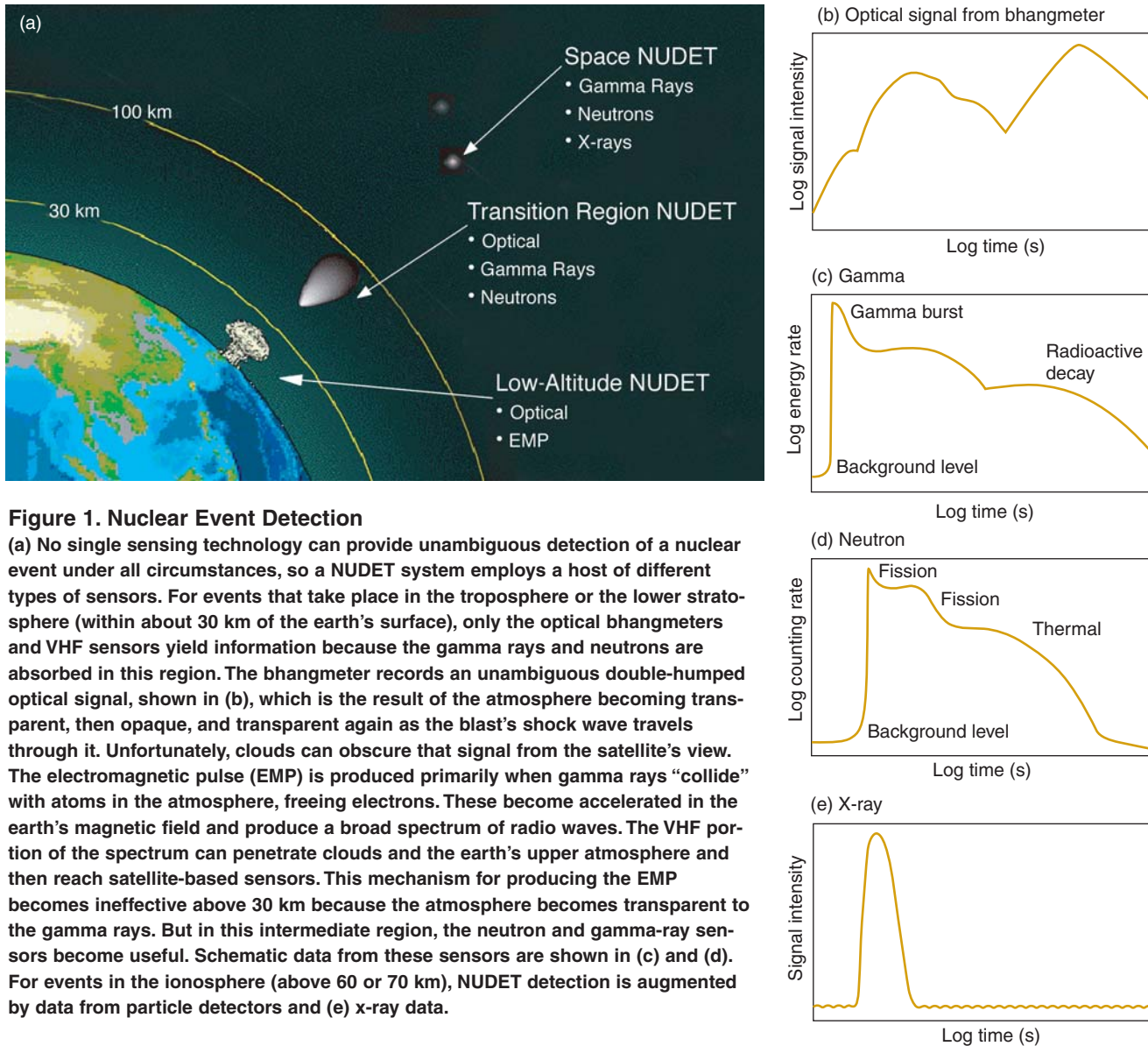
Remote detection of nuclear explosions is accomplished with sensors that measure the different forms of energy coming from the weapon. Neutrons, gamma rays, and x-rays are emitted promptly within about 2 milliseconds of the detonation. Those radiations then interact with their surroundings to produce secondary radiations, including visible light and electromagnetic pulses (EMPs), in the part of the radio-frequency (rf) band below a few hundred megahertz. Delayed gamma rays and neutrons also come

from the nuclear debris. Both the prompt and delayed radiations can be detected by satellite-borne sensors: bhangmeters<sup>2</sup> for detection of optical signals, very high frequency (VHF) radio receivers for measurement of the EMP, plus neutron, gamma-ray, and x-ray detectors.

These sensors studded the surface and filled the insides of the Vela satellites, which were the first used to verify the LTBT. The Velas operated in pairs, with satellites occupying opposite sides of a nearly circular orbit that lay about one-third of the way between the earth and moon. Their sensitive instruments could see the entire surface of the earth, as well as a large region of space surrounding the planet.

<sup>2</sup>The name "bhangmeter" possibly derives from bhang, the Indian name for a type of marijuana. Apparently, some believed that anyone who thought satellite-based optical detection would work must have been smoking something. Equally likely, "b-hang" derives from a two-syllabic way of pronouncing "bang." This pronunciation mirrors the detection of the two distinct optical peaks (one short and one long) characteristic of an atmospheric nuclear explosion.





**Figure 1. Nuclear Event Detection**

(a) No single sensing technology can provide unambiguous detection of a nuclear event under all circumstances, so a NUDET system employs a host of different types of sensors. For events that take place in the troposphere or the lower stratosphere (within about 30 km of the earth’s surface), only the optical bhangmeters and VHF sensors yield information because the gamma rays and neutrons are absorbed in this region. The bhangmeter records an unambiguous double-humped optical signal, shown in (b), which is the result of the atmosphere becoming transparent, then opaque, and transparent again as the blast’s shock wave travels through it. Unfortunately, clouds can obscure that signal from the satellite’s view. The electromagnetic pulse (EMP) is produced primarily when gamma rays “collide” with atoms in the atmosphere, freeing electrons. These become accelerated in the earth’s magnetic field and produce a broad spectrum of radio waves. The VHF portion of the spectrum can penetrate clouds and the earth’s upper atmosphere and then reach satellite-based sensors. This mechanism for producing the EMP becomes ineffective above 30 km because the atmosphere becomes transparent to the gamma rays. But in this intermediate region, the neutron and gamma-ray sensors become useful. Schematic data from these sensors are shown in (c) and (d). For events in the ionosphere (above 60 or 70 km), NUDET detection is augmented by data from particle detectors and (e) x-ray data.

All told, six pairs of Vela satellites were launched between 1963 and 1969. The initial pair (Vela 1 and 2) carried only x-ray, neutron, and gamma-ray detectors. These would see any events that occurred high in the atmosphere (above about 30 kilometers) and also in space (see Figure 1). Even a detonation on the far side of the moon would be detected because the nuclear blast would expel a gamma-ray-emitting cloud of debris that would quickly be seen.

These first detectors were used as

much for system shakedown as for treaty verification. Far from being empty, the space between the sun and the earth is filled with charged particles that boil from the sun’s surface and stream through the solar system at supersonic speeds (the solar wind). Interactions between the solar wind and the earth’s magnetic field create a tenuous and highly variable plasma, known as the magnetosphere, which surrounds the earth. The Velas’ orbit would carry them through the magnetosphere, but in 1963, little was

known about that plasma region or about the effects of that region on sensitive instruments. (The Velas were also subject to hostile cosmic radiation, which comes from outside the solar system. Thus, many skeptics gave the instruments no more than two weeks to live. But most instruments lasted well beyond their design lifetime of six months; some, for as long as a decade.)

Adopting a bootstrap approach, scientists used the data from the first Vela satellites to design new types of



sensors that would monitor the plasma background and track particle fluxes that could cause false signals in the other detectors. These plasma and energetic-particle sensors, plus bhangmeters and VHF sensors that could identify explosions that took place in the lower atmosphere (refer to Figure 1), were fielded along with the other detectors on the Vela 3 through 6 satellites. The last three pairs of satellites (officially known as Advanced Velas) carried improved NUDET systems, plus sensors that monitored solar activity, terrestrial lightning, and celestial x-rays and gamma rays.

As a series, the Velas worked superbly and were widely considered to have seen every aboveground nuclear explosion that was within their field of view. They established the benchmark for surveillance capability, but their legacy was also one of scientific discoveries. As discussed later in this article, much of our early data on the solar wind was obtained by the Velas' particle detectors, whereas their gamma-ray detectors were the first to observe cosmic gamma-ray bursts, an entirely unknown phenomenon that opened a new doorway into the observable universe.

Starting in the 1970s, the Air Force DSP satellites began carrying NUDET systems, which were continually upgraded for sensitivity, dynamic range, and background rejection. But the basic instruments remained the same as those on Vela, even though extending system capabilities into the extreme ultraviolet (soft x-ray) and rf bands had always been a goal. By the late 1980s, we had concepts for new sensors to operate in those extended frequency bands.

Unfortunately, these new devices presented us with a problem. While we could verify their operation in the laboratory, in space they would be subject to large and poorly understood backgrounds. We needed to test them

## The Little Satellite That Could

*Diane Roussel-Dupré*

The two years 1985 and 1986 were bad ones for the U.S. space program. Three major launches failed, and on January 28, 1986, the space shuttle Challenger exploded in full view of the entire world. These calamitous failures stopped all U.S. space launches for more than a year and left the space community cautious and conservative.

Quixotically, it was during this guarded period that our young experimental team at Los Alamos chose to field the Laboratory's first satellite. The ALEXIS satellite was designed to test new soft x-ray and radio-frequency nuclear detonation (NUDET) detectors. It was funded by the Department of Energy and launched by the Air Force Space Test Program. The rocket was the new Pegasus launch vehicle, which had mixed success on its first three outings. Its fourth launch on April 25, 1993, however, went well, and our rocket gracefully ferried ALEXIS aloft to an 800-kilometer circular orbit. But the satellite itself ran into complications caused by the launch forces.



**Figure A. ALEXIS**  
The ALEXIS satellite demonstrated new technologies for treaty verification while carrying out basic research in astrophysics and atmospheric science.

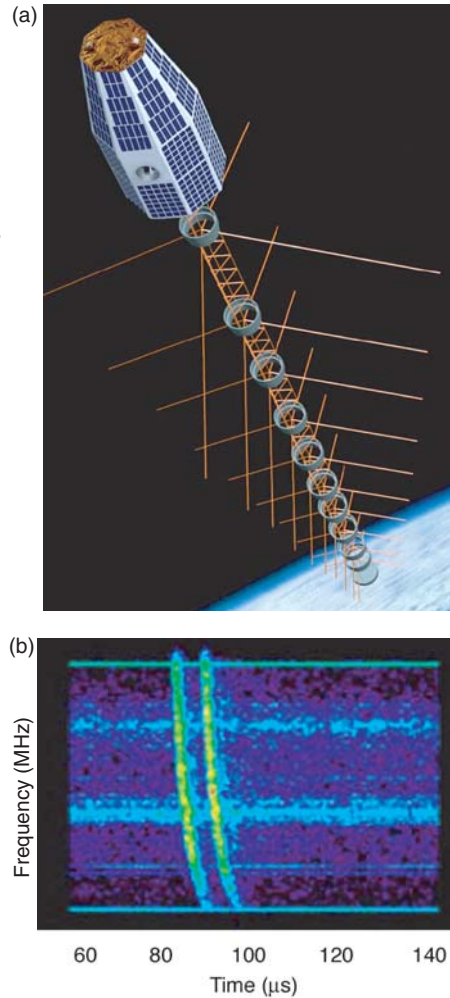
The Pegasus rocket was outfitted with a video camera to monitor the rocket performance and reveal whether the nose cone deployed cleanly. To our horror, the video footage that was transmitted back to the California tracking station showed that one of the solar panels on ALEXIS had broken loose at the hinge and was dangling freely. We could not tell from the video whether any other damage had occurred or whether the satellite was dead or alive. The first attempts to contact the satellite yielded nothing but silence, feeding our team's worst fears.

For six frantic weeks, the team listened for a signal every time the satellite passed over our Los Alamos ground station. We took a second ground station to an Air Force facility, trying to "shout" at the satellite with a bigger dish. We took pictures of our satellite from the Air Force optical tracking station on top of Haleakala in Hawaii to learn about its status, and we optimized our contact strategy. Our persistence finally paid off with a brief contact from our Los Alamos station, followed by a longer contact and an understanding of the satellite's problems. We formulated a recovery plan, and ALEXIS revived as expected.

ALEXIS was planned as a high-risk, one-year mission. However, as ALEXIS approaches its 10th birthday, it is still fully operational, operated by an automatic ground station in the Physics Building at Los Alamos. The solar panels are losing the ability to provide charge to the batteries, the commercial nickel-cadmium batteries have some trouble charging, and protons from recent solar storms have damaged parts of its memory, but ALEXIS is still "the little satellite that could."

### Figure 2. The FORTÉ Satellite and a View of Lightning

(a) The sketch shows an artist's conception of FORTÉ in orbit. The radio antenna, which is pointed toward the earth, is deployed to 11 m in length from a storage container the size of an office wastebasket. (b) This plot shows frequency vs time for a particularly strong NBE collected by FORTÉ. The two pulses correspond to the direct pulse from the lightning and an echo reflected from the ground. The spacing of the two pulses can be used to infer the source height. Free electrons in the earth's ionosphere cause the lower-frequency components of the signal to arrive later than the higher-frequency ones. We quantify and remove this effect to deduce when the event would have arrived at the satellite if the ionosphere were absent. If we see the event from four or more satellites, we can use these timings to solve for  $x$ ,  $y$ ,  $z$ , and  $t$  and locate the event in three dimensions.



in space, but unproved instruments could not be deployed on a commercial or military satellite—the cost of failure was too high.

Unable to fly these sensors on someone else's satellite, we chose to fly them on our own. We assembled a small, dedicated team to design and build Los Alamos' first satellite and called in Sandia National Laboratories and AeroAstro, Inc., a start-up small satellite company, to help. Named ALEXIS (for array of low-energy x-ray imaging sensors), our satellite was launched in 1993. The first of the "faster, cheaper, better" satellites, its sophisticated design included a novel uplink/downlink protocol, similar to the file transfer protocol used on the Internet, which allowed us to have very simple, inexpensive antennae on

the spacecraft and on the ground and to run the satellite almost autonomously. ALEXIS was the first satellite for which the weight and volume of the scientific payload was greater than the nonpayload (batteries, solar panels, structural components, and others) remainder of the spacecraft, and was one of the first to use computer memory instead of a tape recorder for data storage. After a somewhat shaky start—recounted in the box "The Little Satellite That Could" on page 155—it performed beautifully.

ALEXIS carried a set of soft x-ray imaging telescopes and an rf receiver, called Blackbeard, that was intended to help us understand lightning events. Lightning is a common background for our VHF sensors because

the intense electrical discharge produces a burst of rf noise that can mimic the nuclear EMP. The flip side is that our VHF/EMP sensors are excellent lightning detectors that can be used for basic research.

Blackbeard, for example, enhanced our understanding of how the ionosphere modifies lightning-induced rf pulses that pass through it and, in the course of its operation, discovered TIPP's (for transionospheric pulse pairs), or doublets of brief, transient rf events that form in energetic thunderstorms at 8 to 10 kilometers above the earth's surface.

Other successes soon followed. When compared with ALEXIS, the FORTÉ (for fast on-orbit recording of transient events) satellite, which was launched in 1997 and is still operational, was a step-up in size and sophistication. Its primary mission was to demonstrate new rf detection technologies that were to be at the core of the V-sensor, a new EMP sensor that will fly on the next generation of GPS satellites. Over the years, FORTÉ mapped optical and rf backgrounds, tested detection algorithms, and provided a wealth of data on the physics of lightning and the ionosphere.

One of the first and most basic of FORTÉ's findings was an explanation for the TIPP's observed by Blackbeard. A lightning discharge between clouds in the troposphere (the roughly 20-kilometer-thick atmospheric layer closest to the surface of the earth) produces an rf pulse that reflects from the ground, so that a pair of pulses is detected by the sensor. TIPP's are closely related to another unusual lightning phenomenon, narrow bipolar events (NBEs), which are intense, in-cloud rf events that occur during thunderstorms and last less than about 20 microseconds (see Figure 2). They are the brightest, most common form of lightning seen by our orbiting sensors.



**Figure 3. Imaging with MTI**

The MTI is one of the most accurately calibrated thermal imagers ever launched into orbit. It gathers data in 15 frequency bands—from the infrared to the ultraviolet. (a) An optical image taken by MTI of a section of the Lake Ontario shoreline at Rochester, New York, reveals certain types of information, for example, the existence of offshore sandbars. (b) A thermal image of the same area reveals other features. In this false-color image, red represents hot temperatures, whereas blue represents cool ones. We can now see a plume of hot water from a water treatment plant entering the lake. Data from all spectral bands give us valuable information for detecting and characterizing an area or facility.

As it turns out, the occurrence rate and source height of NBEs are excellent statistical indicators of the deep convective strength of the parent storm. Deep convection, or convection between the lower and upper troposphere, is the driving mechanism for several forms of severe weather on the earth and is a primary means by which energy—in the form of latent heat—drives the large-scale atmospheric circulation. It is also the primary means by which the atmosphere injects water into the stratosphere, where it profoundly influences the radiative and chemical balance of the atmosphere. Once the new V-sensor is in orbit, we will be able to use its data to map atmospheric deep convective processes in a near-real-time, global manner, particularly over oceanic regions where weather radar coverage is limited. Such maps will be used to support commercial and military aviation.

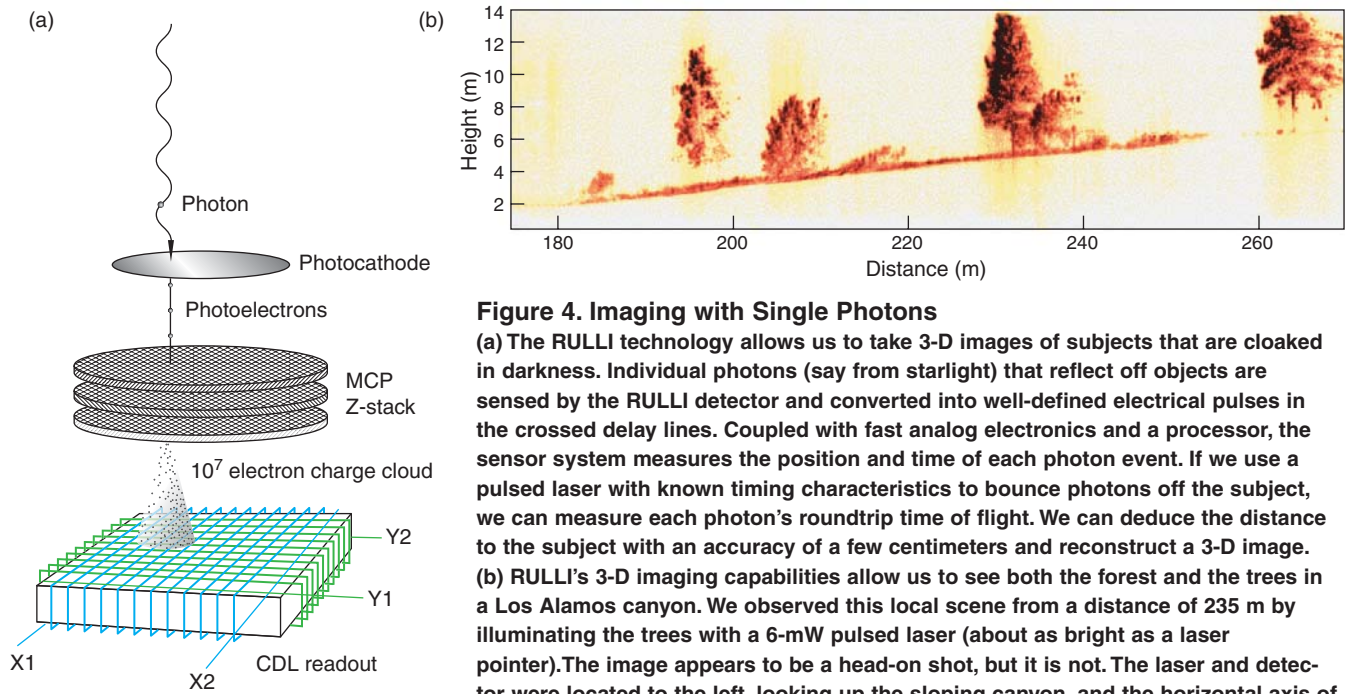
**Advanced Systems.** Although we are still advancing the science of nuclear event detection, the alarming rise of nuclear-capable states in the waning years of the twentieth century called for an expanded mission. We needed to develop surveillance systems that could be used for detecting and characterizing facilities that might be producing weapons of mass destruction. But gleaning information about an unknown facility is far more difficult than gleaning the specifics of a nuclear blast. The latter presents a well-defined signature of gammas, neutrons, and electromagnetic radiations, whereas the former oftentimes presents a patchwork of subtle signals that make sense only after detailed analysis. In general, a modern surveillance system will take images at several wavelengths, or spectral bands. Unfortunately, interpreting and piecing together the spectral information is often hindered by uncertainties in

the spectral calibration and by an inability to fully compensate for atmospheric effects.

The multispectral thermal imager (MTI), developed jointly by Sandia and Los Alamos National Laboratories and launched in early 2000, was meant to demonstrate advanced imaging and image-processing techniques that could be used in future systems. A major component of the MTI project was absolute calibration of the instrument, which is excellent and the best in its class. MTI takes data in 15 spectral bands, ranging from visible to long-wavelength infrared, which, when combined and analyzed, provide information about surface temperatures, materials, water quality, and vegetation health. Additional spectral bands provide simultaneous information about the atmosphere, such as the amount of water vapor and the aerosol content. All this information helps us construct the profile of a remotely located facility or area of interest (see Figure 3).

Multispectral data are also exceedingly useful for conducting basic earth-science research. The satellite doubles as a national and international resource that provides data to a large number of researchers. For example, MTI was used to study the volcanic eruption of Popocatepetl in Mexico in January 2001 and the effects of the Cerro Grande fire that swept through Los Alamos in May 2000. The MTI team at Los Alamos has built the Data Processing and Analysis Center to distribute data to the national user community.

Los Alamos scientists have also developed ground-based advanced imaging systems. Among them is RULLI (for remote ultralow light imaging), a single-photon detector and imager that can accurately and simultaneously measure the position and absolute arrival time of individual photons coming from a target area. The result is a data set that contains full three-dimensional (3-D) informa-



**Figure 4. Imaging with Single Photons**  
**(a)** The RULLI technology allows us to take 3-D images of subjects that are cloaked in darkness. Individual photons (say from starlight) that reflect off objects are sensed by the RULLI detector and converted into well-defined electrical pulses in the crossed delay lines. Coupled with fast analog electronics and a processor, the sensor system measures the position and time of each photon event. If we use a pulsed laser with known timing characteristics to bounce photons off the subject, we can measure each photon’s roundtrip time of flight. We can deduce the distance to the subject with an accuracy of a few centimeters and reconstruct a 3-D image.  
**(b)** RULLI’s 3-D imaging capabilities allow us to see both the forest and the trees in a Los Alamos canyon. We observed this local scene from a distance of 235 m by illuminating the trees with a 6-mW pulsed laser (about as bright as a laser pointer). The image appears to be a head-on shot, but it is not. The laser and detector were located to the left, looking up the sloping canyon, and the horizontal axis of this picture corresponds to distance. We can reconstruct this view through the trees only because we have full 3-D information.

tion about the area (see Figure 4). Because it can detect activities conducted under the darkness of night, RULLI and its successor technologies can be used for various threat-reduction applications, including airborne, large-area surveillance for perimeter protection.

### From Outer Space to Cyberspace

The body of data returned by advanced systems such as MTI, RULLI, or other signal collection and imaging systems is huge. Human analysts face the nearly impossible task of keeping up with this deluge. Increasingly, we must turn to computer-based image-processing tools to automate and assist in the analysis. But a computer’s ability to analyze image data pales in comparison with the remarkable human brain. Hence, we developed GENIE (for genetic image exploitation), a new software tool that

allows translating human knowledge into an algorithm that can recognize objects and patterns in data streams.

At its core, GENIE is a computer program that develops other computer programs (algorithms). It does so by using genetic programming techniques, which are methods for automatically creating a working computer program by combining, mutating, or rearranging low-level, nonspecific computer functions or programs. As its name implies, genetic programming draws its inspiration from biology, where new species emerge through the exchange and modification of chromosomes.

Training GENIE to find selected features in a data set is an iterative, evolutionary process. Starting with a small data set, or even a single image, an analyst marks features of interest—for example, all regions of water. Given this goal and a substantial library of low-level image-processing functions (for example, edge detectors or spectral filters), GENIE uses genetic programming techniques to produce

hundreds of algorithms, each of which finds (to some degree), the regions of water in the training set. The program ranks the algorithms according to a set of “best-fit” criteria.

Although the top-ranked algorithms may work very well, typically they do not find all the features of interest. The analyst then goes through the training set again, retagging missed features or flagging incorrect ones, and GENIE reworks the top-ranked algorithms. After a few such iterations, GENIE “evolves” an algorithm that is optimized to find the features of interest (see Figure 5). The analyst can retrieve the optimized algorithm in human-readable code, automate it, and use it to chew through large, complex data sets.

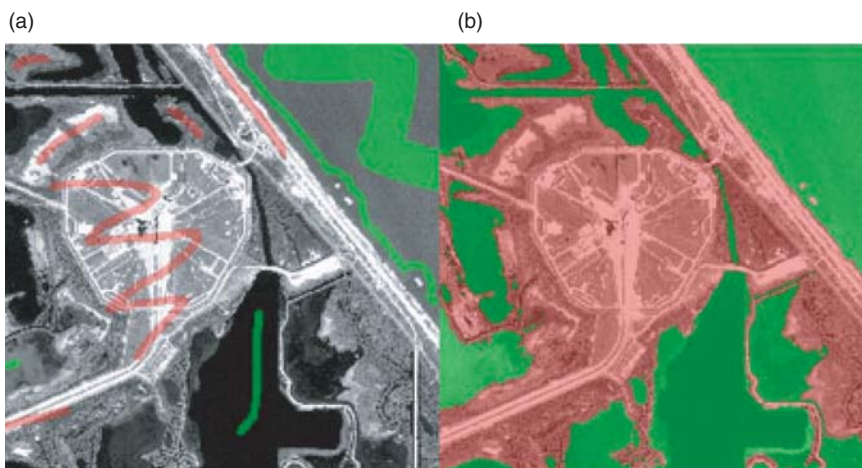
GENIE is a general-purpose tool for feature classification. Aside from threat reduction, it has been used successfully in detecting cancers and pathogens in humans, looking for topographic features and minerals on Mars, and mapping ash and debris

from the World Trade Center after the New York City terrorist attack.

Whereas GENIE enables us to create optimized software, meeting the demands of our expanding threat-reduction mission means optimizing hardware as well. We need to couple a sensor directly to a processor and have the system shoulder much of the real-time data analysis. Unfortunately, in trying to build such a system, we quickly run into size and power restrictions. A general-purpose processing board wastes valuable processing power and real estate because it provides capabilities that are extraneous to our purposes. Our data problems are so supersized that we need every hardware gate to be dedicated to solving our task. Field-programmable gate arrays (FPGAs) deliver this capability.

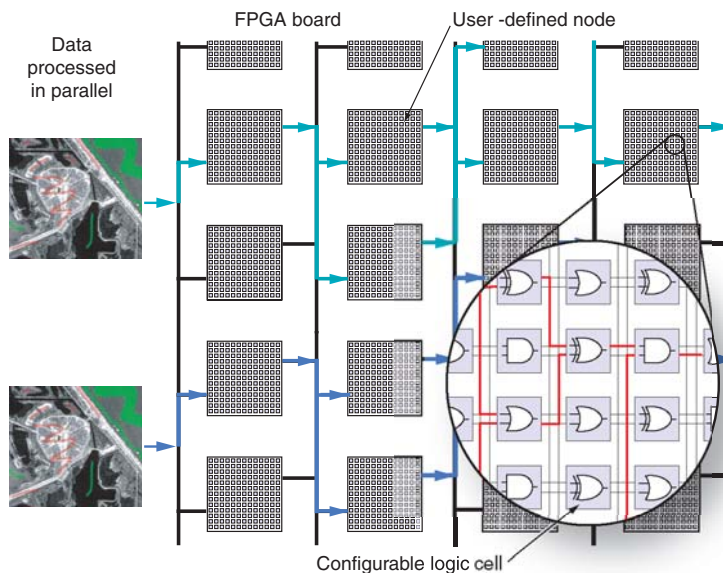
The FPGA consists of cells that implement logical gate functions, such as NAND, NOR, or XOR. Each cell can be configured to perform different logic functions at different times. A programmable matrix connects the cells to each other, and those connections can be altered by signals sent to the FPGA board. Thus, a user can create different logic circuits (nodes). Similarly, the nodes can be linked together to perform all the steps that are needed for the data analysis (see Figure 6). Furthermore, the nodes process large data sets in parallel, greatly reducing analysis time. Once the task is completed, or the search criteria change, the user can reconfigure the FPGA to perform another task.

By adding memory and input/output devices to the FPGAs, we build, in fact, a reconfigurable computer (RCC). One system we have built for an RCC—we called it POOKA—combines genetic programming with reconfigurable hardware and allows us to build a truly optimized analysis algorithm. How much speed can POOKA bring to feature classification tools such as GENIE? A lot! With a



**Figure 5. GENIE**

GENIE is a computer program that develops pattern recognition algorithms from a limited body of analyst-supplied training data. (a) For a water-finding task, an analyst tags pixels of interest (water) in green and undesired pixels (anything but water) in red. GENIE used this initial information to evolve the mask shown in (b), which includes all the water and nonwater in the image. The user is able to influence the evolution of algorithms by providing additional information and by interactively providing additional training data.



**Figure 6. Reconfigurable Computing**

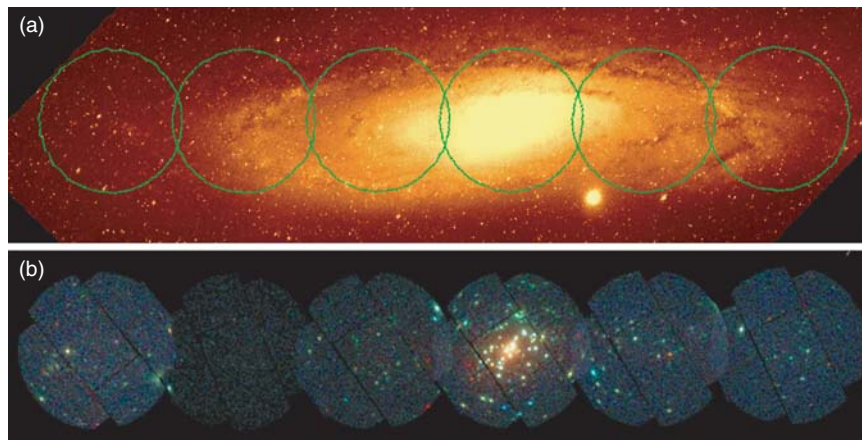
The heart of an RCC is the field-programmable gate array (FPGA) circuit board, whose function can be modified by software. The FPGA consists of millions of system gates, which are the basis for the reconfigurable cells. Each cell can be reconfigured to perform a different low-level logic function, such as AND or OR. Many cells are grouped together into a node that performs a complex function, such as edge detection or spectral filtering. A genetic algorithm reconfigures the collection of nodes to create an optimized analysis algorithm. The advantage of the RCC is that many subtasks can be done in parallel. Information is also pipelined so that new data can start to be analyzed even as old data move through the system. The RCC allows us to do complex analysis tasks much faster and with far less hardware than was previously possible.

small data set, new algorithms can be obtained 100 times faster on POOKA than on a conventional computer. Once the system is trained, the optimized algorithm applied to a new data set runs 20 times faster. POOKA is so fast that we are able to search in real time for features in video data streams, for example, from a surveillance camera on an unmanned aerial vehicle. Thus, we can train the algorithm to recognize not just spatial or spectral features but also features that vary between video frames.

The ability to couple a processor to a sensor and optimize the processor to perform specific tasks has allowed us to do multispectral analysis in real time. This achievement has revolutionized our surveillance capabilities and has also opened up amazing opportunities for basic research. (See the box “Gotcha! You Blinked!” on the opposite page 161.)

### Fundamental Space Science and Astrophysics

In 1973, a Los Alamos team announced that the gamma-ray detectors aboard the fifth and sixth pairs of Vela satellites had detected 16 very intense “bursts” of celestial gamma rays, each lasting about a minute but consisting of a number of quick, sharp pulses. The astounding feature of the bursts was their unbelievable brightness—often brighter than the rest of the gamma-ray universe combined! The discovery of bursts immediately raised two scientific questions: What astrophysical sources could emit such rapid, potent spikes of energy, and where were those sources located? Because the intensity of light falls off inversely as the square of the distance, the questions were related. Cosmic sources located millions, or even billions, of light-years away would have to emit orders of magnitude more energy compared



**Figure 7. X-Ray Map of M31, the Milky Way's Neighbor**

The X-Ray Multimirror Mission Observatory allows seeing the Andromeda Galaxy (M31) in (a) optical light and (b) x-rays. Although the early Vela x-ray instruments could not even detect M31, the observatory is so sensitive that it can resolve 600 x-ray sources within that galaxy. Most of these sources are rare double stars that contain a neutron star or black hole.

with a source located within or near our galaxy.

Theoreticians and experimentalists at Los Alamos were extremely active in trying to shed light on the phenomenon. Collaborating with other institutions, Los Alamos researchers fielded increasingly sensitive gamma-ray detectors aboard the Pioneer Venus Orbiter (launched in 1978), the third International Sun-Earth Explorer spacecraft (ISEE-3, also launched in 1978), and the Ginga spacecraft, which was launched in 1987. But because of their small size, those instruments were insensitive to all but the largest bursts. In addition, the instruments had limited spatial resolution, so data had to be combined with that from other spacecraft to allow accurately locating the burst in the sky. Unfortunately, the initial data analysis often took weeks to complete, far too long to permit follow-up studies by higher-resolution x-ray and optical telescopes. For many years, those deficiencies limited the amount of information available to the gamma-ray burst

community.

Things began to change in 1991, after NASA launched the Compton Gamma-Ray Observatory. The satellite viewed the entire sky with an array of relatively large detectors and recorded hundreds of gamma-ray bursts. The data clearly showed that bursts came from all parts of the sky, without any preference for the plane of the Milky Way or for regions around the Andromeda Galaxy. The likely explanation was that sources were uniformly distributed throughout the universe. That view was solidified by the Italian-built BeppoSax satellite, launched in 1996. Data from the satellite could be analyzed fast enough (within 5 to 8 hours) that ground personnel could direct onboard x-ray instruments to observe the source. BeppoSax was the first to detect an x-ray “afterglow” following a gamma-ray burst. The x-ray data allowed researchers to extract redshifts and hence deduce a distance scale. Most physicists now agree that the bursts come from sources located billions of light-years away.

Scientists are still searching for a

complete picture of how bursts are produced and are relying on data from the latest generation of spacecraft. The HETE-2 (for High-Energy Transient Explorer) satellite, for example, launched in 2000 with Los Alamos instrumentation and software, processes burst data within tens of seconds. A fast trigger on the gamma-ray detectors quickly relays to observers worldwide event information, which elicits fast responses from ground-based robotic telescopes (such as the RAPTOR system discussed in the box “Gotcha! You Blinked!” on this page). Spectral information can be gathered during the crucial first minutes of the event, while the burst is still happening. In December 2003, NASA will launch the Swift satellite. With its enormous burst alert telescope and Los Alamos triggering and imaging software, Swift will have an even greater opportunity to locate and observe hundreds of bursts per year.

Gamma-ray bursts are but one area of fundamental space research that was advanced by Los Alamos instruments. Another is in the field of x-ray astronomy. This work started with a simple x-ray telescope that flew on the Vela satellite. Although modest in size and limited in performance, that telescope proved to be exceedingly useful because it operated for more than 10 years. It allowed us to do long-term studies of x-ray binaries (peculiar double stars containing a black hole or neutron star) and active galactic nuclei (supermassive black holes at the center of galaxies). That telescope was the forebear of the optical-ultraviolet monitor telescope that we helped develop for the giant X-Ray Multimirror Mission Observatory, a satellite launched by the European Space Agency in 1999. The observatory has studied the x-ray source population in the Andromeda Galaxy, the Milky Way’s nearest large neighbor (see Figure 7).

Closer to home, research on the

## Gotcha! You Blinked!

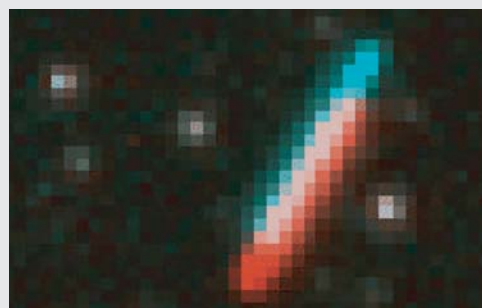
*W. Thomas Vestrand*

We take for granted that the stars in the night sky are stable from night to night and year to year. But also populating the heavens are short-lived optical transients such as the bright optical flash of January 23, 1999, that lasted about 90 seconds and reached an apparent magnitude in brightness of 9. Estimated to have originated at a distance of 10 billion light-years, it was the most luminous optical object ever measured by humankind. Unfortunately, witnessing similar events is frustratingly difficult. The flashes are generally not preceded by other events and are often over by the time we can train a telescope to the right spot.

The solution is to adapt technology that is used to fulfill our threat reduction mission and couple optical sensors to real-time processors. This procedure has allowed us to develop the first of a new generation of “smart” telescopes that can locate, in real time, celestial optical transients that come and go in less than a minute.

Our sky-monitoring system, RAPTOR (for rapid telescopes for optical response), is best understood as an analogue of human vision. The human eye has a wide-field, low-resolution imager (rod cells of the retina), as well as a narrow-field, high-resolution imager (cone cells of the fovea). Both eyes send image information to a powerful real-time processor, the brain, running “software” for the detection of interesting targets. When a target is identified, both eyes are rapidly turned to place the target on the central fovea imager for detailed “follow-up” observations with color vision and higher spatial resolution. Because we have two eyes viewing the same scene, we can eliminate such image faults as “floaters” and extract distance information about objects in the scene. Similarly, RAPTOR employs two primary telescope arrays that are separated by a distance of 38 kilometers to provide stereoscopic imaging (see Figure A). Each telescope array has a wide-field imager and a central, narrow-field fovea imager. Both arrays are coupled to a real-time data analysis system that can identify transients in seconds. Instructions are then relayed to point the high-resolution fovea telescopes at the transient.

The RAPTOR sky-monitoring system, which collected its first data in the summer of 2002, will give astronomers the first unbiased global look at the variability of the night sky on timescales as short as a fraction of a minute. It has already imaged an asteroid approaching the earth (see the figure), which stands out from the stars in its field because of the parallax (position shift) between the images taken by the two telescopes.



**Figure A. RAPTOR**

**This double image of an asteroid approaching the earth was taken by RAPTOR. Two telescopes, 38 km apart, took each image (shown in red and blue, respectively). Unlike the distant stars, the asteroid position shifts from one telescope to the other.**

### Near Space and ENA Imaging

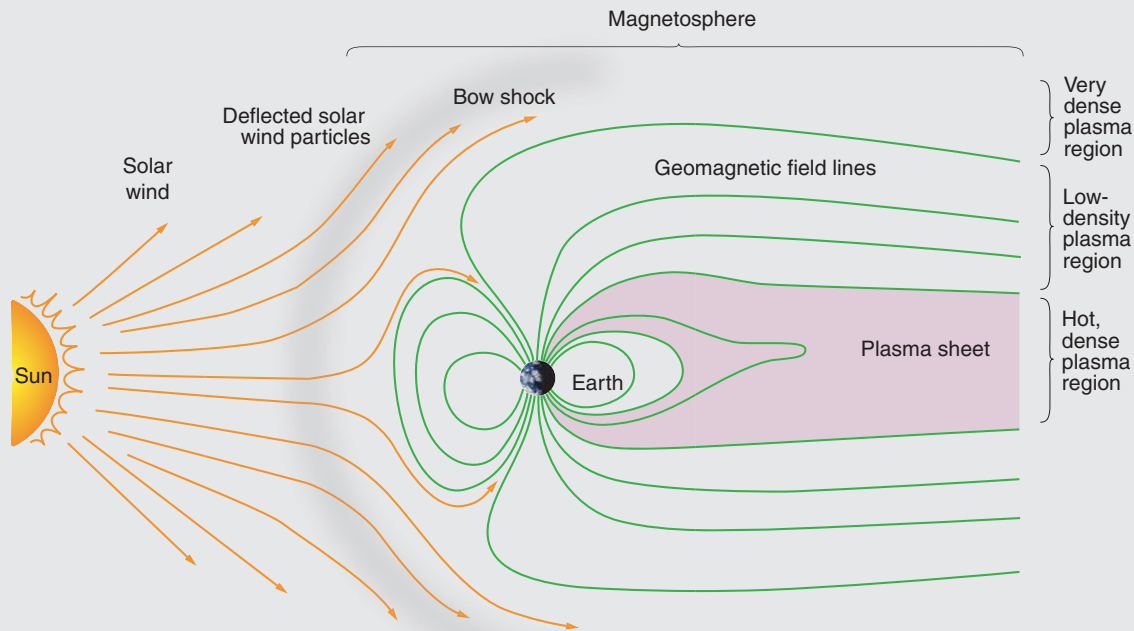
*John T. Gosling and Geoffrey D. Reeves*

Far from being empty, the near-space environment is filled with magnetic fields and solar wind (see Figure A). The latter is a magnetized plasma consisting primarily of protons and electrons that flee the sun’s surface at supersonic speeds. As an ionized gas, the bulk of the solar wind cannot penetrate directly into the earth’s magnetosphere and, therefore, must be diverted around it. Because the solar-wind flow is supersonic, a bow shock stands off upstream of the earth to cause the solar wind to divert around the magnetosphere.

As a result of its interaction with the solar wind, the day side of the earth’s magnetosphere is compressed. Some of those field lines, through the process of magnetic recombination, become interconnected with the magnetic field carried by the

heated solar-wind plasma, along with plasma of ionospheric origin that also resides in the geomagnetic tail, is further energized and accelerated toward the earth, where it collides with and excites particles in the upper atmosphere. The excited particles then emit light that we see as auroras.

Los Alamos pioneered an effort to image and map the earth’s entire magnetosphere at one time, a feat that will revolutionize our understanding of this plasma environment. We proposed our innovative imaging technique—known as energetic neutral-atom (ENA) imaging—nearly 20 years ago, demonstrated the principle in the 1990s, and have begun to demonstrate its full potential in the last two years. ENA imaging relies on the exchange of an electron between an



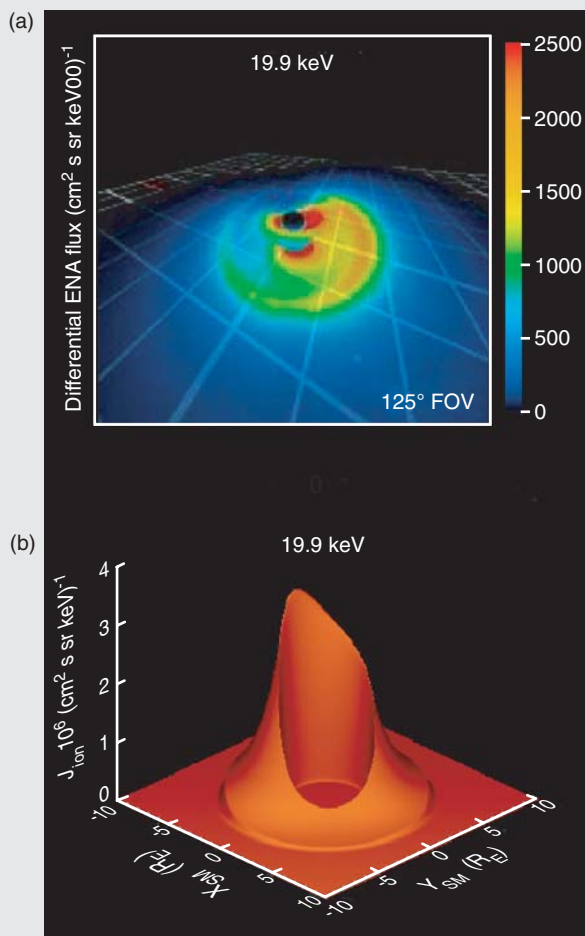
**Figure A. Magnetic Fields and Solar Wind in the Near-Space Environment**

The solar wind flows at supersonic speed and is deflected around the earth’s magnetosphere by a bow shock. The solar wind compresses the day side of the magnetosphere. Field lines from the earth’s day side recombine with the magnetic field and form a geomagnetic tail on the earth’s dark side. The tail encloses a plasma sheet of hot, high-density solar-wind plasma.

solar wind and are carried far past the earth. The result is a long geomagnetic tail on the earth’s dark side. Far downstream, the magnetic interconnection with the solar wind is broken, and field lines can return to the earth. This enclosed area within the geomagnetic tail is called the plasma sheet, and it holds a relatively high density of captured and heated solar-wind plasma. During geomagnetic disturbances, this

energetic ion and a cold neutral atom. Neutral atoms in space are extremely rare, and they seldom collide with ions. But when they do, the ion gives up its charge and breaks free from the confines of the planetary or interplanetary magnetic fields. Except for the weak effects of gravity, the neutral atom travels in a straight line and can be imaged by a detector to “take a picture” of the distant plasma.





### Figure B. ENA Imaging

(a) How would the magnetosphere look if seen from a satellite 53,000 km above the earth?

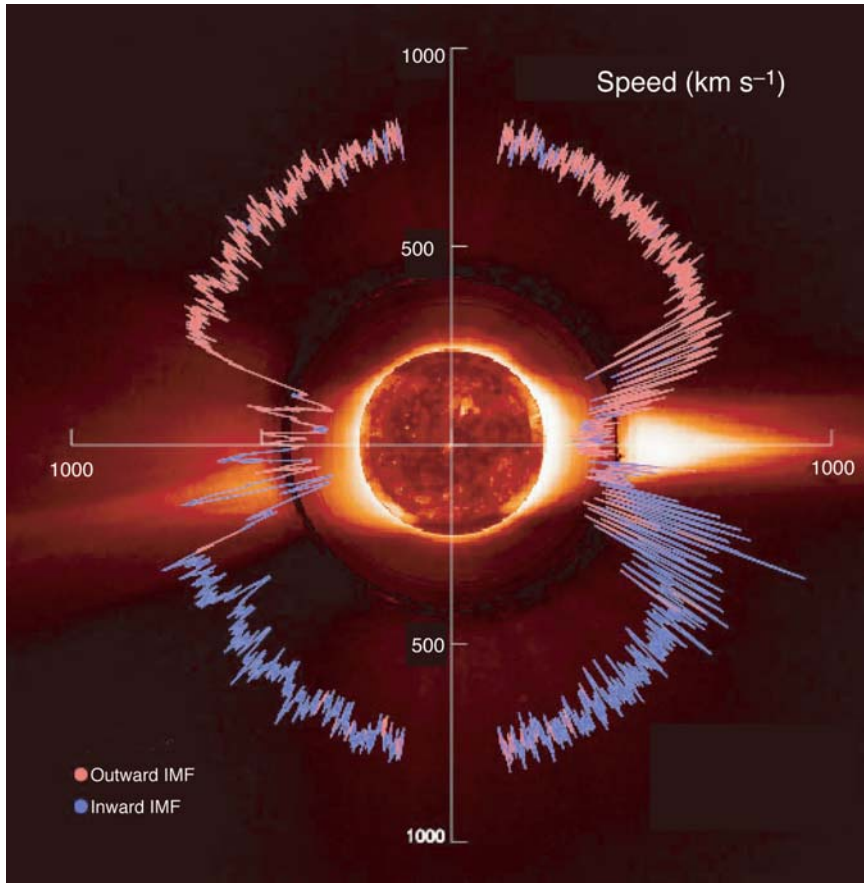
The earth is seen at the center, and the northern polar cap, where ENAs are not produced, is seen as the black, oval shape in the figure. An asymmetric ion ring, the ring current, produces bright emissions around the north and south polar caps as it sweeps around the earth. Ions in the low-altitude “horns” of the magnetic field lines interact with the dense portion of the atmosphere. (b) Viewed from one direction, the ENA emissions are an integral of a 3-D ion distribution convolved with a 3-D atmospheric distribution. A model of the inferred equatorial distribution is shown here.

We produced the first dynamic images of the earth’s magnetosphere in the 1990s, using a satellite-based instrument that was originally designed to measure charged particles. In 2000, NASA’s IMAGE (for imager for magnetopause-to-aurora global exploration) mission was launched with three types of ENA imagers (for high-, medium-, and low-energy atoms), each fully optimized to image the magnetosphere. Figure B(a) shows what the magnetosphere would look like if you were on a satellite 53,000 kilometers above the earth. Models of the magnetic field and the distribution of atmospheric neutrals can be used with these images to determine the distribution of ions as a function of radius and local time. One such distribution is shown in Figure B(b).

The upcoming TWINS (for two wide-angle imaging neutral-atom spectrometers) mission, which will be launched in 2003 and 2005, will have two medium-energy ENA instruments on separate satellites. The stereoscopic data will allow us to create the first 3-D images of the magnetosphere. The data will also help advance our understanding of “space weather,” or the variations in the plasma environment that can adversely affect, among other things, satellite communications and operations, radio and television transmission, the power network, and even the safety of our astronauts in space.

near-space environment has been extremely productive and has led to a number of fundamental discoveries about the sun’s extended atmosphere, the solar wind, and the interaction of that atmosphere with the earth’s magnetic field. Measurements by instruments on the Vela satellites revealed some of the complexity of the earth’s magnetosphere and led directly to our discovery of the earth’s plasma sheet, a region of concentrated plasma that extends far downstream on the night side of Earth (see the box to the left). Other measurements by Los Alamos instruments on Vela led to the discovery that the sun often impulsively ejects into interplanetary space large amounts of material, which have come to be known as “coronal mass ejections.” Los Alamos work in the early 1990s revealed that these ejections, and not solar flares, are the prime cause of major solar-wind disturbances and large geomagnetic storms.

The considerable success of the early Vela measurements prompted NASA to use Los Alamos plasma sensors on a series of satellites launched in the early 1970s. That was the beginning of a long and fruitful collaboration between our two institutions to study the near-space environment, a collaboration that continues to the present. Our instruments have sampled all the different regions of the earth’s magnetosphere and have explored the solar wind in considerable detail. Figure 8 shows the solar-wind speed as a function of solar latitude. The data were obtained by Los Alamos instrumentation on the Ulysses spacecraft, a joint endeavor between NASA and the European Space Agency. Ulysses was launched toward Jupiter from the space shuttle Discovery in October 1990. The giant planet’s gravitational field deflected the craft out of the ecliptic and into a 5.5-year-long orbit over the poles of the sun. It is the first-ever polar orbit of the sun by a manmade object.



**Figure 8. Solar-Wind Speed as a Function of Solar Latitude**

A polar plot of the solar-wind speed as a function of solar latitude was measured by Los Alamos plasma detectors on Ulysses. The speed trace is color-coded according to the observed polarity of the interplanetary magnetic field (IMF) that threads the heliosphere. Underlying the speed trace is a set of concentric images of the solar corona, the source of the solar wind, obtained from a combination of space and ground-based telescopes. A striking aspect of the plot is the high and nearly constant speed of the solar wind, outward in the northern hemisphere and inward in the southern hemisphere, observed at high heliographic latitudes throughout the orbit. This high-speed wind originates from relatively dark regions in the solar atmosphere known as coronal holes. Low-speed wind originates in the bright coronal streamers prevalent at low solar latitudes at this phase of the solar cycle. The alternating flows at low latitudes reflect the fact that the solar magnetic dipole had a sizable ( $20^\circ$ -to- $30^\circ$ ) tilt relative to the solar rotation axis during the interval shown, and as the sun rotated with a periodicity of 25 days, high- and low-speed flows were directed toward Ulysses at regular intervals.

Ulysses is now nearing completion of its second trip over the sun, during which time the 11-year solar-activity cycle rose and peaked. As a pointed reminder of the variability of our local environment, the relatively organized nature of the solar wind measured during the first orbit (refer to Figure 8)

was noticeably more complex on this second pass. This change was due to the considerably more complex nature of the sun's magnetic field and corona and the increased number of solar-wind disturbances produced by solar activity at this time.

Near-space research has con-

tributed substantially to a fundamental understanding of, among other things, magnetic reconnection, collisionless-shock formation, and charged-particle acceleration and transport. These phenomena, in turn, have helped us construct models of basic astrophysical plasma processes. Magnetic reconnection, for example, is a restructuring of a plasma's magnetic field, in which field lines oriented in opposite directions break and reconnect to each other with a subsequent release of stored magnetic energy. Magnetic recombination, which has been evoked as the likely power source for the acceleration of charged particles into space during solar activity, is also believed to power the relativistic jets of matter that shoot out from quasars. Although magnetic reconnection cannot be studied directly from a quasar located billions of light-years away, our own near-space environment provides us with a remarkable laboratory to study the phenomenon.

## Epilogue

As the Laboratory celebrates 60 years of serving society, it also celebrates 40 years in space. In those 40 years, we have strengthened the national security with sensor and processing systems and used the same capabilities to explore our world from the earth outward to the early universe. With each generation of nuclear detection sensors, we do more with less, driving our systems into progressively smaller but more capable packages, thanks to advances in onboard event detection, device miniaturization, and background processing. The curve of performance shows no sign of turning over. As a result, we are confident of many more discoveries in the decades to come. ■

For more information, please visit <http://www.lanl.gov/orgs/nis/>.

*For further information, contact  
William Friedhorsky (505) 667-5204  
([wpriedhorsky@lanl.gov](mailto:wpriedhorsky@lanl.gov)).*

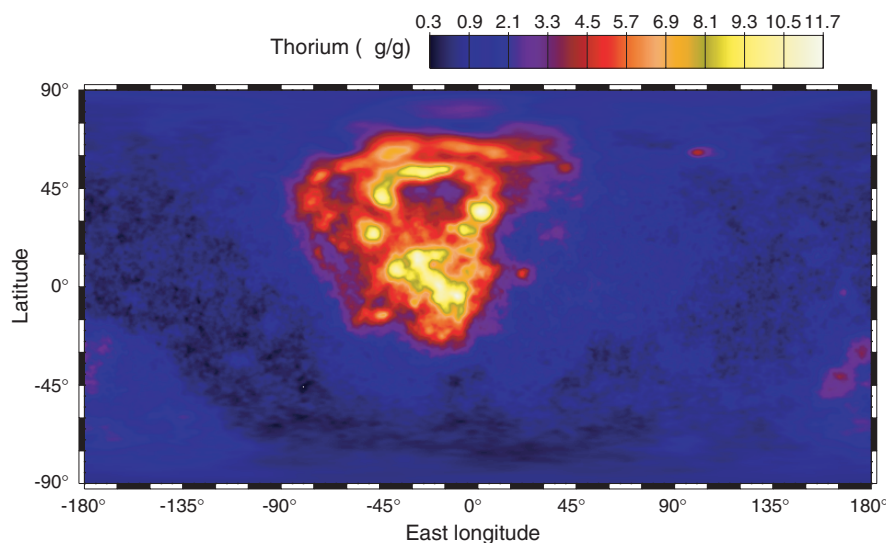
# Geochemical Studies of the Moon and Planets

*William C. Feldman*

**A**n outgrowth of the space-based Treaty Verification and Proliferation Detection Programs at Los Alamos was the fielding of experiments to determine the elemental composition of the moon and Mars. Our instruments, which included neutron, gamma-ray, and alpha-particle spectrometers, flew aboard Lunar Prospector and Mars Odyssey missions sponsored by NASA and were tailored into a small package to meet the scientific objectives of these missions. All four Los Alamos experiments aboard the Lunar Prospector mission were extremely successful. Measurements from the neutron and gamma-ray spectrometers enabled the first global mapping of the seven most abundant rock-forming elements (oxygen, magnesium, aluminum, silicon, calcium, titanium, and iron); the trace radioactive

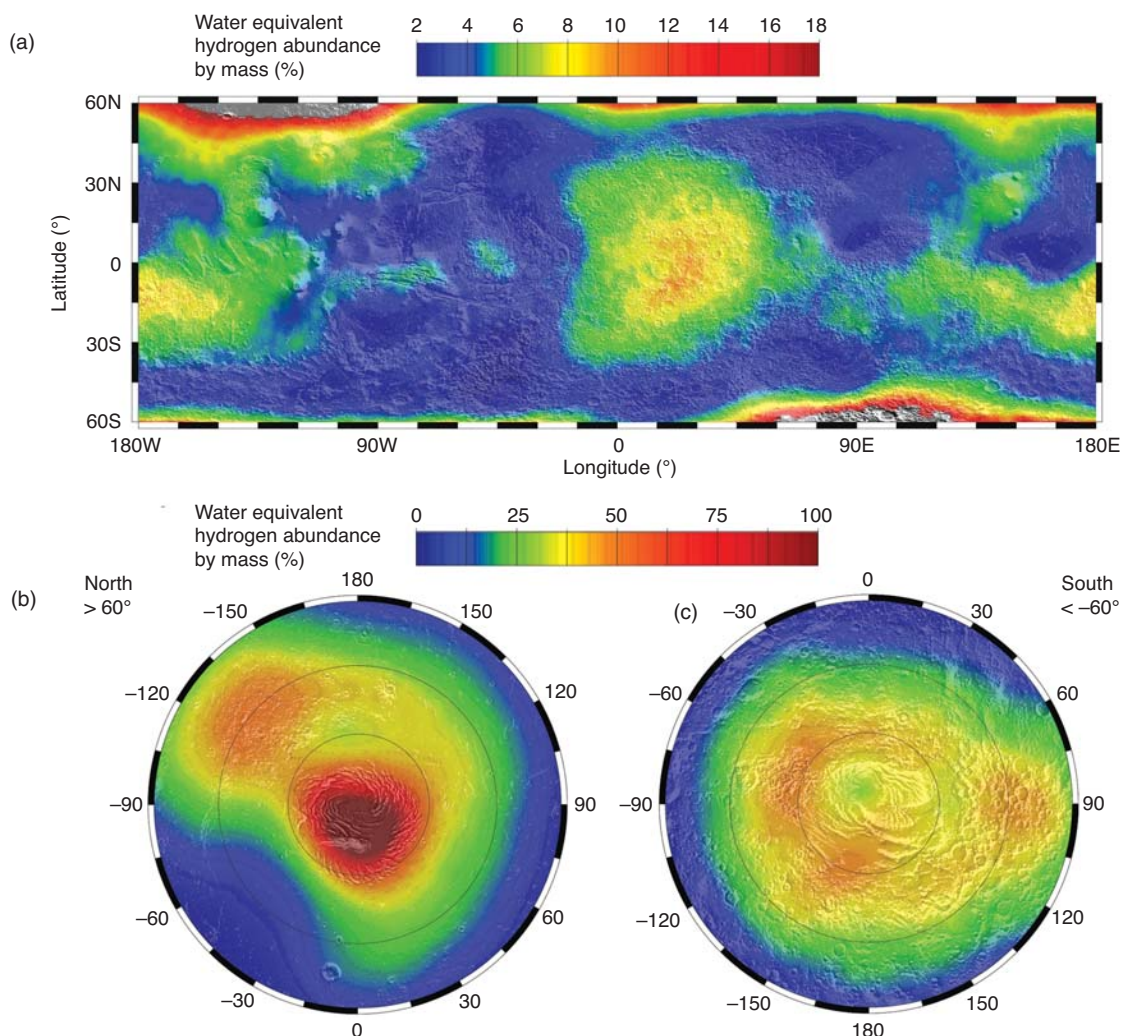
elements thorium (refer to Figure A for thorium's spatial distribution on the moon), uranium, and potassium; and the minor elements hydrogen, gadolinium, and samarium. This combined package of instrumentation marked the first application of neutron spectroscopy to planetary exploration. Several highlights emerged from our data collecting. We learned that a unique, thorium-rich geochemical province exists on the earth-facing side of the moon. Heat from the decay of radioactive elements is no doubt responsible for the iron-rich basalt flows that give the moon its black-splotted appearance. In another landmark success of these experiments, we discovered water-ice deposits that reside within the permanently shadowed craters near both lunar poles. Asteroids, comets, and interplanetary dust grains striking the surface were most likely

responsible for delivering the water to the moon. In addition, using the alpha-particle spectrometer, we detected radon and its daughters that escape from vents in the lunar crust. Our experimental results help unravel the origin and evolution of the moon and show the existence and locations of lunar resources that could be used to support the manned exploration of the moon. First returns from the Los Alamos neutron spectrometer aboard Mars Odyssey have proved the power of neutron spectroscopy to map the volatile inventory of nearly airless planetary bodies. We have discovered that a vast region of Mars south of  $-60^\circ$  latitude is rich in buried water ice. Figure B shows lower-bound estimates of water on Mars. Although water ice has been predicted to be stable at these cold Martian latitudes, we were surprised by the extent and richness of this deposit (up to 50 percent water ice by mass). Initial data were taken during the southern late summer and early fall, when the southern cap is smallest; our analysis clearly shows that the residual south polar cap of Mars is permanently covered by dry ice—or frozen carbon dioxide ( $\text{CO}_2$ ). A much more extensive dry-ice cap covers the northern region, poleward of about  $55^\circ$ , which was in the grip of late winter and early spring. As time went on and the sun returned to the north, the northern cap shrank and the southern cap grew, as  $\text{CO}_2$  evaporated and precipitated, respectively. This exchange of  $\text{CO}_2$  between the Martian poles is a major driving force for Mars' atmospheric circulation, and a factor in the long-term climate variability on Mars (which is also driven by the changing obliquity, eccentricity, and perihelion of the Martian orbit). The shrinking north polar dry-ice cap reveals a basement as



**Figure A. Global Map of Thorium Abundance on the Moon**

The spatial distribution of thorium looks asymmetric, being strongly concentrated in a single province on the earth-facing side of the moon. This province witnessed much of the volcanism that has distributed large quantities of basalt filling many of the large-impact basins, which are also concentrated on the earth-facing side of the Moon.



### Figure B. The Distribution of Water on Mars

Data taken between February 2002 and April 2003 by the neutron spectrometer aboard the Mars Odyssey were used to determine the minimum water distribution on Mars. Measurements of the epithermal neutron flux streaming from the planet's surface allow us to determine the amount of hydrogen trapped in the upper one meter of soil. We then assume that the hydrogen is most likely in the form of water ice and uniformly distributed throughout the soil layer, and convert the data to water mass equivalent. (a) This cylindrical projection map of the midlatitudes shows the water equivalent hydrogen abundances by mass overlaid on a shaded relief map of the planet's surface. The relief map was derived from tomography data taken by the Mars Orbiter laser altimeter (MOLA), an instrument designed at the Goddard Space Flight Center. The data are presented in stereographic projection in (b) and (c), in which the latitudes poleward of  $+60^\circ$  and  $-60^\circ$  are shown, respectively.

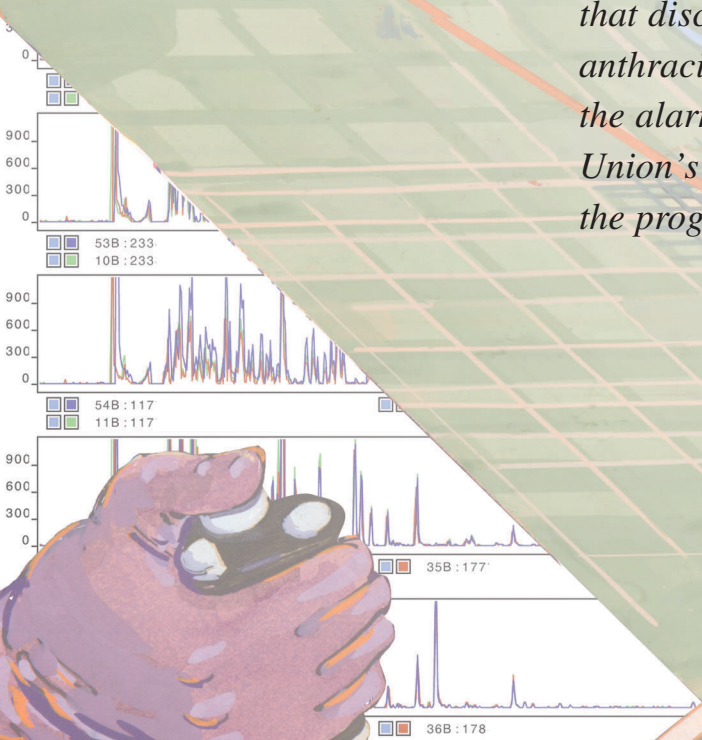
rich in water ice as that in the southern cap. Future work for our group includes the development of a neutron spectrometer, to be launched to Mercury in 2004, and a combined neutron and gamma-ray spectrometer, which is scheduled for launch in 2006 for a rendezvous with the asteroids Vesta and Ceres. ■

# Reducing the Biological Threat

*Detection, characterization, and response*

*Paul J. Jackson and Jill Trehwella*

*Los Alamos has worked for over a decade to develop DNA-analysis tools that can distinguish one pathogen from another. That effort has already paid off. Assays that discriminate between individual strains of *Bacillus anthracis*, the pathogen that causes anthrax, revealed the alarming sophistication of the former Soviet Union's bioweapons program and exposed elements of the program in Iraq.*



Los Alamos  
NATIONAL LABORATORY

From our beginnings, disease-causing microbes have taken their toll on human populations, sometimes in devastating numbers. In the fourteenth century, the Black Plague killed 25 million people, or one-third of the European population, and the 1918 Spanish Flu pandemic resulted in 30 to 40 million deaths worldwide. As man evolved and looked to defending or expanding his territory, it did not escape his attention that the agents that cause disease could be used effectively as weapons. As early as 1346, during the siege of Kaffa, the Tartar army hurled corpses of plague victims over the walls of the city, an action that led to the eventual surrender of Kaffa's inhabitants. In the spring of 1763, during the French and Indian War, blankets laden with the remnants of smallpox sores from infected British troops were collected and given to the Indians allied with the French while they attended a conference at Fort Pitt. Thus, a devastating smallpox outbreak was unleashed in the previously unexposed Native American population.

In the nineteenth century, Louis Pasteur, Robert Koch, and others established the relationship between microbial pathogens and disease. From that point in history, efforts were made to isolate and propagate cultures of the different pathogens both for research into ways to understand them and protect human health and for use of those pathogens as agents of war. During World War I, German saboteurs in France infected horses and mules with *Bacillus anthracis* and *Burkholderia mallei*, the microbes that cause anthrax and glanders, respectively. Beginning in the 1930s and continuing through World War II, the infamous Imperial Japanese Army Unit 731 experimented with biological warfare in Manchuria. Those experiments, some of which were gruesome, resulted in the deaths of thousands of Chinese nationals. It is now well established that the former Soviet Union supported one of the largest and most

sophisticated biological weapons efforts, whose legacy in terms of control and accountability of materials and expertise is a concern today. In the 1990s, the United Nations Special Commission, or UNSCOM, inspectors in Iraq eventually forced Saddam Hussein to acknowledge an active and diverse biological warfare program that included having agents loaded in munitions ready for delivery.

In September 1984, the prospect of non-state-sponsored terrorism gained attention as the Rashneeshee cult contaminated salad bars in ten restaurants in The Dalles, Oregon, by pouring vials of liquid *Salmonella typhimurium* culture over the foods. This contamination caused an estimated 751 cases of salmonella poisoning and is believed to be an attempt at influencing the outcome of the November elections. Most vivid in our memories, however, is the impact of five deaths and the infections resulting from the letters laced with *B. anthracis* spores and mailed to people in the media and Congress soon after the September 11 terrorist attack on the World Trade Center. Although quick medical responses contained the number of deaths, the closing down of the Senate Hart Building, the disruption and loss of confidence in the safety of the U.S. mail, the cost of cleanup, and treatment of tens of thousands of potentially exposed people raised our awareness of the potential impact of individual or state-sponsored bioterrorism. Los Alamos scientists became actively involved in the nation's response to the anthrax mail attacks because of expertise we started to develop more than twelve years ago.

### The Challenge of Biothreat Reduction

As early as 1991, scientists at Los Alamos began developing DNA-based methods for detecting and characterizing those biological agents that can be used as weapons. Our motivation was twofold. In addition to possibly identify-

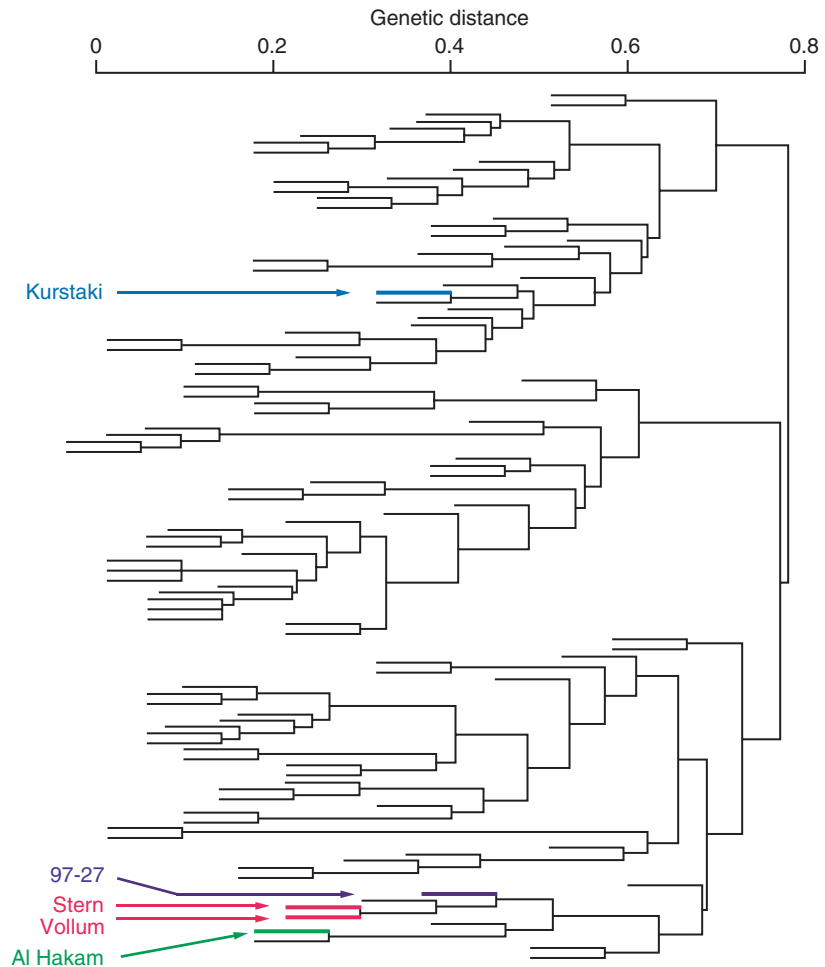
ing the presence of a threat agent and how we might deal with it, we needed to know whether the pathogen had been deliberately released, and if so, where it might have come from. Attributing an outbreak to its source could provide information to help mitigate the spread of disease in the population. In the realm of threat reduction, "attribution" can also deter individuals or nation states from using biological weapons for fear of being caught and having to bear the consequences.

Los Alamos was a logical place for federal agencies to come and ask for these capabilities. Almost from its inception, the Laboratory has had a significant bioscience program. At that time, our efforts were directed toward understanding the health effects of radiation. By the 1980s, this long-standing effort was intensely focused on molecular and cell biology and on exploring the hypothesis that individual susceptibility to radiation was programmed into our DNA. In particular, Charles DeLisi of the Department of Energy was among those arguing that the basis for this genetic susceptibility could be discovered if there were a complete human genome reference sequence. The ambitious idea to sequence the human genome later evolved into the multinational Human Genome Project, which led to unprecedented capabilities for genomic analysis. But this capability could be applied equally to microbial genomes, and as a result, Los Alamos entered the genomic era with two complementary missions: advancing the Human Genome Project and developing methods to detect and identify pathogens in environmental and laboratory samples.

Pathogen detection is challenging, however, because there are few genetic differences that distinguish a pathogen from a closely related nonpathogenic organism. It is not enough to detect the genes that make the organism threatening—the so-called pathogenicity or virulence genes—because these are often found in nonpathogenic microbes and

**Figure 1. Genetic Variability and Phylogenetic Trees**

A phylogenetic tree depicts the evolutionary relationship between different species. This is a partial tree, showing some of the pathogenic subgroup I of the bacillus family. Each branching point results from one or more mutations that generate a new, genetically distinct organism. Species that are near each other on the tree are closely related. “Closeness” is quantified by genetic distance, which can be estimated by summing the horizontal lengths of the branches that go from one species to another. (Vertical lengths carry no information.) The red branches are the *B. anthracis* region of the tree. (The blue, purple, and green branches are species discussed later in the text.) All *B. anthracis* species and strains are closely related and show limited genetic variability. On this tree, they fall within the two branches labeled Stern and Vollum. *B. thuriangiensis*, however, exhibits high genetic variability and is difficult to identify. Many bacteria on this tree have been called strains of *B. thuriangiensis* because of phenotypic (nongenetic) properties. Interpreting the results of a test without a thorough understanding of the tree is difficult even with DNA-based detection assays.



sometimes in microbes that are not even closely related to the target pathogen. Furthermore, given the aforementioned genetic mutability of microbes, there may be numerous “strains” of the same organism, some of which may be pathogenic and others not.

One must therefore know something about the pathogen’s genetic diversity as such knowledge affects the detection results. Unfortunately, we still do not completely understand the diversity of all the threat agents. We know, for example, that a species such as *Burk. pseudomallei* (the causative agent of melioidosis, an infectious disease similar to glanders) is very variable. If *Burk. pseudomallei* must be detected and characterized, reagents that detect one strain are not likely to detect the others. We also

know that *B. anthracis* exhibits very limited genetic variability, so reagents that detect one strain will virtually always detect the others. But there is less than 0.3 percent DNA sequence difference between *B. anthracis* and one of its closest nonpathogenic relatives. This common near neighbor is often found in environmental and sometimes even medical samples. If only the near neighbor is in the sample but the detection reagents cannot distinguish between it and *B. anthracis*, the sample will give a false positive.

The strategy to detect and identify pathogens must therefore incorporate several steps. One must first identify species-specific markers that are present in all strains of a pathogen but are not present in its close relatives. One then

needs to develop a comprehensive phylogenetic tree (refer to Figure 1) that maps out the relationships among the pathogen strains and closely related species. Then one can begin to build a capacity for attributing a microbe to a particular source by overlaying information about the global distribution of species and strains on this tree and including isolates from known biological weapons activities.

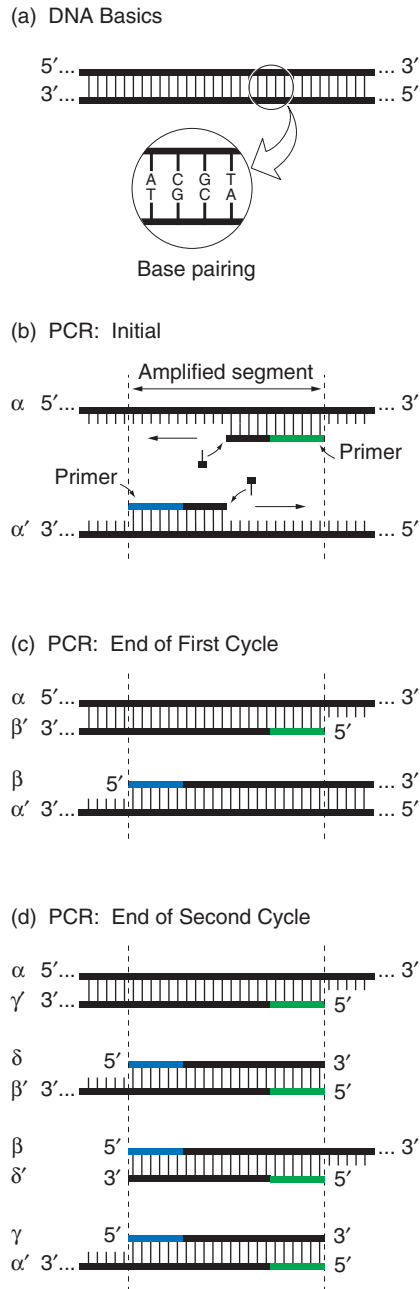
Before addressing the science and technology challenges inherent in DNA-based detection and identification, we will note that antibody-based assays provide an alternative means of detection. Because they are easy to implement, such assays are commonly used in the field by first responders. Typically, a solution containing the



**Figure 2. DNA Basics and PCR**

(a) Double-stranded DNA is a sequence of nucleic acid bases (represented by the letters A, C, G, and T) attached to a sugar-based backbone. The bases always pair up: A to T and C to G. Pairing means that each strand can act as a template to replicate the other strand. Because of its molecular structure, the backbone runs in a particular direction, designated as 5' → 3', and the two strands run in opposite directions. (b) PCR is a technique for producing large quantities of specific DNA segments. Double-stranded DNA is placed in a vial, together with short pieces of single-stranded DNA (primers), an ample supply of DNA bases, and enzymes known as polymerases. The PCR cycle starts when the vial is heated to allow the DNA to split into its constituent strands (labeled  $\alpha$  and  $\alpha'$ ). The temperature is lowered. Because of base pairing, the primers bind to specific sites that flank the segment to be amplified. The primers provide a starting point for the polymerase (not shown), which moves along each strand in the 5' → 3' direction, affixing the proper base to the growing DNA. (c) Both strands get replicated and form two double strands. (d) A second PCR cycle starts when the vial is reheated. The DNA splits into four single strands ( $\alpha$ ,  $\alpha'$ ,  $\beta$ , and  $\beta'$ ), which are all templates for replication. At the end of the second cycle, there are four double strands, and two of the eight single strands ( $\delta$ , and  $\delta'$ ) are the exact DNA segment of interest. After three cycles, there are 16 strands and eight copies of the desired segment. Thereafter, the desired segment amplifies exponentially and after about 30 cycles completely dominates the product.

sample is placed on an antibody-laden matrix (paper, plastic membrane, and others). In principle, the antibodies will lock only onto proteins that are associated with a specific pathogen, and that binding will trigger some detectable event, such as a change in the color of the matrix. But to date, antibody-based assays are neither specific enough nor sufficiently reliable to form the basis



of a detection strategy. (This lack of reliability caused the government last year to recommend that HAZMAT teams stop using antibody tickets for *B. anthracis* testing.) Los Alamos is developing a new assay that has many of the desired properties of antibody detection but with improved reliability and specificity (see the article "Fluorobodies" on page 178).

**Science and Technology Challenges**

Our choice of DNA-based methods for pathogen detection stems from the relative stability of DNA in the environment. The DNA molecule contains coded information that can potentially be linked to a specific pathogen, and we can extract that information from samples even when the DNA is badly degraded, that is, broken into pieces and partly destroyed.

An initial concern was whether our methods would require more DNA than was available in the samples. Fortunately, at the time we were focusing on this problem, a newly developed method, the polymerase chain reaction (PCR), was showing great promise for amplifying specific DNA sequences from samples containing millions of other microbial species (see Figure 2). Our first breakthrough was demonstrating that PCR-based methods could indeed perform this feat and amplify selected portions of pathogenic DNA sequences from very complex mixtures.

In principle, genomic sequencing can be used to uniquely identify every pathogen. But currently, high-resolution sequencing of even a small genome is expensive and time-consuming. Furthermore, so many microbes share so many different DNA sequences that most of the information is not very useful for species identification. For an unknown pathogen, finding the unique stretches of DNA is very difficult. A fast and efficient method of characterizing DNA sequences from a large number of different strains and species is to cut the total DNA from each microbe into fragments and look at the pattern of fragment lengths that is generated for each isolate.

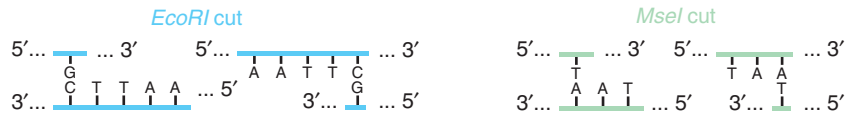
We use restriction enzymes to cut DNA. These proteins cut the DNA double helix in two whenever they happen upon a specific, short sequence of typi-

cally 4 to 6 bases. On average, enzymes that recognize a sequence of 6 bases cut every 4100 bases, whereas those that recognize 4 bases cut every 250 bases. If one starts with a single DNA molecule of several million bases, then digestion with a single restriction enzyme will generate several thousand DNA fragments whose lengths are defined by the locations of the enzyme recognition sites within the original DNA molecule. Because of differences in their respective genomes, each microbial species or strain will have a different distribution of fragment lengths.

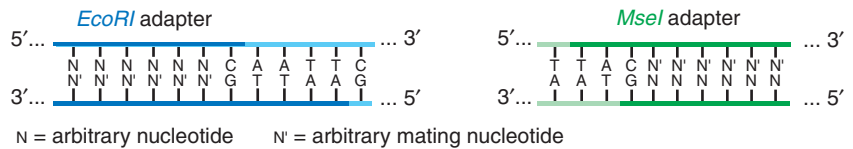
We create a “fingerprint” of a microbe or pathogen’s genome by using an electric field to push the fragments through a slab of jellylike substance called a sizing gel. Because smaller fragments move through the gel faster than longer ones, the fragments separate over time. The emerging pattern corresponds to the original distribution of fragment lengths. This pattern is called a fingerprint and can be used to identify the microbe if a matching or similar fingerprint exists in an archive. Numerous methods for pathogen identification and strain discrimination depend upon this kind of DNA fingerprinting.

At Los Alamos, we have had considerable success developing and applying the fingerprinting approach known as amplified fragment length polymorphism (AFLP) (Jackson et al. 1999). AFLP is a way of culling from the large pool of DNA fragments a very small subset of fragments that have optimal lengths for sizing (see Figure 3). The technique uses a battery of reagents that are not specific to a single microbial species and can therefore identify pathogens even when one does not know what might be present in a sample. (The sample, however, must contain only a single species; otherwise, the result is an unidentifiable mishmash of fragments.) Furthermore, entirely new microbes (not previously observed)

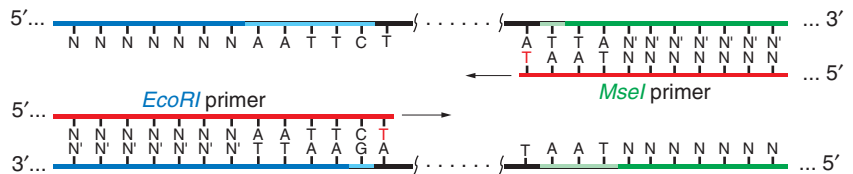
(a) Restriction Enzyme Digestion



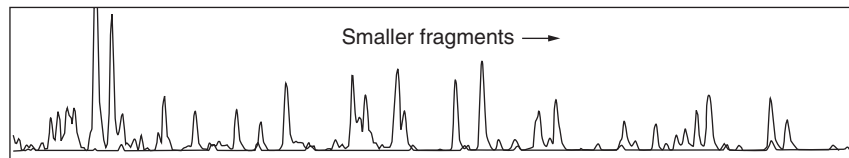
(b) Adapter Ligation



(c) PCR Amplification

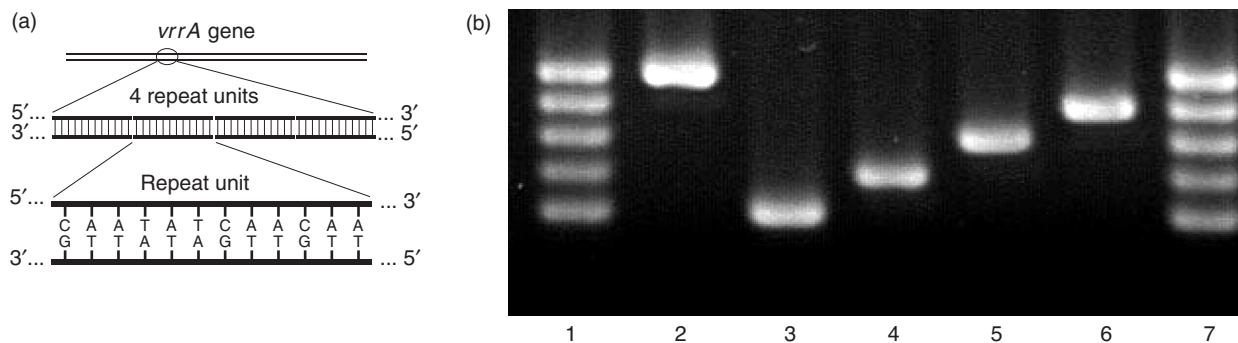


(d) AFLP Profile



**Figure 3. Identifying Pathogens by AFLP**

(a) AFLP is a DNA fingerprinting technique that uses two restriction enzymes (for example, *EcoRI* and *MseI*) to cut a pathogen’s genome into thousands of fragments. Each enzyme recognizes a short sequence of paired bases and cuts the double-stranded DNA at that site. The 5’ end of each cut strand is left with a few unpaired bases. (b) A set of double-stranded “adapters,” which mate to the unpaired bases, are added and then “glued” (ligated) to each fragment. (c) A special set of PCR primers are added: one matching the *EcoRI* and adapter site and the other, the *MseI* and adapter site. The primers also have one or several extra bases (shown in red). Only the subset of fragments that have an *EcoRI* end, an *MseI* end, and the correct extra base (or bases) will get amplified. Enzymes, adapters, and the number of extra bases are chosen in such a way that PCR produces 100 to 200 fragments. (d) The fragments are sized on a capillary electrophoresis system (similar in function to the sizing gel discussed in the text). Smaller fragments travel through the capillary faster, so a detector focused on one spot of the capillary detects progressively larger fragments. A trace of the detector signal vs time consists of a series of peaks, and each peak corresponds to a different fragment size. The height of the peak is related to how much of a particular DNA fragment is present. To identify the pathogen, we compare this genetic “profile” with others from a database.



**Figure 4. VNTR Strain Analysis**

(a) The *vrrA* gene in *B. anthracis* contains a 12-base-long “unit” of DNA that will repeat, back to back, from 2 to 6 times. Different *B. anthracis* strains are associated with the number of repeats. (b) We use PCR to amplify the repeat region of the gene. The resulting DNA fragment will have one of five different lengths. We typically run fragments on a sizing gel to determine their size. Columns 2 through 6 in this gel show five different strains. Each band in the vertical direction corresponds to a fragment with an additional repeat unit. Columns 1 and 7 contain DNA fragments that were created to calibrate the size.

can be located on a phylogenetic tree and placed into a genetic context. For example, if we analyze an unknown microbe and 90 percent of its fragments match the AFLP fragments generated by *B. anthracis*, then this unknown is very closely related to *B. anthracis*. (In fact, the tree shown in Figure 1 was constructed with information obtained from AFLP profiles.) AFLP can also be used for strain discrimination, and was the first method to identify strain-variable DNA target sequences in *B. anthracis* (Keim et al. 1997).

A detailed analysis of AFLP profiles from different *B. anthracis* isolates showed that, although most DNA fragments in the profile were identical across all isolates, a minor set of fragments showed variations. These fragments contained what we call a variable number tandem repeat (VNTR) (Jackson et al. 1997). The VNTRs are relatively fast mutating loci in the genome that contain a variable number of short sequence repeats. For example, the VNTR found within the *vrrA* gene of *B. anthracis* can have five different lengths, depending on how many times a 12-nucleotide repeat is present in a particular strain (see Figure 4). These different lengths allow us to place all known

*B. anthracis* isolates into one of five groups. (A group in this context contains those isolates that have the same number of repeats in the *vrrA* gene and can include different strains that are distinguished by variations in other parts of the genome.)

Methods focusing on VNTRs, such as multiple-locus VNTR analysis (MLVA), have been further developed and very successfully applied to recent real-world samples. Typically, pathogens have numerous locations in their genomes that harbor repeating units of DNA, and each pathogen has a unique set of loci. Together, the VNTRs from this set constitute a unique microbial signature that can be used for precise strain identification. For example, analysis of *B. anthracis* AFLP profiles from many different isolates identified eight VNTRs (Keim et al. 2000). These eight genetic markers divided all known *B. anthracis* isolates into 89 different groups.<sup>1</sup> Similar VNTRs are being identified for *Yersinia pestis*

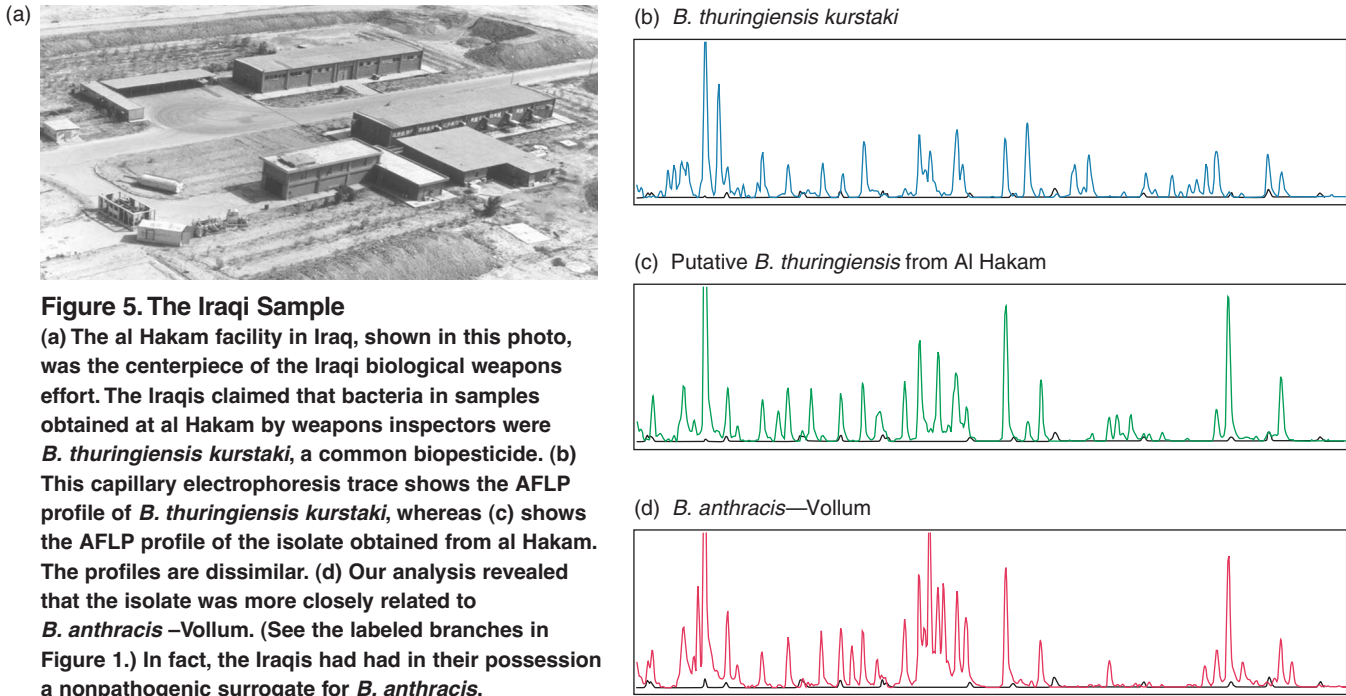
<sup>1</sup>A survey of the recently completed *B. anthracis* whole genomic sequence identified 28 more VNTR loci (Paul Keim, personal communication), increasing the number of different possible categories significantly. However, to date, all *B. anthracis* isolates fall into one of more than 100 categories.

(Klevytska et al. 2001) other pathogens (Farlow et al. 2001).

Providing one knows what pathogen to look for, MLVA has some advantages over AFLP analysis. Because we design PCR primers to amplify only the VNTR-containing regions of a targeted pathogen, purified DNA from a single microbial species is not required for the analysis. However, in the absence of information about the target species present in the sample, MLVA analysis has limited value.

Recent advances in sequencing technology and methods for automation have presented us with another approach to MLVA that uses single nucleotide polymorphisms (SNPs) in genetic sequences. These are loci in the genome that show variations in sequence between strains involving only a single nucleotide substitution. In *B. anthracis*, such changes are rare and provide far less resolution than MLVA analysis. However, in more highly variable species, SNPs may provide as much or more resolution to allow distinguishing among strains. Where there is significant genetic overlap between closely related species, SNPs may provide the only means of differentiating between such species.

Largely on the basis of extensive AFLP and MLVA archives and analy-



ses, *B. anthracis* has become one of the best characterized pathogens in terms of its genetic diversity and its relationship to its close relatives in the bacillus family (Jackson et al. 1997, Keim et al. 2000). The results of this mapping demonstrate that the genomic sequences of the closest *B. anthracis* relatives differ from the true pathogen by less than 1 percent, making identification of *B. anthracis*-specific DNA signatures quite difficult. To date, only a very few truly species-specific sequences have been identified. But recent work shows great promise in using computational methods to identify DNA signatures that can be used to position an unknown pathogen on a phylogenetic tree. (See the article “Analyzing Pathogen DNA Sequences” on page 182.)

Work to develop a detailed understanding of the phylogeny of different pathogens considered candidates for use as weapons continues today and is critical to the successful implementation of DNA-based methods for reliable detection and characterization of the range of pathogens we are concerned about.

Detection systems that do not detect all strains of a pathogen may miss the presence of the pathogen when it is really there. Conversely, the use of DNA sequences or other targets that are not threat-agent specific can result in false positives, causing disruption and expense for those who rely on them.

### Applying Our Methods to Bioterrorism

The first real-world application of our methods for pathogen detection and strain discrimination was the analysis of tissue samples from the victims of an anthrax outbreak in the former Soviet Union, near Sverdlosk, in 1979. Soviet authorities attributed the outbreak to contaminated meat. We analyzed tissue samples at Los Alamos in the 1990s (Jackson et al. 1998). VNTR profiles showed that the victims were infected by multiple *B. anthracis* strains. Because all natural outbreaks tested until then resulted from only a single *B. anthracis* strain, this finding strongly

suggested that there had been intentional mixing of strains. Thus, our finding validated other indicators that the outbreak was due to an accidental release of *B. anthracis* spores from a Soviet biological weapons production facility. This scenario was eventually proved and acknowledged around the time of the fall of the Soviet empire.

We learned much about the sophistication of the former Soviet Union’s bioweapons program from the analysis of these samples. For example, infection with multiple strains would complicate initial sample analysis. As a result, selection of an effective therapy would become problematic, opening up the possibility for the spore population to become resistant to multiple drugs or vaccines.

In the mid and late 1990s, we conducted an AFLP analysis on samples collected by UNSCOM inspectors in Iraq. In one case, a putative *Bacillus thuringiensis* isolate was collected at the Iraqi al Hakam facility. The Iraqis claimed the isolate was a subspecies of *B. thuringiensis* known as *kurstaki*, a bacterium widely used as a biopesti-

cide. Standard analysis methods could not dispute this claim. But AFLP analysis demonstrated that the sample was neither *B. thuringiensis kurstaki* nor any other closely related *B. thuringiensis* isolate. Comparison of its AFLP profile with our continuously growing collection of bacilli AFLP profiles shows that the sample was a nonpathogenic close relative of *B. anthracis*, with the same growth and spore production properties (see Figure 5). The Iraqis had an excellent surrogate for *B. anthracis* that had proved hard to identify by standard assays. Al Hakam was eventually acknowledged as the centerpiece of the Iraqi biological weapons production effort.

Another intriguing sample came to us even more recently (1998), again with a *B. thuringiensis* label. This time it was cultured from the infected wounds of a French soldier in Bosnia. Initial antibody-based analyses by other laboratories suggested that this was *B. thuringiensis*. Such results led to public concerns that *B. thuringiensis*-based biopesticides might be dangerous to humans in spite of 40 years of apparently safe use. As a result, European and North American regulatory agencies have been re-evaluating the use of this microbe. However, AFLP analysis of this sample shows that, like the al Hakam isolate, it is not closely related to the *B. thuringiensis* isolates that are used as biopesticides. Instead, its AFLP fingerprint suggests a very close relationship to *B. anthracis* and not to the insecticidal bacilli. (The purple arrow in Figure 1 indicates the identified species.) The story illustrates the problems that can arise when an assay with poor specificity yields a false positive (or when microbes with the same species name are erroneously assumed by regulatory agencies to have the same pathogenic properties).

## New and Improved Pathogens

Until recently, the prevailing view has held that naturally occurring pathogens are sufficient to generate an arsenal of biological weapons and that there is no need to further manipulate these microbes to enhance their effectiveness. However, rapid developments of molecular tools to make specific changes to a microbe's genome, the expansion of our knowledge concerning the biochemical bases for virulence and pathogenesis, and identification of genes and pathways that affect these factors provide opportunities to deliberately enhance pathogenicity or perhaps to introduce the genes that can make a normally innocuous microbe pathogenic. These developments, coupled with publication by scientists from the former Soviet Union's biological weapons program of methods that outline the results of genetically engineering *B. anthracis* to confer resistance to the current anthrax vaccine (Pomerantsev et al. 1997), strongly suggest that genetically engineered biological agents for weaponization are already a reality.

Resistance to a variety of antibiotics can result from inserting specific genes into a microbe or from selecting naturally occurring resistant isolates from laboratory-grown cultures. Methods that detect these changes are essential to ensure the best medical response to an outbreak. Los Alamos assays can detect eight different single nucleotide changes within *B. anthracis* responsible for conferring ciprofloxacin resistance (unpublished results). We have also developed rapid PCR-based assays that detect many of the DNA molecules used to genetically manipulate this pathogen. As the scientific community continues to unravel the mechanisms underlying pathogenicity and virulence, such knowledge can be used to defeat these microbes—or to enhance them. We must continue to develop assays that will detect such manipulations.

## The Future

The first step in responding to a naturally occurring disease outbreak or to an intentional release of a biological threat agent is detecting the presence of the microbe and determining those genetic characteristics that will provide information about its pathogenic properties and, perhaps, its source. However, the nature of biological agents requires their release before they can be detected. Therefore, we must pursue two important strategies, one being long term and the other, short term. Our short-term strategy should be to continue the development of rapid, sensitive, and accurate detectors and to develop strategies for their effective deployment. It should also include developing an effective means of intelligence to better track activities associated with the intentional production of biological threat agents. Our long-term strategy will be to better understand the pathogenic microbes responsible for these diseases so that we can effectively treat or prevent the diseases they cause. Our ultimate goal is to remove each of the threat agents from the threat list. The best approach to this end is to develop methods of effectively treating those who are exposed to the agent so that there is no health impact from an exposure. Such treatment strategies require a thorough understanding of the mechanisms underlying pathogenicity and virulence and the genes that encode these mechanisms.

The challenges to achieving these goals are huge and worthy of a major initiative. In the 1960s we set out for the moon, in the 1970s we started the "war on cancer," and in the 1980s we took on the fight against AIDS. At the beginning of the twenty-first century, we have been challenged to limit the spread of infectious diseases and prepare to defeat an adversary that might use biological weapons against our citizens, our crops, or livestock. As did the

preceding challenges, this new challenge will help drive advances in understanding and technology that will have broad benefits. While we are striving to more rapidly detect and characterize these pathogens and to understand the mechanisms underlying virulence and pathogenicity, our work will drive advances in biology, medicine, instrumentation, information technology, communications systems, and public health protection that will benefit us all. ■

### Further Reading

Farlow, J., K. L. Smith, J. Wong, M. Abrams, M. Lytle, and P. Keim. 2001. *Francisella tularensis* Strain Typing Using Multiple-Locus, Variable-Number Tandem Repeat Analysis. *J. Clin. Microbiol.* **39** (9): 3186.

Jackson, P. J., K. K. Hill, M. T. Laker, L. O. Ticknor, and P. Keim. 1999. Genetic Comparison of *Bacillus anthracis* and Its Close Relatives Using Amplified Fragment Length Polymorphism and Polymerase Chain Reaction Analysis. *J. Appl. Microbiol.* **87** (2): 263.

Jackson, P. J., M. E. Hugh-Jones, D. M. Adair, G. Green, K. K. Hill, C. R. Kuske et al. 1998. PCR Analysis of Tissue Samples from the 1979 Sverdlovsk Anthrax Victims: The Presence of Multiple *Bacillus anthracis* Strains in Different Victims. *Proc. Natl. Acad. Sci. U.S.A.* **95** (3): 1224.

Jackson, P. J., E. A. Walthers, A. S. Kalif, K. L. Richmond, D. M. Adair, K. K. Hill et al. 1997. Characterization of the Variable-Number Tandem Repeats in *vrrA* from Different *Bacillus anthracis* Isolates. *Appl. Environ. Microbiol.* **63** (4): 1400.

Keim, P., A. Kalif, J. Schupp, K. Hill, S. E. Travis, K. Richmond et al. 1997. Molecular Evolution and Diversity in *Bacillus anthracis* as Detected by Amplified Fragment Length Polymorphism Markers. *J. Bacteriol.* **179** (3): 818.

Keim, P., L. B. Price, A. M. Klevytska, K. L. Smith, J. M. Schupp, R. Okinaka et al. 2000. Multiple-Locus Variable-Number Tandem Repeat Analysis Reveals Genetic Relationships within *Bacillus anthracis*. *J. Bacteriol.* **182** (10): 2928.

Klevytska, A. M., L. B. Price, J. M. Schupp, P. L. Worsham, J. Wong, and P. Keim. 2001. Identification and Characterization of Variable-Number Tandem Repeats in the *Yersinia pestis* Genome. *J. Clin. Microbiol.* **39** (9): 3179.

Pomerantsev, A. P., N. A. Staritsin, Yu. V. Mockov, and L. I. Marinin. 1997. Expression of Cereolysine AB Genes in *Bacillus anthracis* Vaccine Strains Ensures Protection against Experimental Hemolytic Anthrax Infection. *Vaccine* **15** (17-18): 1846.

**Jill Trehwella** received her bachelor's degree in physics and applied mathematics in 1975 and master's in physics in 1978 from the University of New South Wales. She received her Ph.D. in chemistry from the University of Sydney in 1981. Following a postdoctoral research position in molecular biophysics and biochemistry at Yale University, Jill came to Los Alamos as a staff member in then Life Science Division. She had a joint appointment in physics to build a program in structural biology using neutron scattering. Jill has published a body of work focused on the structural biology of intracellular messengers and signal transduction in biomolecular networks. For her contributions, Jill was named Laboratory fellow in 1995 and a fellow of the American Association for the Advancement of Science in 1999. She is currently the Bioscience Division Leader.



**Paul Jackson** graduated from the University of Washington in 1974 with a bachelor's degree in cellular biology and received a Ph.D. in molecular biology from the University of Utah in 1981. Paul came to Los Alamos National Laboratory as a Director-funded postdoctoral fellow in 1981 and became a technical staff member in the Life Sciences Division in 1983. He was appointed a Laboratory fellow in 2001. He has worked on research related to biological threat reduction for many years and is recognized for his contributions to understanding variations among different strains of selected threat agents and for development of DNA-based detection and characterization of *Bacillus anthracis* and other threat agents.





# National Health Security

I. Gary Resnick

The terrorist use of *Bacillus anthracis* in the wake of the September 11 attacks heightened the national awareness of the threat posed by bioterrorism. It is also true, however, that the formerly clear lines between criminal uses of biological threat agents and natural disease outbreaks have become blurred. The advent of new diseases such as the acquired immune deficiency syndrome, or AIDS, and the severe acute respiratory syndrome, or SARS, the development and rapid spread of antibiotic-resistant *Mycobacterium tuberculosis*, and the rapid spread of West Nile virus are vivid reminders of our society's acute vulnerability to infectious diseases, regardless of whether those diseases arise naturally or by design.

The inevitability of emerging and reemerging diseases and the possibility of bioterrorism dovetail with increased health-care spending in the United States. Health costs now represent 14.1 percent of the gross domestic product, and in 2001, \$1.42 trillion was spent on health care. The economics alone invite a new approach to national health.

Advances in microbiology and computational science, in addition to those in secure communications, provide a compelling opportunity for detecting and combating disease through a national, and eventually international, health-security program. The goals of such a program would be the early detection of a natural epidemic or terrorist incident, appropriate diagnostic evaluation of affected individuals, identification of effective

treatment (secondary prevention), and implementation of appropriate primary prevention and control strategies.

National health security may be possible because technology that will allow cost-effective global surveillance and response is rapidly being developed. A comprehensive environmental monitoring system for detecting infectious agents in air, water, and food sources, coupled with population-based medical surveillance, could result in early identification of the initial phase of a natural or intentional epidemic or incident. Early warning makes it possible for the health care system to assess the medical and public health significance of these events and to respond as appropriate. Monitoring the population holds the potential for achieving the vision of dual-use systems that will help maintain and improve public health while allowing quick response to bioterrorist acts.

Although the availability of sensitive and specific sensors is central to this vision, just as critical is the capacity for high-throughput laboratory analysis and algorithms for analyzing the large body of data that will be produced in real time. But the age of modern medicine is fueled by rapidly advancing molecular applications of genomics, proteomics, and information management and analysis (or bioinformatics), as well as high-throughput screening of candidate vaccines and therapeutics. In this context, developing biomedical tools to reduce morbidity, mortality, and health care costs can become a reality.

In addition to facilitating effective response to biological threats in the United States, a national health security program can provide biomedical tools and systems for the developing world, where many of the potential bioterrorist agents are endemic and are a major cause of morbidity and mortality. Implementing a national health security program can therefore reduce the degree of economic and social asymmetry between developed and developing nations that fuels international bioterrorism.

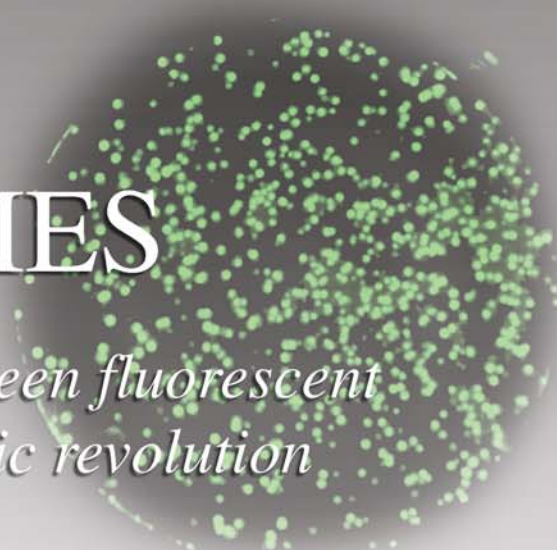
Ironically, the challenge presented to us by bioterrorism, intensified by rapid scientific advances in biotechnology, creates a crossroad between the threat of disease and the promise of national, if not global, health security. If society is to enjoy the potential long-term benefits afforded by current and future technological advances, the public health and national security communities will have to work hand in hand. Being the wealthiest country in the world, the United States has a unique role to play in determining which path the world will follow. ■

For further information, contact Gary Resnick (505) 665-0770 (resnick@lanl.gov).

# FLUOROBODIES

*Mixing antibodies and the green fluorescent protein to unravel the genomic revolution*

*Andrew M. Bradbury, Geoffrey S. Waldo, and Ahmet Zeytun*



**A**ntibodies are proteins produced in animals and humans in response to infection. They bind tightly and selectively to infecting agents (antigens, such as surface proteins found on bacteria or viruses), providing the mechanism by which those agents can be destroyed. Because they bind so well to their targets (usually proteins), antibodies are used extensively in biological research to identify proteins. They also form the basis for many important diagnostic tests, such as the pregnancy test, as well as almost all tests involving infectious diseases.

We have combined the selective binding properties of antibodies with the extraordinary fluorescent properties of a protein found in a jellyfish to synthesize a new type of protein called a fluorobody. Like antibodies, fluorobodies bind tightly and selectively to antigens, but unlike antibodies, they glow a very visible green when illuminated by blue light. They will therefore become a powerful tool for biological research, for which seeing the location of proteins and having a means to track their interactions are most significant. Fluorobodies also have the potential for use in biosensors to detect infectious and biothreat agents.

## Antibody Libraries

Antibodies bind to antigens and subsequently initiate immune responses as a result of their modular structure, which is basically the same for all antibodies (see Figure 1). A region found at one end of the antibody—the variable region—varies (as its name suggests) between different antibodies. This is the part responsible for recognizing and attaching to antigens. At the other end of the antibody is the so-called constant region. In humans, there are seven different types of constant regions, and these are the parts that activate the killing mechanisms once the variable region has latched onto its target.

Although antibodies are produced naturally by the immune system against infecting microorganisms, they can also be made artificially against almost any target if that target is injected into an animal. Such a process is called immunization and involves three or four injections of the target into the animal (usually mice or rabbits) until antibodies can be detected in the blood. The ability to artificially induce antibody production has been a major reason to

use antibodies so heavily in biological research.

Whereas animals are very effective at generating antibodies by immunization and have been used for this purpose for decades, 10 years ago a method to create specific antibodies that did not require the use of animals was developed in a laboratory in Cambridge, the United Kingdom. At any one time, 100 million to one billion different antibodies are thought to be present in human blood. Each is produced by a specialized cell, called a B cell, which circulates in the blood. Each B cell has a different specific “antibody gene” (the gene is a section of DNA that describes how to make the antibody) that is made of a unique arrangement of variable gene regions (refer to Figure 1). Because of its unique antibody gene, each B cell makes antibodies of only one specificity.<sup>1</sup> By taking blood from many different human volunteers and harvesting all the B cells, the Cambridge researchers could extract millions of different antibody genes. Then they

<sup>1</sup> B cells will clone themselves; therefore a small number of B cells will produce the same antibody.

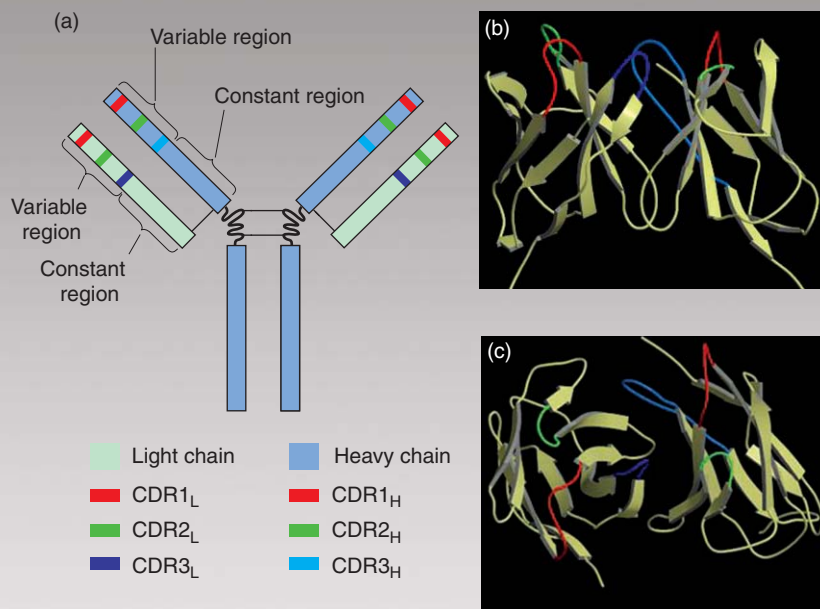


isolated the genes that made the variable portions of antibodies and separated them from the genes that made the constant region. In this way, they created a “library” of human variable-region genes.

It was hypothesized that any specific antibody could be produced if its antibody gene could be isolated from the library and inserted into bacteria, which would then synthesize the antibody encoded by the gene. The problem was that the library was simply a collection of unlabeled molecules. Although techniques were available to insert genes into bacteria, identifying the single, correct antibody-producing bacterium among millions and millions of different bacteria was extremely difficult.

The Cambridge researchers surmounted this difficulty by adopting a strategy of physically coupling the antibody genes to the antibodies they make. They carried out this coupling in the laboratory using a phage, which is a virus that infects only bacteria. The particular phage, called fd, has only five different proteins on its surface, one of which is called gene 3 protein. Using gene-splicing techniques, the researchers fused an antibody gene to gene 3 of the phage. The genetically engineered “antibody phage” was identical to the normal fd phage, except that it displayed an antibody on its surface, as seen in Figure 2. The important point is that phages can easily be made to replicate in the laboratory. By inserting the antibody gene into the phage, the Cambridge researchers had, in effect, created an antibody that could be replicated.

Then they created a phage antibody library consisting of millions of phages, each containing a different antibody gene and hence displaying a different antibody on its surface. It is possible to select specific antibodies from the library by mixing the entire library with a protein target of interest (see Figure 3). Some of the



**Figure 1. Antibody Structure and Binding**

(a) All antibodies have the same basic structure: two heavy chains and two light chains covalently joined together to form a complex. The figure shows a linear representation of the amino acid sequence of the most common antibody type, immunoglobulin G (IgG), which has a Y-shaped structure. At the ends of the arms of the Y are the so-called variable regions of the heavy and light chains. This is the part responsible for recognizing infectious targets (antigens) and attaching to them. The recognition of such targets is largely mediated by the three complementarity-determining regions (CDRs), which are hypervariable regions found interspersed within each of the variable regions. (b) This ribbon diagram shows a side view of the three-dimensional structure of the variable regions, whereas (c) shows a top view. The CDRs form protruding looplike structures, creating a unique surface that will bind to a specific antigen with high affinity.

phages would bind to the target, whereas the nonbound ones could be removed by washing. The bound phages could then be eluted from the target and allowed to replicate in a controlled environment. Thus, the library could be used to produce antibodies that have specificity for the target of interest.

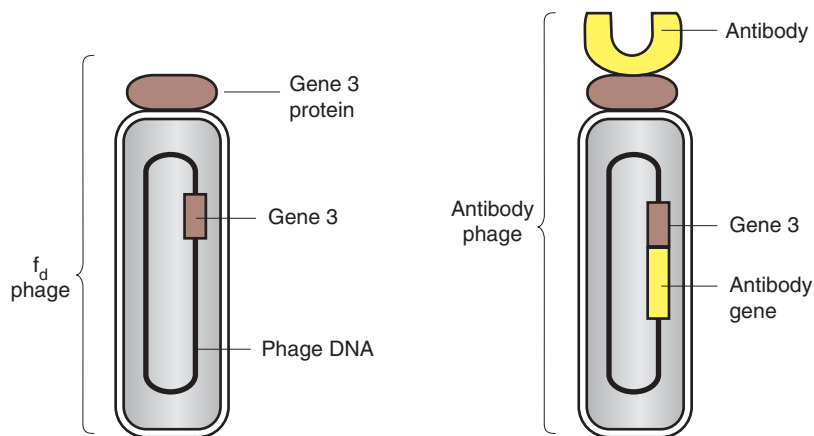
## Fluorobodies

One problem with the antibodies generated by the library technique is that they tend to fall apart relatively easily. Furthermore, to be detected, they require the addition of other reagents—a problem common to all

antibodies, even those derived from rabbits or mice. Our group at Los Alamos addressed these problems by inventing fluorobodies.

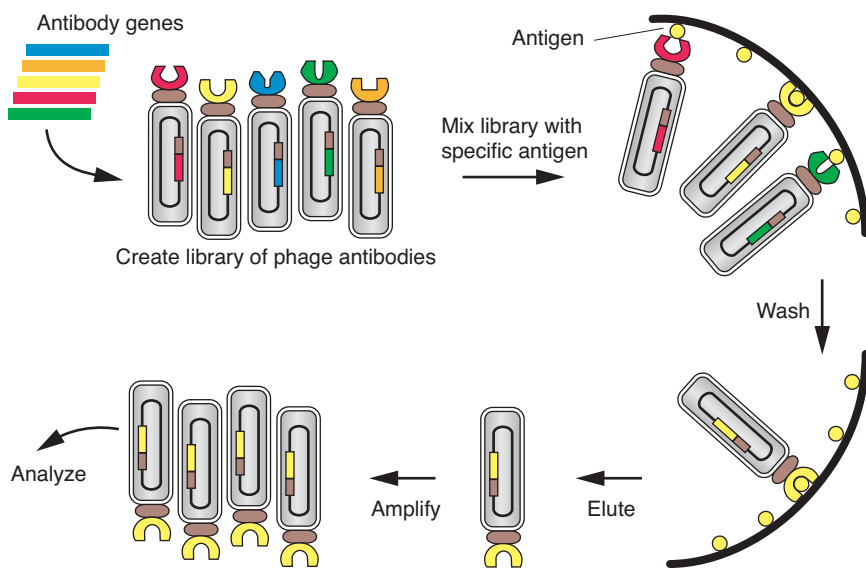
The underlying structure of the fluorobody is a remarkable protein called green fluorescent protein (GFP), which is obtained from the jellyfish *Aequorea victoria*. Expression of this protein alone will render fluorescent any tissue, cell, or animal when it is viewed under blue light. GFP has an extremely stable canlike structure, as seen in Figure 4.

Our working hypothesis was that the CDR3 hypervariable loop of an antibody (refer to Figure 1) could be spliced into GFP, thus creating a fluorescent protein with antigen-binding



**Figure 2. Phage Library Detection Technique**

By fusing an antibody gene to one of the genes that produces a phage surface protein, we create an antibody phage that displays a specific antibody.



**Figure 3. Selecting Specific Antibodies**

One can select antibodies that bind to specific targets by mixing the entire phage antibody library with a target of interest. Only those phage antibodies that have an affinity for the target will bind; all others can be washed away. The selected phage antibodies can then be eluted from the target and cloned to large numbers in the laboratory for subsequent use.

capabilities. It was not at all obvious that such a splice would produce a viable product. Previous attempts at inserting amino acids (the constituents of proteins) into GFP had typically resulted in unstable, nonfluorescent

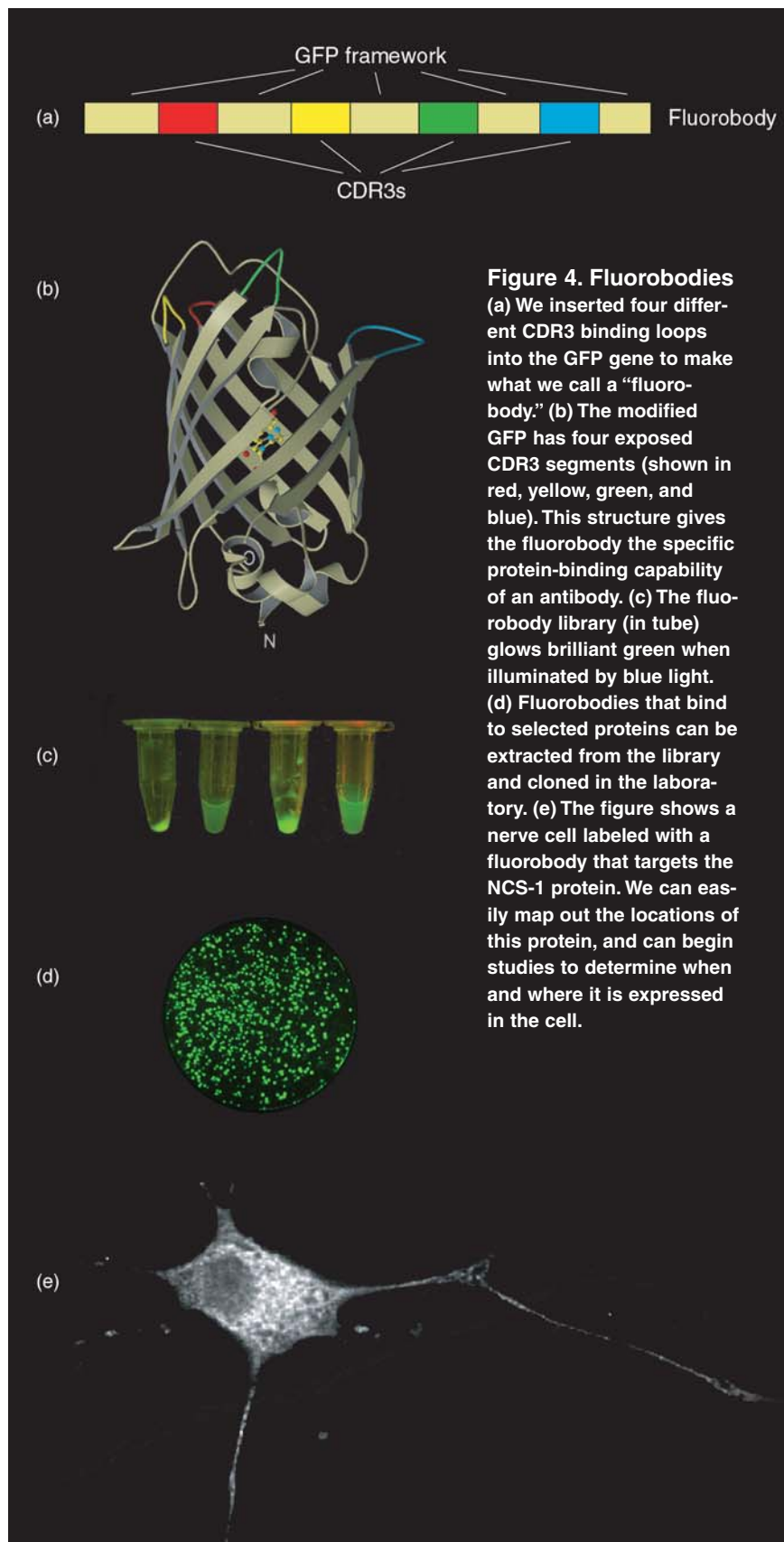
molecules. But the CDR3 loop tends to be very “floppy.” Our guess was that it could be inserted into the relatively exposed loops at the top of GFP and would not affect the canlike structure responsible for fluorescence.

We first developed a novel protocol that would insert CDR3 genes into the GFP gene at four different sites. The result was what we hoped would be a “fluorobody gene” that produced a modified GFP protein with four exposed antigen binding loops. Then using tens of millions of different CDR3 genes, we synthesized millions of different fluorobody genes. We next coupled the fluorobody to its own DNA by fusing each gene to gene 3 of the fd phage, thus creating a phage fluorobody library. We were gratified to see the library glowing brilliant green in its sample tube. Then following the same selection procedure described above, we obtained a number of highly specific fluorobodies that bound as tightly to their targets as antibodies. We had successfully created stable, easily produced, and easily detectable protein markers.

We can now use fluorobodies to track specific proteins and observe in which tissue, cells, or organelles they are expressed and with which molecules they interact. This capability is crucial for the next phase of the Human Genome Project, for which Los Alamos researchers are developing methods to select fluorobodies against hundreds of targets simultaneously. Fluorobodies will likely prove to be very powerful in the development of novel diagnostic tests and should accelerate drug development, because the ability to observe the effects of a drug in real time will allow very rapid screening against millions of compounds. ■

**Acknowledgments**

This work has been supported by the Department of Energy and by Laboratory Directed Research and Development funds.



For further information, contact  
 Andrew Bradbury (505) 665-0281  
 (amb@lanl.gov).

# Analyzing Pathogen DNA Sequences

Thomas S. Brettin

**D**NA-based experimental techniques allow rapid detection and identification of pathogens for medical treatment, criminal forensics, and possibly attribution (that is, finding the source of an outbreak). Bioinformatics provides the computational tool for that identification process. Loosely defined as the merger of computers and biology, bioinformatics evolved significantly during the Human Genome Project in response to the biologists' need to assemble a complete genomic sequence from DNA fragments. In the area of bioterror reduction, we are primarily concerned with analyzing the information contained in a pathogen's genome. Bioterror reduction, therefore, requires a collection of computational tools and databases that are different from those used to determine the genomic sequence.

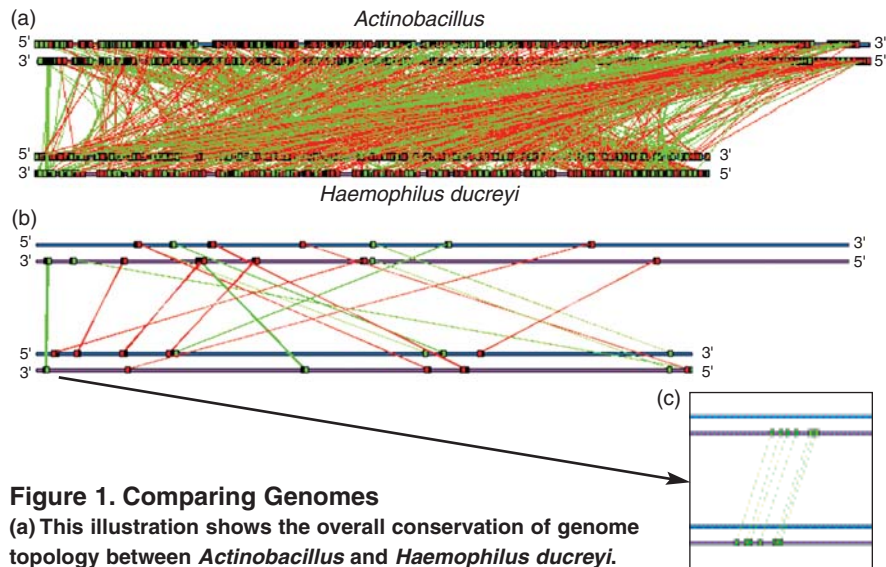
The initial step in creating an identification tool for reducing the biological threat is to identify the complete set of genes contained in the pathogen's genome. This task is generally accomplished by use of hidden Markov-based techniques, which enable us to calculate the probability of finding the next base in a DNA sequence, given the bases directly preceding it. It turns out that in regions of DNA that encode genes, the next nucleotide in a sequence can be predicted with high probability based on the previous four to seven nucleotides. In genomic sequences that do not code for genes, this predictability is significantly less. DNA sequences that have high predictability are therefore identified as gene sequences. For organisms such as the common bacterium *Escherichia coli*, this approach has shown better than 95 percent accuracy. We routinely apply this

approach to pathogenic bacteria that represent a threat to our safety, such as *Bacillus anthracis* and *Yersinia pestis*, the causative agents of anthrax and plague, respectively.

With the gene predictions in hand, we are in a position to interpret the functions of these genes in the cell, as well as identify genes that are unique to a pathogen. At its simplest, this procedure is known as functional annotation. It is a difficult process, in part because genes from two organisms that perform the same function rarely have the same DNA sequence. We use various statistical techniques to assess the degree to which a gene sequence is similar to anything in a database of known gene sequences. Even in a single-cell organism like *B. anthracis*,

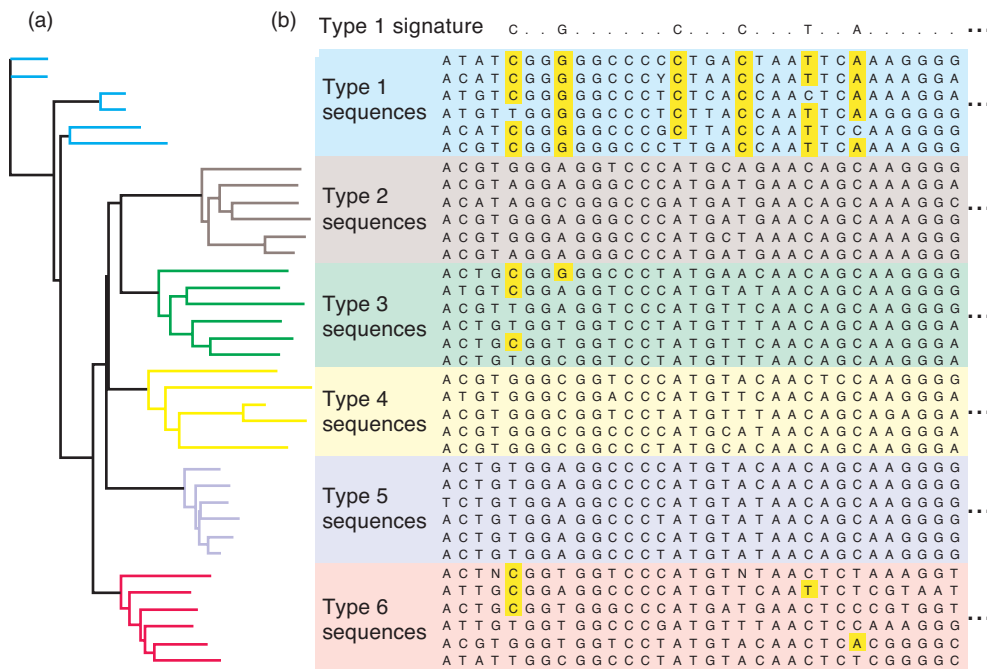
which contains between 5000 and 6000 genes, about 40 percent have no statistically significant similarity to anything in the database. Recent studies have shown, however, that the functions of the protein products of two genes can be remarkably similar, despite any discernable sequence similarity between the genes. Thus, our techniques look not only for overall similarities to specific genes but also for similarities to conserved functional domains within the genes.

Current technology allows us to complete 95 percent of the DNA sequence of a bacterium of five million bases in a matter of days. Because decreasing that time even further is expected, studies that compare the genomes of closely related



**Figure 1. Comparing Genomes**

(a) This illustration shows the overall conservation of genome topology between *Actinobacillus* and *Haemophilus ducreyi*. Both bacteria have only one chromosome, which is represented by the two heavy lines (one for the forward strand and one for the reverse). Genes appear on both strands. The green lines connect similar genes that appear on the same strand in both genomes, whereas red lines connect genes appearing on opposite strands. Most genes are present in both genomes, yet overall, the gene order, or the location of the gene, is not conserved. (b) There are, however, contiguous stretches of six or more genes that are conserved in gene order. (c) The detail shows a stretch of six genes that are involved in cell envelope biosynthesis and antibiotic susceptibility. These genes are conserved in both gene order and function and therefore point to an important evolutionary link between the two bacteria.



**Figure 2. DNA Signature**  
 (a) This phylogenetic tree shows the evolutionary relationships of a subset of hepatitis C viruses. (b) Given the tree, we can align partial DNA sequences and identify individual bases that can be used to distinguish a particular clade. The signature for Type 1 hepatitis C viruses is shown below the Type 1 sequences. It can be used to identify such viruses, or when used in conjunction with other signatures, to help characterize unknown viral strains.

bacteria will become more common. Such studies may help us understand the remarkable, possible physiological differences between closely related organisms. For example, although the genomes of *B. anthracis* and its near neighbor *B. thuringiensis* are greater than 99 percent similar, the former can kill humans, whereas the latter is harmless to humans and is widely used as a pesticide. Our first examination of these two genomes has revealed small sections of unique DNA, scattered throughout each genome, that range from small mobile DNA elements (insertion elements, transposons, and phages) to regions of around 25 genes of unknown function. The unknown genes presumably relate to existing physiological differences.

Other comparisons such as those between *Actinobacillus* and *Haemophilus ducreyi* show small regions of conserved gene order in an otherwise nonconserved genome topology (see Figure 1). In addition to their potential evolutionary significance, those conserved regions may also serve as gene targets for detection assays and disease mitigation strategies.

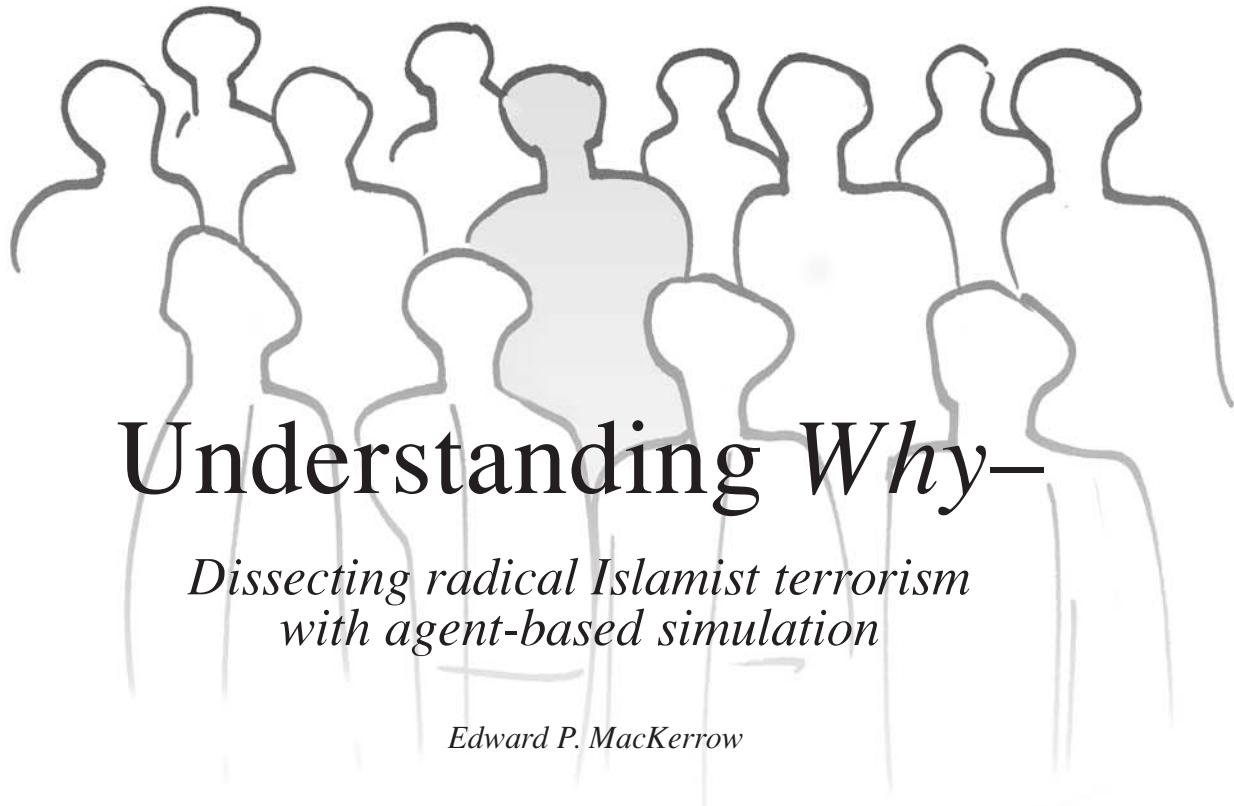
The article “Reducing the Biological Threat” on page 168 describes how variable number tandem repeats and single nucleotide polymorphisms can be used as “signatures” to differentiate among different strains of a single organism. Given the genomic sequence, we can easily locate these simple repeats and supply laboratory personnel with the necessary information to design and test an assay. But we are also developing techniques to construct new signature sets that will identify pathogen strains.

The DNA sequences from thousands of related pathogens are first used to construct a phylogenetic tree, as seen in Figure 2(a). (Note that the tree seen in the figure is “pruned” to show only a small subset of branches. Each branch end therefore corresponds to a specific DNA sequence that is representative of many similar sequences.) Overall, the branches cluster into groupings, or clades. We examine the DNA sequences that make up each clade and identify a set of individual bases that is highly likely to be seen in the chosen, but not in other sequences—see Figure 2(b).

This set of bases—the signature—provides a statistically powerful means to discriminate among clades and allows us to place new strains/DNA sequences into their genetic context.

Genome sequencing has opened a new era in the study of pathogens. Soon, hundreds of genomes from closely related pathogens will be available, and understanding what makes pathogens both similar and different will start at the DNA level. The looming challenge will be to develop methods for rapid detection and identification of pathogens, as well as new treatments. Making good use of genomic sequence data is a significant step forward in our goal of meeting that new challenge. ■

For further information, contact Thomas Brettin (505) 665-3334 (brettin@lanl.gov).



# Understanding *Why*—

## *Dissecting radical Islamist terrorism with agent-based simulation*

*Edward P. MacKerrow*

**T**he events of September 11, 2001, were a wake-up call for Americans. All of us remember where we were and what we were doing when we learned of the horrific attacks. On that terrible day, I was presenting results of a financial agent-based simulation to corporate executives on the top floor of a Houston skyscraper. I had no idea why the United States was attacked, but I had plenty of time to contemplate the question as I drove back to Santa Fe in my rental car. What had we done to motivate terrorists to take such hateful actions against us? What message were they trying to send us? Why did these attacks happen?

As the world strives to understand terrorism better and struggles to identify its many faces and forms, few would deny that religiously motivated terrorism is becoming increasingly prevalent. But the association of religion with violent, radical groups, many of which have their own interpretation of a religion, needs to be examined carefully. The violent actions of radical Islamist groups, for example, have led to the mistaken

association of terrorism with Islam.

One-fifth of the world population is Islamic. Dispersed around the globe, the largest concentrations of Muslims are in Indonesia, Pakistan, India, Bangladesh, and the Middle East. In Western society, the largest concentration is in the United States. The religion itself is based upon peace (the word Islam means “self-surrender” in Arabic, and the universal greeting of Muslims is “*salaam alaikum*,” which means “peace be upon you”). In many Islamic societies, however, the passive or neutral behaviors of the peaceful majority often become obscured by the attention-seeking acts of a “noisy minority.” Although the point is debated, the general understanding is that most Muslims are peaceful because of their Islamic beliefs, and that the “noisy minority” has misinterpreted Islamic teachings.

Muslims are taught that Muhammad was sent by Allah to spread belief in a single God—as opposed to the multitude of pagan rituals honoring a variety of deities at his time (A.D. 610). For many modern-day Muslim radicals, especially

those in traditional societies, American pop culture may be perceived as being similar to old-fashioned paganism, a cult that worships money and sex. Some modern-day militants may perceive themselves as following a path similar to Muhammad’s in cleansing the Islamic world from the infiltration of the “pagan” West. Other Islamic people may fear that their culture, traditions, and beliefs are being replaced because globalization imposes Western values on them. Being mutually understanding of religious sensitivities, as well as responsible and respectful of each other’s influence, will help establish a peaceful coexistence between the West and the Islamic world. Finding ways to move toward this goal requires careful analysis and discussion.

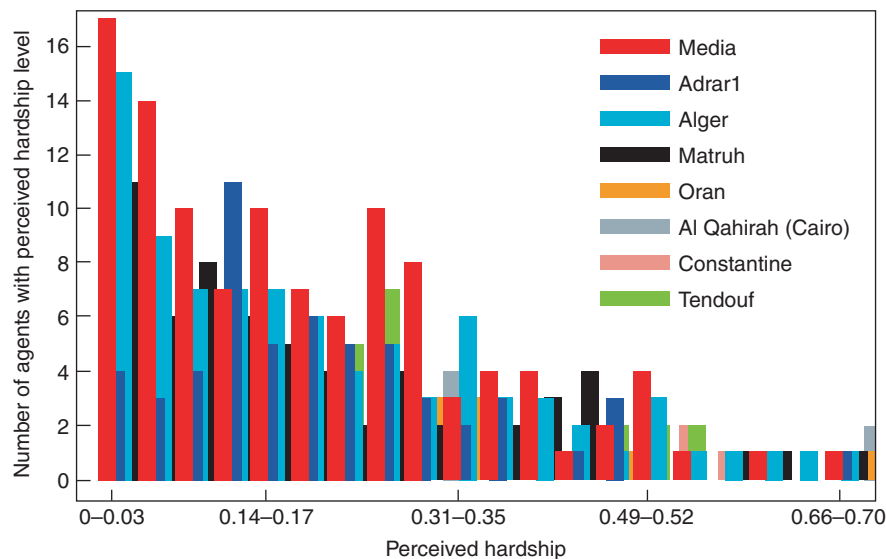
The Complex Systems Group at Los Alamos has been examining questions related to the “why” behind terrorist organizations in the Middle East. Borrowing tools from the field of computational economics and sociology, we are developing agent-based models that simulate social networks and the spread of social grievances

within those networks. Our computer-generated “agents” are humanlike, endowed with personal attributes and allegiances that statistically match the demographics of a specified region and, like people, interact with one another and respond to societal pressures. The defining feature of our agents, however, is that their “behaviors” are allowed to change during a simulation run. For example, an agent may “learn” during the simulation not to interact with agents of a certain social class, or an agent may develop deep “feelings” of oppression and grievance based on its experiences.

We do not know *a priori* the life stories of our agents, but after tens of thousands have interacted, we have produced a scenario, or a virtual history, for a region of interest. The plausibility of this scenario is normally assessed by human experts who have complete knowledge of the model assumptions and the rules followed by the agents. By replaying any particular simulation, the experts can observe how the agents behaved and examine why they behaved in a certain way.

We can expose our agents to a variety of determinants—new government policies, different media exposure, economic pressures, and others—and quickly generate hundreds of new scenarios. Thus, we can conduct computational experiments that can be analyzed statistically and objectively to increase our insight, support decision making, and aid policymakers (see Figure 1). Scenarios can even be used to gain insight into actual events that have little or no historical precedence. It should be emphasized, however, that the goal of these simulations is not to predict *specific* events and not to estimate the probability or frequency of terrorist acts, but to generate scenarios and analyze them.

Our work is part of the Defense Threat Reduction Agency (DTRA) research effort known as the Threat



**Figure 1. A Sample Distribution from the TAP Model**

The TAP model places thousands of agents throughout the Middle East, endows them with numerous properties and behaviors, and allows them to interact for number of simulation years, thus creating a short, virtual history of the region. We can dissect that history and analyze various social metrics (such as perceived social disadvantage, directed grievances, and allegiances) to gain insight into extremist behaviors. The figure shows a distribution for perceived hardship, a metric that might be obtained in reality by a pollster asking the question, “On a scale of 0 to 1, what is your level of hardship?” The colors correspond to some districts in Algeria and Egypt. Because agent behaviors are not preprogrammed, each simulation using the TAP model will produce a different distribution of perceived hardship (or any other social metric). Human experts must assess whether the envelope of results is valid, that is, whether our model correctly brackets the possible levels of real-world social grievances. That task is difficult because the model can be validated with only one data point—the real-world poll made under specific conditions.

Anticipation Program (TAP).<sup>1</sup> Built on a multidisciplinary team of scholars from the arts and literature, history and psychology, Middle Eastern and Muslim cultures, religion, economics, and sociology, TAP aims to develop algorithms and software frameworks that can generate the most likely models of terrorism and terrorist scenarios in order to catch the precursor signals of the next terrorist attack. We hope that TAP can eventually provide us with insight into potential prevention,

interdiction, and mitigation policies.

## Modeling Complex Socioeconomic Systems

The social tensions in the Middle East emanate from many different yet interrelated conflicts, and each Middle Eastern nation has a unique history in relation to those conflicts (Miller 1997). Therefore, the underlying social processes cannot be understood by a simple linear combination of separate sociologic, economic, demographic, religious, cultural, and political subprocesses.

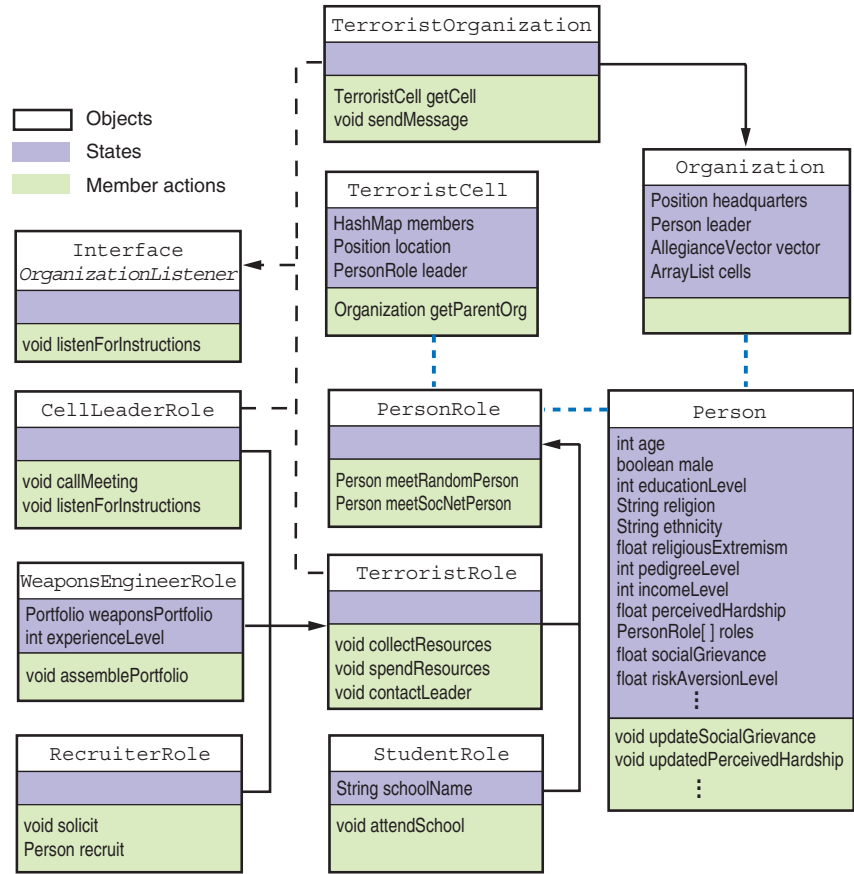
Agent-based simulation provides

<sup>1</sup>TAP was originated by Dr. Stephen Younger, director of DTRA. Younger is a former director of the nuclear weapons program at Los Alamos.

us with a methodology for modeling complex socioeconomic phenomena. Agent-based simulation was first introduced into economics to address shortcomings with economic simulations, which in early versions assumed homogeneous populations of idealized, perfectly rational agents who had perfect information about perfect markets. The results of those simulations, though frequently used, were often incorrect because of the flawed representation of real-world agent behavior. See Shubik (1997) for a discussion of issues associated with game theory applied to real-world applications.

As the computer became further integrated into the social sciences, more realistic socioeconomic models were attempted, and the methods of agent-based mathematics began to develop. Agent-based models have now found widespread use in economics and allow agents to act with bounded rationality, based on imperfect or incomplete information, and to act on chance and perceived economic utility. In addition to more realistic representations of individuals, agent-based simulation allows for analysis of nonequilibrium conditions compared with the historical practice of analysis made at equilibrium points—many real-world socioeconomic systems are not in equilibrium and may never reach equilibrium.

Currently, the major difficulty we face in building a model of a complex socioeconomic system is in quantifying social situations. Not only do we need better models that show how to represent social interactions, but we also need better empirical analysis of actual real-world studies. A fundamental problem is that real-world “observables” may be generated by many different interaction processes; therefore, empirical findings are open to different interpretations. My belief is that certain social micromodels apply better than others, depending on the context in



**Figure 2. Objects in the TAP Model**

Object-oriented programming is a natural fit to agent-based modeling because we can design objects that learn and adapt based on their history, their current state, and the states of other objects. This class diagram shows some of the object types used in the simulation. The simulation is built upon many different instances of these object types, each with different attributes. The object architecture allows for flexibility; the `PersonRole` class, and its inherited subclasses, allow a construct where any one `Person` object can play multiple roles. Interfaces allow for specification of required actions that can be implemented differently, depending upon the type of object implementing the interface (interface relationships are shown by black dotted lines). Objects can be composed of other objects allowing for new objects to build upon the structure of existing objects (shown by blue dotted lines).

which they are applied—and a good model will utilize a broad suite of social micromodels.

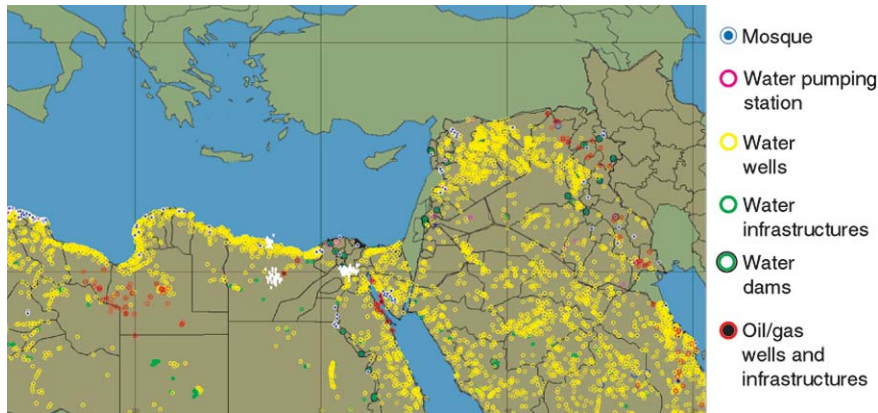
### The Los Alamos TAP Agent-Based Model

The TAP simulation is built with object-oriented software and implemented in the Java programming lan-

guage. (The development of object-oriented software has been directly related to rapid advances in agent-based simulation over the last 15 years.) Although, in theory, any computer language can be used to represent a social system, object programming fits naturally with modeling social systems and greatly reduces development time.

In a simplified description, we first





**Figure 3. Region and Other Objects in the TAP Model**

A screen capture from the TAP model shows many of the geographical region objects used to delineate different groups of agents as well as other objects. Each region is described by a different demographic composition. Agents are instantiated inside each region on the basis of these regional demographics. The white areas in northern Egypt, for example, indicate clusters of agents.

**Table I. Attributes Used to Set the State of Agents**

Agent Attribute	Distribution Type	Data Source
Age	Empirical (discrete)	U.S. Census Bureau
Sex	Empirical (male, female)	U.S. Census Bureau
Education	Empirical (years)	CIA World Factbook
Education type	Estimated (religious, secular)	CIA World Factbook
Ethnicity	Empirical (percent by group)	CIA World Factbook
Religion	Empirical (percent by group)	CIA World Factbook
Extremism	Estimated ([0,1])	Interviews and readings
Pedigree	Estimated ([0,1])	Interviews and readings
Income	Empirical (\$ per year)	World Bank
Married	Empirical (Boolean)	The Economist
Employment	Empirical (Boolean)	The Economist, World Bank
Location	Empirical (# per km <sup>2</sup> )	GIS Data Sources

establish the initial properties of “objects.” The most important objects in the TAP model are Nation, District, Organization, Mosque, and Person. (In this article, we identify all software objects by capitalization and by using a Courier font.) Each object contains attributes that support a corresponding abstraction representation and modes of actions that represent agent behaviors (see Figure 2). We divide each Middle Eastern nation into a series of administrative District objects, as

defined by standardized Geographical Information System data (see Figure 3). Each District is then populated with a number of agents. Empirical distributions derived from regional demographic and ethnographic data (see Table I) are used to initialize agent attributes. Different types of agents are instantiated with different data sets and different rules of behavior.

For each District, we define relative weighting factors to estimate the “social welfare,” or “social capi-

tal,” of religious and ethnic groups in that region. We posit that social capital is a weighted sum of income, ethnicity, religion, education, and pedigree, where pedigree represents inherited or appointed social wealth. (A Saudi prince would have a very high pedigree value.) We allow for flexibility of social weighting factors across different districts; for example, in some districts, ethnicity may not influence social status as much as does income.

We also take into account the important aspect of social rank in a society. Each agent updates both social capital and social rank on a regular basis during a simulation run. As discussed later in the text, social rank plays an important role in the theory we use to model social interactions—although by itself it is not necessarily a determining attribute of a terrorist.

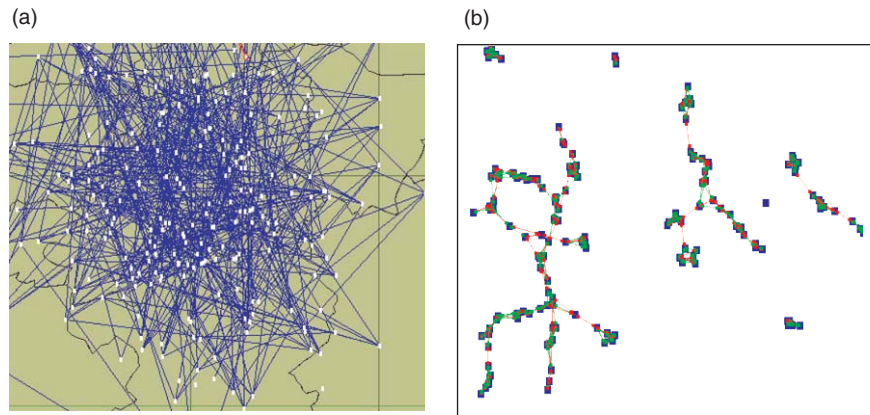
## Modeling Interactions

Many of the behaviors and opinions of individuals are rooted in the social structures to which they belong. For example, a young adult’s proclaimed dislike toward the United States may be socially inherited from parents and reinforced through family circles, friendship circles, and by the media. Understanding which social experiences, conditions, and interactions increase the likelihood of becoming a terrorist is very difficult. Lay explanations for an individual’s choice to become a terrorist include desperation, poverty, mental illness, and lack of education. Statistics of suicide attackers, however, show that low income levels are neither necessary nor sufficient to explain suicide attacks. Education is not a determining factor either, and ironically some data may suggest that higher education levels are positively correlated with terrorist attributes. So what is it then? The answer may be found by a

closer examination of the terrorist organizations themselves.

Many terrorist organizations use charismatic leaders to cultivate and indoctrinate small cells of young recruits to become martyrs for the overall cause (Atran 2003). (Note that, for terrorist organizations, the use of martyrs is very economical, approximately \$150 per suicide bomber attack.) These leaders, who rarely become martyrs themselves, seek to establish a commitment between the members of the cells in the form of a social contract—usually sealed with a video testimony. A cell member develops a sense of obligation to the fictive “kinship” of the cell. The social cost to individuals for renegeing on commitments is very high; they risk being labeled “kafir”—an infidel, or nonbeliever. Defectors from some terrorist organizations can even be killed by the organization. Peer pressure from the cell is a motivating factor for some terrorists; others are fully committed on their own and believe strongly in becoming terrorists. The latter appears to be particularly true with respect to terrorists who oppose Israeli settlements in the occupied territories of Palestine. In the TAP model, we include different types of agents and allow some agents to be more self-motivated toward terrorism and others to require more or less active recruitment.

Structural and institutional constraints induce individuals to act in a manner most consistent with the preferences of the social structure or institution—in this case, the terrorist organization. In the TAP model, we introduce these constraints using social network representations for establishing probabilistic social rule-sets. To include realistic interactions, we construct several different types of social networks between agents: kinship, religious, organizational, and friendship. Although meetings can occur between any two agents or



**Figure 4. Evolving Networks**

(a) This screen capture shows a large friendship network of agents in the TAP model at the end of the preprocessing stage. At this point in time, the network is rather dense; each agent is connected to an average of five other agents. A few agents initialized as Armed Islamic Group (GIA) terrorists are shown in red. (b) As agents interact with other agents during the simulation, the network evolves. The colors of the lines connecting the agents reveal how the agents became friends—either through random meetings or through mutual friendships. Some agents become isolated in stranded cliques and associate with very few other agents.

between an agent and a group, we assume that interactions are more probable between agents on the same social network than between random agents in the local population, and we weight the interactions accordingly.

A screen-capture image of an initial friendship network of agents is shown in Figure 4(a), where we have used the algorithm of Jin et al. (2001) to construct the network. It is obvious how dense the friendship network is at the end of the preprocessing stage—each agent has a large number of social contacts. During the following stages of the simulation, this network will evolve into cliques and a less uniform density as agents evolve their social networks—as shown in Figure 4(b).

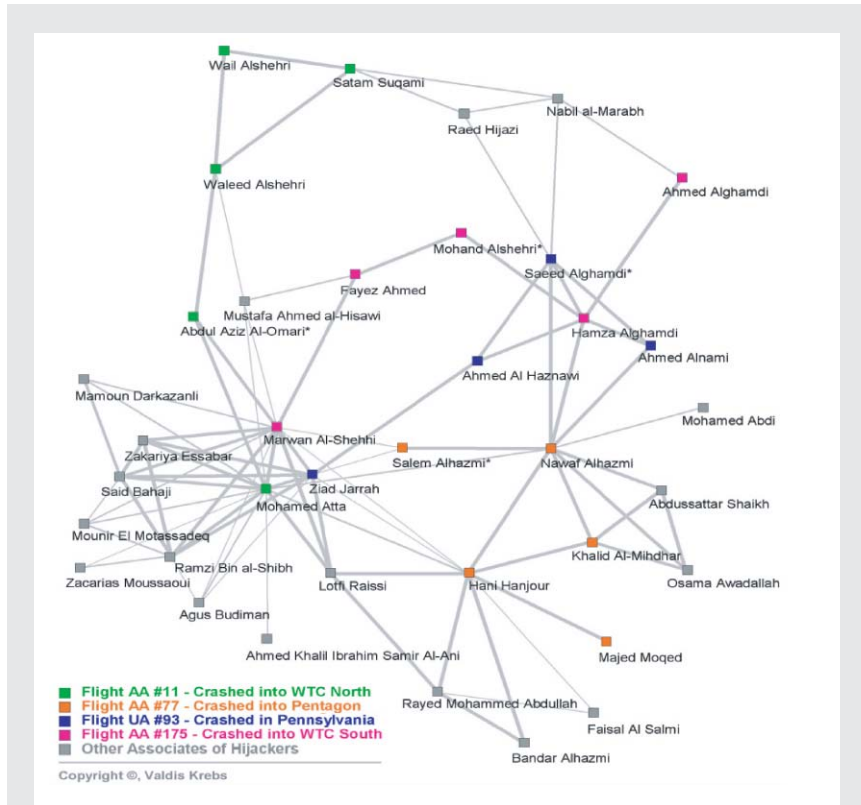
We include simulated social networks as structures that affect social interactions in our model, although we have also included the ability to populate the model with known terrorist networks. Obtaining data and characteristics on the actual social structures of the various terrorist organizations around the world is difficult because of their covert nature. In place of existing terrorist net-

works, we use surrogate information to model these social networks. (An example of a surrogate network is given in the box, “An al-Qaida Network” on the opposite page.)

## Interacting Agents

When an agent meets another agent in our simulation, a social interchange occurs. Depending on the outcome of the meeting, a certain amount of interactive learning occurs. Because interactions are part of the larger social structure, a form of “social learning” occurs throughout the agent population—observable in part through their heterogeneous allegiances.

Each agent in the TAP simulation carries an “allegiance vector.” The elements of this vector contain an integer value representing that agent’s allegiance toward specific nations, organizations, ethnic groups, or religious groups. A positive (or negative) value of allegiance suggests a positive (or negative) allegiance for the group associated with that element. Figure 5(a) shows a schematic of the alle-



### An al-Qaida Network

The compartmentalized social network of the al-Qaida cells involved in the September 11 hijackings, researched by Valdis Krebs (2002), is shown above. Krebs obtained open-source data on the hijackers as those became available after the attacks. Although some nodes are likely missing, his analysis and construction of the network are very useful in understanding some features of terrorist cells.

The social network of the hijackers was very loosely connected and sparse. Whereas Mohamed Atta is clearly seen as the ringleader, many of the hijackers were separated by a few degrees—more than one step away from each other. This separation even applied to hijackers on the same flight. The strategy ensures that the entire network is robust to the capture or compromise of a cell member. Usama bin Laden described this strategy on a videotape that was found in a deserted al-Qaida house in Afghanistan. This type of covert social network suffers from reduced informational efficiency and information sharing, although the hijackers were clearly able to mitigate those deficiencies and ensure some level of communication and resource planning. (Courtesy and permission of Valdis Krebs.)

giance vector concept.

The Gallup Polls of the Middle East are aids to understanding the way Islamic nations feel toward other nations on a variety of topics. We use these polls in the TAP model for esti-

imating our aggregate allegiance vectors. Figure 5(b) shows an example of data from a Gallup poll.

### Social Bargaining and the Nash Demand Game. Many theories of

social interaction can be used for modeling ways in which beliefs, opinions, and values are communicated, shared, and modified during social meetings. We require a theory that allows us to model how allegiance values are transferred between interacting agents.

One theory we use is a social bargaining theory, established by H. Peyton Young (1998), that is based on the one-shot Nash demand game. During an interaction, each one of two agents places a bid for some portion of an abstract available “property,” where the bid is related to the estimated value of establishing a social contact with the other agent. Both parties get their demands if the sum of the bids is less than the total property available; that is, successful bargaining occurs if the two agents are not too greedy. If one agent’s bid is  $b_i$  and the other’s is  $b_j$ , then a successful bargaining process occurs if  $b_i + b_j \leq 1$ . The condition for a pure Nash equilibrium (where the agents are at their best-bid positions, and a change in bid by either agent will lower the overall payoff) is  $b_i + b_j = 1$ .

Agents learn from past interactions how to bid optimally in Nash demand games. Each agent retains a memory of its past  $m$  meetings. Agents associate some attributes of those agents that they interacted with in these past  $m$  meetings and judge how well they did in demand games with agents of similar attribute types. This information is then used as a basis for future bids with agents of similar type—in a form of social learning.

If the one-shot social bargaining game is successful, then allegiance values are transferred in an asymmetric, bidirectional manner between the two agents. The agent with the lower social status “absorbs” more of the other agent’s allegiance values, whereas the agent with the higher status absorbs fewer. In this way, social norms can emerge from the continued

interactions between agents. Note that this is but one heuristic from a set of “social rules” used by the agents.

We hope that this methodology will represent one component related to the spread of beliefs through social structures. Some Islamic organizations establish relief efforts and supply resources for citizens of low socioeconomic status. In addition to providing needed welfare, these actions also improve public support of and opinion about Islamic organizations. Some organizations rely upon a “bottom-up” approach to instill the populace with their doctrine (that is, Islamic Law—Shari’a), whereas others may choose a more revolutionary “top-down” approach by replacing a secular regime with a more Islamic ruling party. The social bargaining approach fits closer with the “bottoms-up” approach of belief spreading.

### Social Repression and Estimation of Social Grievance

I have briefly outlined how allegiances are transferred through agent populations by social learning. This learning is affected by who meets whom, how socioeconomic status is valued relative to allegiance adoption that occurs between agents, and past experiences of agents in meeting with agents that are considered similar. At the end of the simulation, however, we want to quantify relative measures of propensity toward terrorism across the agent population.

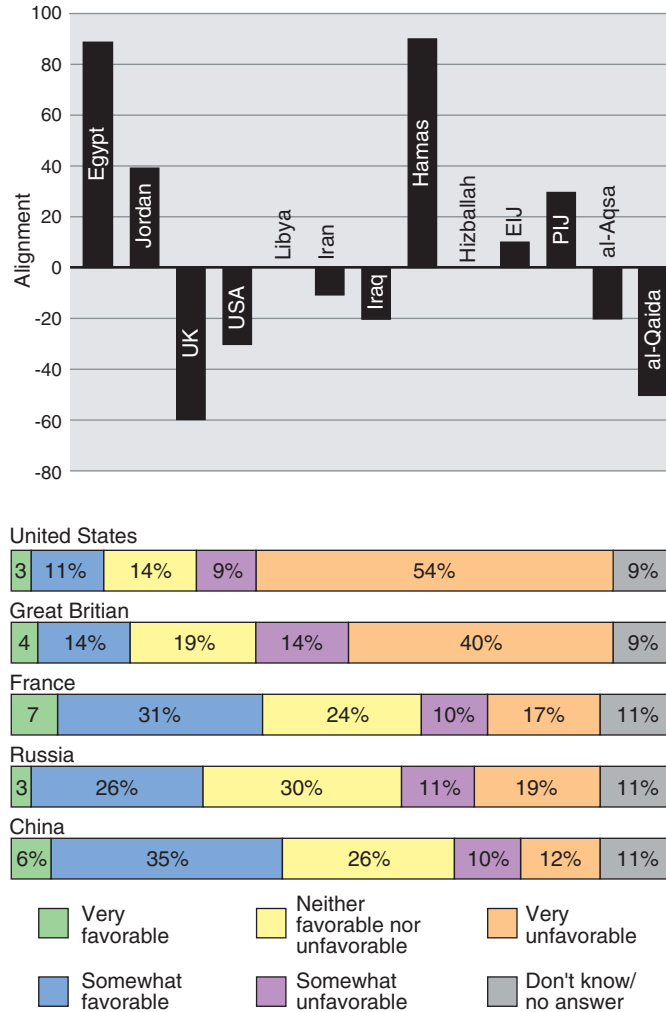
Social grievance is one of the “summary” metrics we use to determine when an agent, or a collection of agents in a region or organization, is considered to have a propensity for protest and therefore has a higher potential to become a terrorist. It is calculated from metrics that include an agent’s sense of social repression. Social grievance is directed against a

particular group or organization. In the simulation, repression stems from social disadvantage, inherited allegiances, cultural penetration, repression from the regime, and media influences.

We calculate a composite socioeconomic disadvantage metric based on the agent’s social ranking and the social ranking of the groups that the agent identifies with most. An agent may have high social status but may identify

with a social group that has low social status—thereby increasing the level of disadvantage felt for other members of the group.

Agents perceive repression by a corrupt regime. We quantify this level of repression by calculating the overlap between the agent’s and the regime’s allegiance vectors, weighted by a corruption factor. Corrupt governing regimes in the Middle East that are secular and aligned with Western



**Figure 5. Typical Allegiance Vector**

(a) All agents have allegiance vectors, which contain more than 100 elements. The values of individual vector elements indicate the feelings of an agent with regard to countries, organizations, and groups. These values change as agents interact with each other, thus simulating the process of social memes or “contagion” effects. (b) The results of Gallup Middle Eastern opinion polls help us quantify allegiance values in our model.

nations for commercial trade and military support, and who do not support the underlying social welfare of their populations, are considered to be an important root cause of militant Islamist terrorism against the West.

The infiltration of some aspects of Western culture, whether physically or media based, into Islamic regions is considered to be another source of repression to Islamic groups. In the TAP model, we estimate the contribution of cultural penetration to social repression as a weighted sum of the fraction of the local population that is nonindigenous and a media presence factor, which represents the relative amount of external (foreign) media influence in that region. The weighting function depends on the relative allegiance values—influences from cultures with high allegiance values are considered good. We estimate the media influence factor from “surrogate” data. For many Middle Eastern nations, we have measures of the percentage of households that have access to television, radio, and the Internet. Though not always the case, a household that has access to these information sources usually has access to the media of all other cultures. (In some Middle Eastern regions, the regime may censor external media influences.)

After calculating an agent’s social disadvantage and level of repression, we calculate agent  $A$ ’s time-dependent grievance  $G$  toward the social group  $M$  as

$$G = d(t) \times a(M, t) \times o(t) \times p(t) \times f(M),$$

where the first term is the socioeconomic disadvantage of  $A$ , the second term is the dislike  $A$  feels for group  $M$ , the third and fourth terms are the regime and cultural-penetration contributions to  $A$ ’s perceived oppression, and the last term is a measure of  $A$ ’s perception of group  $M$ ’s level of corruption. Social grievance of the agents

in each region is monitored in order to obtain a social grievance potential as the simulation progresses. An indicator of terrorist instability exists if agents with high levels of social grievance have access (through their social networks) to a terrorist organization with similar grievance targets. This represents increased public support and more probable recruitment for terrorist organizations.

## Summary

I have briefly described a complex agent-based model of a complex situation—terrorism associated with militant groups. My hope is to give a flavor of the methodology used in constructing agent-based simulations of socioeconomic systems and to show how this methodology is being applied to the challenges in developing a detailed understanding of the sociodynamics of militant Islamist terrorism.

I see agent-based simulations as computational experiments that convey a great deal of scenario information in a timely, efficient, and safe manner. Examining the path-dependent time evolution of a particularly interesting simulation result (a virtual history) in replay mode allows us to analyze the “how” and “why” of agent behavior. If the TAP agent-based model can supplement policy-making in the turbulent clash between the West and Islamist radicals and can help policy makers visualize and understand important yet currently unknown interrelationships, then this work will be a success. Although agent-based simulation can help us gain insight into complex system behavior, my longer-term hope is that the “tit-for-tat” pattern of violence in the Middle East will be quenched by taking an honest look at both sides of the issues at hand through peaceful negotiation and mutual respect.

## Acknowledgments

I thank the DTRA Advanced Systems and Concepts Office for their kind support and valuable expertise in constructing the TAP model. I also thank Gary Ackerman, Anjali Bhattacharjee, and many others at the Monterey Institute of International Studies for sharing their valuable expertise on terrorism and the Middle East; Charles Macal and Michael North of Argonne National Laboratory for their expert modeling advice; and the Santa Fe Institute for its support and useful input; Amir Mohagheghi, Faraj Ghanbari, Jessica Turnley, and Nancy Kay Hayden of Sandia National Laboratories for their useful discussions. I thank Waleed El-Ansary of the Islamic Research Institute for his useful insights into Islamic economics and game-theory. Finally, I thank the Los Alamos team of Joe Holland, Karl Lautenschlager, Merle Lefkoff, Ayla Matanock, Dennis Powell, Brian Reardon, David Sharp, and Zoltan Toroczka for their important contributions to this work. ■

*For further information, contact Edward MacKerrow (505) 665-3491 (mackerrow@lanl.gov).*



# The Los Alamos Center for Homeland Security

*Thomas W. Meyer, J. Wiley Davidson, I. Gary Resnick, Ray C. Gordon III, Brian W. Bush, Cetin Unal, Gaspar L. Toole, L. Jonathan Dowell, and Sara C. Scott*

**O**n November 25, 2002, the president of the United States signed a bill creating the Department of Homeland Security (DHS), thus initiating the most significant transformation of the U.S. government since 1947, when Harry S. Truman merged the various branches of the U.S. armed forces into the Department of Defense. Planning for the DHS began in the aftermath of September 11, 2001, and was first codified in the July 2002 National Strategy for Homeland Security released by the White House. When it came into existence on January 24, 2003, the DHS consisted of 170,000 employees from 22 agencies and had an annual budget of \$38 billion.

The DHS has three primary missions: prevent terrorist attacks within the United States, reduce America's vulnerability to terrorism, and minimize the damage from potential attacks and natural disasters. Los Alamos National Laboratory has a rich history in developing technologies that can be brought to bear on these DHS mission areas and over the past few years has pursued activities that paralleled the evolution of the new department. As the national strategy for homeland security was evolving, Los Alamos had already decided to create the Center for Homeland

Security (CHS), which was formally established in September 2002.

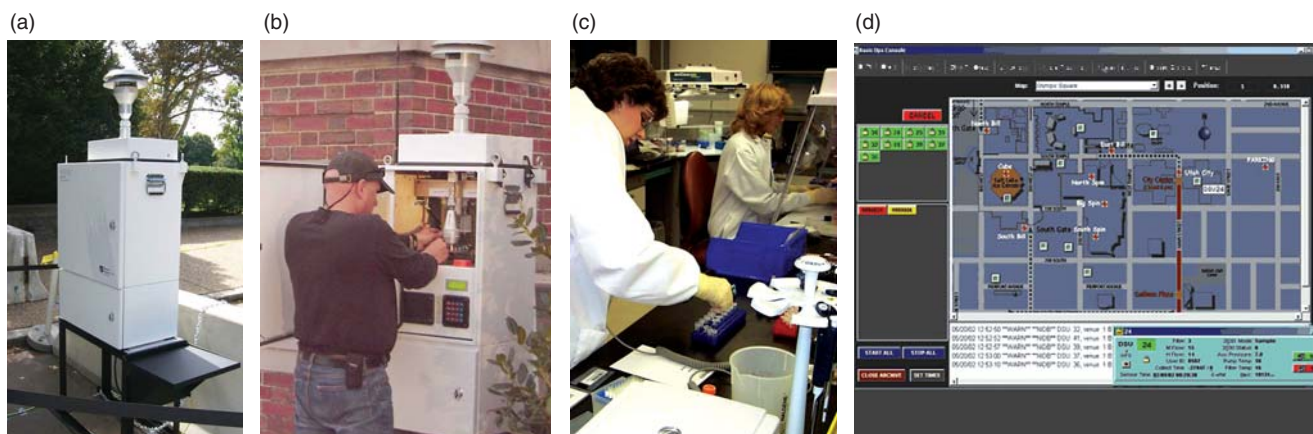
The CHS is responsible and accountable for all Los Alamos programs for the DHS. It applies the Laboratory's science and technology capabilities toward homeland security and seeks to provide solutions. It also helps streamline operations, since it is the sole point of contact to the DHS and other agencies involved in homeland security. In addition, the CHS provides the opportunity to leverage institutional relationships within New Mexico at the state, regional, and local levels.

The CHS was established as a small program office that would oversee three critical focus areas: chemical and biological; nuclear and radiological; and systems analysis, integration, and infrastructure. Technologies that Los Alamos had been developing for decades under sponsorship of the Department of Energy (DOE), the Department of Defense, and other government agencies were evaluated and their associated programs transferred into one of the three focus areas. This action was taken in anticipation of the transfer that was mandated by the creation of the DHS. Although each focus area encompasses many research efforts that can address the challenges of homeland security, we highlight only a few in the sections that follow.

## **Technology against Bioterrorism: BASIS**

A bioterrorist attack with aerosolized biological threat agents could have a catastrophic impact in an urban environment. In collaboration with scientists and engineers at Lawrence Livermore National Laboratory, we have developed a wide-area environmental monitoring system called the Biological Aerosol Sentry and Information System, or BASIS, which will provide early warning of biological attack. Early detection and rapid response is crucial because the identification, treatment, and possible isolation of exposed individuals are most effective if they occur within the first few hours following a biological attack. Unfortunately, awareness of an attack typically comes only after individuals begin displaying symptoms, when it is too late to save a large percentage of those exposed. BASIS can determine the time and place of a bioattack within 12 hours, well before the onset of most symptoms and in sufficient time to warn public health and safety organizations.

Figure 1 provides an overview of BASIS. Distributed sampling units (DSUs) that sample the air are located at specific sites in a city or in a mobile unit. A suction pump in the DSU draws outside air through filters



**Figure 1. The Biological Aerosol Sentry and Information System**

**BASIS** is a suite of integrated technologies developed to provide timely detection, identification, and characterization of bioagent aerosol releases. (a) DSUs continuously collect aerosol samples in and around selected sites. (b) The Sample Management System helps to coordinate the periodic retrieval and delivery of the samples and is responsible for maintaining and archiving information. Here, a support team member scans an aerosol filter holder with a laser barcode reader. (c) Samples are analyzed at the RFL (or possibly at

existing local laboratories), where high-sensitivity, high-specificity bioassays provide bioagent detection and identification. Samples are saved and inventoried to provide opportunities to confirm and reanalyze the findings. (d) All operations, including sample management and testing in the laboratory, are monitored at the command console of the BASIS Operations Center. The operations center has links to external agencies, and in the event of an attack, it can initiate and help coordinate a rapid response.

that capture any aerosolized threat agents. Two filter systems operate simultaneously: a “holder,” which typically collects samples for four hours, and the “magazine.” The latter contains several filters, each of which typically collects samples for one hour. The filters from the DSU are periodically retrieved and delivered to a relocatable laboratory, where they are analyzed for multiple biothreat agents by identification assays based on polymerase chain reactions. (Similar assays are described in the article “Reducing the Biological Threat” on page 168.) Only if a holder filter tests positive for an agent are the magazine filters tested, a procedure that enables prompt biothreat detection while minimizing the number of assays.

A command and control center oversees the collection and analysis of the samples and maintains communication links to federal, state, and local agencies. In the event a biothreat agent is detected, appropriate public health and safety organizations can be alerted in time to initiate effective medical treatment and other responses.

**Successful Demonstrations.** BASIS proved its operational capabilities at the 2002 Winter Olympic Games in Salt Lake City. The system went into full operation on January 21 and ran continuously until February 26, when it was shut down. Sixteen DSUs were deployed at key indoor and outdoor locations in Salt Lake City and Park City. The sampling was performed 24 hours a day, except at ice skating venues when the coverage started one hour before the beginning and finished one hour after the end of an event. Each sample run lasted four hours, except during the night when the run was extended to eight hours. During the Winter Olympics, the Sample Management System coordinated the loading, replenishing, and tracking of approximately 10,000 filter cassettes.

A relocatable field laboratory (RFL) was set up at the Utah Department of Health. The RFL ran two production lines that analyzed samples for threat agents and operated 20 hours a day, processing approximately 2100 samples during the five weeks of operation.

The BASIS Operations Center operated continuously. Each DSU was in constant contact with the Operations Center, which monitored the airflow rates and particulate concentrations. The replenishment operations were confirmed, and the barcodes on samples were recorded. If a sampling unit had problems, service teams were sent out immediately. The Operations Center was also in contact with the Sample Management System and the laboratory. All sample transactions were recorded in a master database for forensic purposes.

By all measures, the performance of BASIS was superb. During the Winter Olympics the overall time to detect was a minimum of two hours and a maximum of eight hours. For overnight sampling, the time to detect was increased by four hours because of the extended sampling time. The level of detection has been studied in field tests with surrogate agents and with live agents at Dugway Proving Ground. The system is proving to be both sensitive and specific.

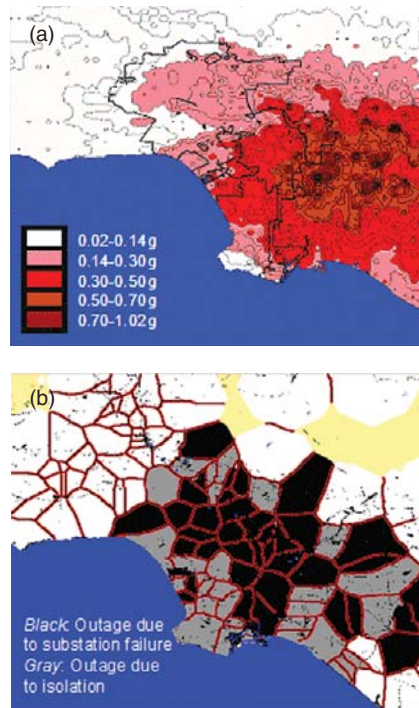
In collaboration with the

Department of Defense, Lawrence Livermore and Sandia National Laboratories, and the New Mexico State Department of Health, we established an operational systems-level test bed in Albuquerque, New Mexico. BASIS was deployed in the test bed and expanded to include autonomous sampling units at the Albuquerque airport, and in December 2002 a demonstration was conducted. Again, BASIS performed successfully. After this demonstration—in collaboration with the DHS, the Environmental Protection Agency, and the Centers for Disease Control and Prevention Laboratory Response Network—we deployed BASIS to numerous urban centers in the United States as part of Project BioWatch.

### Systems Analysis: Modeling the Nation's Energy Infrastructure

Since the middle of the 1980s, researchers at Los Alamos have modeled and simulated energy transmission networks, with a long-standing focus on electric power systems. During this period, we have developed an extensive set of databases, analysis tools, and science and engineering expertise to answer a broad range of questions that are important to decision makers; various local, state, and federal agencies; and the nation as a whole. Our work is typically done in collaboration and coordination with other national laboratories, industry organizations, and government agencies.

**Electric Power Grid Modeling.** Much of our analysis of the electric power industry has focused on possible outage events that could interrupt the reliable supply of electric power. Inherent attributes of the electric supply system, in addition to natural or man-made breakdowns, are possible sources of disturbances in the power



**Figure 2. Simulating the Effects of an Earthquake**

(a) The figure shows the calculated peak ground acceleration for an earthquake of magnitude 6.75 on the Richter scale, occurring at the Elysian Park fault under downtown Los Angeles. (b) The quake damages equipment in some areas (black), creating power outages. Neighboring areas (gray) subsequently become disconnected from the grid and also lose power.

system. We typically construct detailed models of the utilities of interest and then analyze the models using state-of-the-art power-flow simulation tools. Thus, we are able to identify the service and outage areas, estimate how long the outage lasts, identify critical system components, and recommend restoration strategies or mitigation options. In general, our goal is to evaluate the performance of the system and determine the electric industry's ability to supply sufficient electric power to its customers, given all the demands and energy requirements and taking into account the breakdown of system elements.

For example, Figure 2 illustrates

the effect a large earthquake under downtown Los Angeles might have on the electric power grid. The analysis starts with the evaluation of ground motion and acceleration. We then estimate the damage to electric power substations using "fragility curves" that approximate the probability of a certain level of damage to the equipment, based on the previously estimated ground motion. Those estimates of equipment damage provide a basis for simulations of the earthquake's effect on the overall operation of the power system. Using a Los Alamos-developed cellular automaton algorithm that calculates the area that an electric power substation can serve, we predict the geographic extent of the power outages that might occur.

We have also examined how deregulation and mergers in the electric power industry have affected the reliability of the power grid and performed simulations to understand how the structure of a deregulated energy market influences the day-to-day operation of the power system. Aside from the technical challenges, this problem is politically complex because it involves differences in state and federal guidelines or policies, differences among state deregulation policies within the same geographic region, planned new regional transmission organizations, and new independent system operators. In another project, we have undertaken an extensive series of case studies to document the robustness of the energy supply to important government facilities.

### Interdependent Infrastructures.

Over the past five years, Los Alamos work on the electric power grid has expanded into the broader area of energy transmission infrastructures in general. We now model natural gas pipeline networks and petroleum liquid networks and have plans to model the coal infrastructure within the United States. These energy networks typically

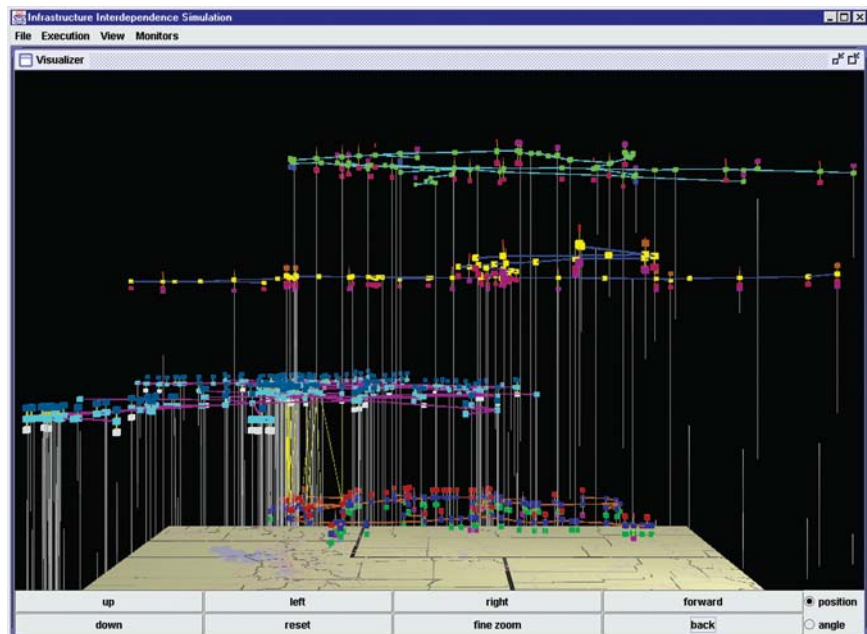


## Countering Nuclear and Radiological Threats

Nuclear and radiological threats exist now, and there is concern that more will occur over the coming decades. The quantity of nuclear material is increasing. Worldwide weapons information is available in the public forum, and terrorist organizations—some of which are well funded—have stated their interest in obtaining nuclear and radiological devices. (The United Nations reports that 130 terrorists groups may be capable of developing a homemade atomic bomb.) Although preventing threats is optimal, we must also be prepared to detect and respond to threats that develop and evolve to crises.

The CHS works with customers and end-users to develop and implement technologies and approaches that affect all aspects of nuclear and radiological terrorist threats. (See Figure 4 for an overview of our focused efforts.) Overviews of five of our thrust areas follow.

**Prevention.** Safeguarding fissile and radioactive materials is important in preventing nuclear terrorism. The Laboratory's current safeguards mission is in part to assist with the global control of nuclear material and expertise that is accomplished through several venues: the implementation of treaties and agreements, worldwide export control, research and development, and a new effort to counter nuclear terrorism. Since 1966, Los Alamos has had active programs to develop methods to track, secure, and account for fissile material, including the Material Protection, Control, and Accounting (MPC&A) Program, and the nuclear safeguards programs. We also provide technical support for actively monitoring the export of sensitive equipment and raw materials and deploying capabilities for detecting the clandestine production of



**Figure 3. Analysis of Interdependent Energy Networks**

The IEISS is a set of software tools that helps us analyze interdependent energy networks. This screen capture shows an abstract three-dimensional visualization of major energy networks laid over a map of Utah. The network of crude oil pipelines is displayed in the upper layer, then the petroleum product pipelines, the electric power transmission lines, and finally the natural gas pipelines. The vertical lines identify interdependencies between the systems.

depend on each other to deliver their product. A gas-fired electric generating plant, for instance, requires a steady supply of natural gas, and the natural gas pipelines may possess electrically powered compressors to maintain sufficient pressure. Because traditional tools that modeled single infrastructures were severely limited when applied to such interdependent networks, we have developed new tools to address the earlier shortcomings.

The Interdependent Energy Infrastructure Simulation System (IEISS) is a suite of analysis software tools developed by Los Alamos in collaboration with Argonne National Laboratory. We intend to develop a comprehensive simulation of the nation's interdependent energy infrastructures that will include all components and couplings, in a manner far beyond what could be done previously. The IEISS will help us understand

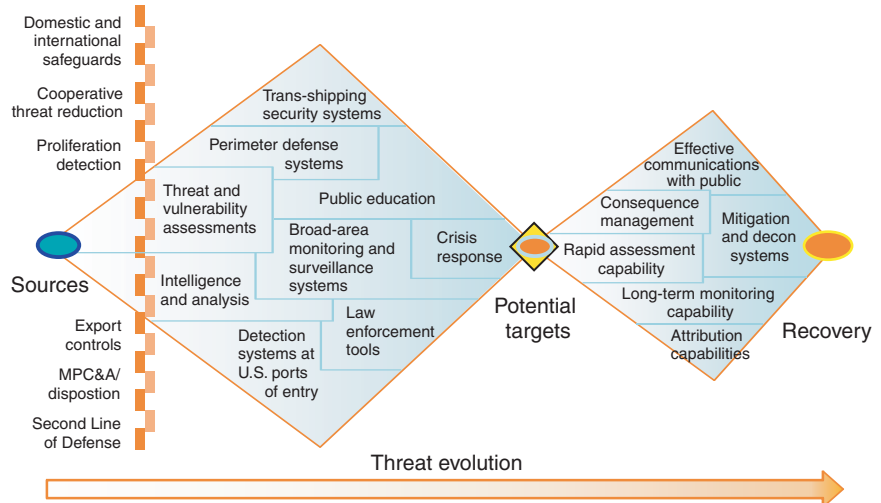
in depth the normal operations of the infrastructures and help us develop insight into disrupted operations. In addition, it allows us to assess the technical, economic, and national security implications of the interdependencies. Figure 3 is a screen capture from a prototype of the IEISS analysis tool that was used in preparation for the 2002 Winter Olympic Games in Salt Lake City.

In addition to identifying critical components and vulnerabilities in coupled infrastructures, we hope to use the IEISS to assess how future investments in the systems might affect quality of service and to perform integrated cost-benefit studies, evaluate the effect of regulatory policies, and aid in decision making during crises. Additionally, IEISS is a research tool for investigating fundamental issues related to real-life, complex networks.

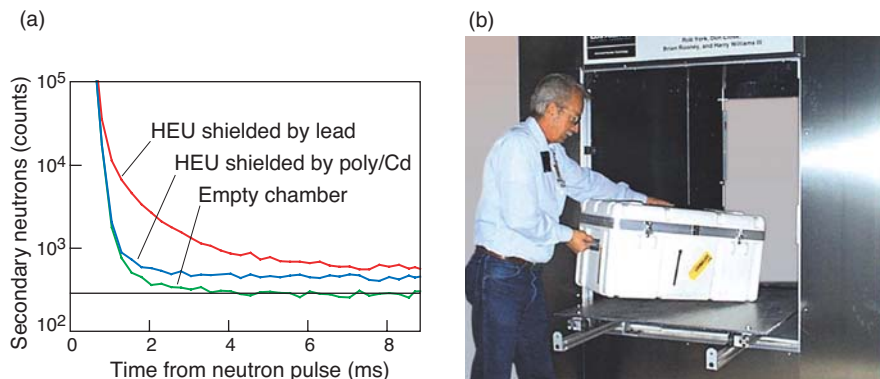
nuclear materials. Additionally, in the mid-to-late 1990s, the safeguards program assisted in the down-blending of large amounts of Russian weapons-grade uranium, wherein the highly enriched material was diluted to produce a mixture that could not be used in nuclear weapons. More recently, the international safeguards program, motivated by experience in Iraq in the early 1990s, has developed technologies that support additional protocols for detecting undeclared nuclear activity.

Our nuclear safeguards programs also have an extensive training component. All IAEA inspectors have been trained by Los Alamos, as have personnel from the National Nuclear Security Administration and the Nuclear Regulatory Commission. We have similarly trained state authorities in effective management of state systems for nuclear material accounting. In the near future, the program will expand to include the development of threat analysis methodologies for subnational units, the development of new “proliferation resistant” fuel cycles, new safeguards approaches for future large-scale nuclear facilities, and the creation of new technologies for mitigation and detection of nuclear noncompliance.

**Monitoring and Assessing.** We spend considerable time at Los Alamos monitoring nuclear programs worldwide. After the disintegration of the Soviet Union in 1991, we became very concerned about the fate of their nuclear weapons, especially those that were left within the borders of the newly formed nations of Kazakhstan, Belarus, and the Ukraine. Over the course of several years, and with the cooperation of the countries involved, we participated in an unprecedented reversal of nuclear proliferation and helped denuclearize the three new nations. Their weapons were destroyed or returned to Russia. We



**Figure 4. The Los Alamos Nuclear and Radiological Systems Strategy** We contribute to multiple programs (left) that are geared towards monitoring and controlling nuclear and radiological materials and technology. Protection at the source is one key to preventing nuclear and radiological threats. If a threat were to develop, then a different set of capabilities would address the problem, initially with a broad range of potential responses, but focusing to more specific actions when a specific threat is identified. Should an incident occur, responsibility evolves to programs that are concerned with immediate postevent mitigation and with decontamination, attribution, and recovery.



**Figure 5. Active Interrogation of Packages** (a) Pulses of electromagnetic or neutron radiation induce highly enriched uranium (HEU) or plutonium to produce characteristic emissions. This graph shows that even shielded material is discernable. (b) Los Alamos has developed and demonstrated this active detection system for monitoring luggage and packages; it could readily be extended to monitor air cargo containers. Gram quantities of shielded HEU can be detected within seconds.

then assisted Russia in securing these and other weapons. To the best of our knowledge, all nuclear weapons in the Russian stockpile are accounted for and secure.

We also monitor worldwide nuclear smuggling on a continuing

basis. With ties to the intelligence community and to national and foreign law enforcement communities, we combine information as it is made available with our own understanding of smugglers. Our immediate goal is to assess current terrorist capabilities

so that the United States can take immediate action, but we also try to estimate future capabilities. In addition, working with other government agencies, we field equipment and protective measures and help establish security measures and procedures to keep ahead of the terrorists.

**Detection.** The United Nations reported that attempts to smuggle radioactive material have doubled over the past five years. Since 1993, IAEA's database has recorded 550 incidents of illicit trafficking in nuclear materials. (More than 370 of those incidents have been confirmed.) For these and other reasons, Los Alamos has been developing and installing radiation monitors at Russian and U.S. border crossings.

The technology to monitor for radioactive materials is either passive or active. In passive monitoring, one looks for neutron or gamma radiations that are emitted naturally from the radioactive material. This is a mature technology that began at the Laboratory 25 years ago under the direction of Paul Fehlau. Passive monitors have been installed in many airports to scan baggage and people, and whole-vehicle and train monitors have been placed in strategic positions around the world, including the Russian–North Korean border.

The downside of a passive monitoring system is that fissile materials can be shielded, reducing the emissions reaching the detectors. Active detectors emit pulses of electromagnetic waves or neutrons that induce detectable electromagnetic and neutron emissions from highly enriched uranium and plutonium, even when the materials are shielded (see Figure 5). Active detectors are more capable of detecting small amounts of shielded weapons-grade uranium. The Laboratory has already developed and fielded prototypes of these next-generation portal monitors.

**Response.** Los Alamos plays an essential role in responding to nuclear and radiological threats. We are active participants in the Nuclear Emergency Support Team (NEST), a DOE–National Nuclear Security Administration umbrella program that includes the Joint Technical Operations Team, which provides advice on how to “render safe” terrorist nuclear devices and provides nuclear safety assessments for the safe disposition of devices; the Accident Response Group, which is responsible for incidents involving U.S. nuclear weapons; and the Radiological Assistance Program (RAP), which upon request, provides assistance to local, state, tribal, and federal government officials or to private individuals. The RAP would respond, for example, to a radiation alarm at a landfill, an abandoned radiation source, or a transportation accident involving radiological materials. Another NEST program in which we participate is TRIAGE, which provides technical assistance to front-line personnel (such as customs officers) should they need help in evaluating data from fielded radiation monitors.

**Recovery.** Consequence management (CM) focuses on rapid and prepared responses to the tragic reality of an executed terrorist attack. CM preparedness provides direct support to first responders; playbooks for directing response activities; readily available public education information; rapid postevent simulation capabilities; and triage, mitigation, and decontamination technologies. Los Alamos emergency response and science experts are making significant contributions to increase our nation's CM preparedness. Advanced mitigation and decontamination technologies are required to prevent or minimize resuspension of radionuclides, to protect people and the environment, and to provide risk-based strategies for postevent cleanup. The

Laboratory is addressing development challenges, including understanding the molecular-based interactions between radionuclides and building materials; establishing performance criteria; developing test and evaluation protocols; and creating new technologies to clean buildings, ground coverings, water supplies, and runoff from first responder activities.

## Concluding Remarks

Since September 11, 2001, there has been a tremendous need to confront the international problem of terrorism. In line with our Manhattan Project roots, the Laboratory and the CHS are accepting that responsibility by again creating new technologies that aid in the nation's defense. For additional information about the CHS and our ongoing programs, please visit [www.lanl.gov/orgs/chs/](http://www.lanl.gov/orgs/chs/). ■

*For further information, contact Wiley Davidson (505) 665-8031 (wdavidson@lanl.gov).*

# Strategic Investments

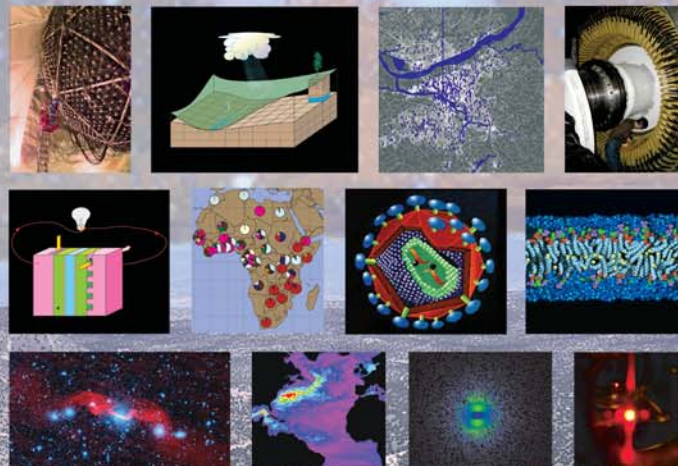
At Los Alamos, we make continual investments in fundamental science and innovative technologies as well as in development of our research staff. Those investments are strategic because they provide the intellectual foundations for our national security missions, present and future.

The images on the facing page represent a sampling of projects in fundamental science and advanced technologies.

(Top row, left to right): Studies of the solar neutrino puzzle with the SNO neutrino detector; simulations of regional watersheds to aid water management; simulations of epidemics to aid first responders; and research on materials using ultra-high magnetic fields from the country's largest electric generator.

(Middle row, left to right): Development of the PEM fuel cell for energy sustainability; development of the HIV DNA database, a mainstay in the fight against AIDS; studies of HIV dynamics in the human body; and simulations of the insertion and folding of a protein into membranes.

(Bottom row, left to right): Investigation of the magnetic fields observed in astrophysical radio jets; high-resolution modeling of the ocean circulation; three-dimensional simulations of supernova dynamics; and development of tunable lasers from nanocrystal quantum dots.



# Strategic Research at Los Alamos

*Rajan Gupta and David E. Watkins*

Los Alamos National Laboratory is a scientific institution whose primary mission is to be a steward of the nation's nuclear deterrent. In a broader context, the Laboratory's mission is to preserve the security and safety of the United States against present and future threats. To anticipate future threats and develop the necessary ideas and technologies to detect, foil, and mitigate possible attacks, we must be working at the cutting edge of research in many branches of engineering and science, and we must simultaneously integrate a wide range of research advances from academia and industry.

The long-term success of any scientific organization is tied to its ability to recruit and retain exceptional people and to foster collaboration and meaningful relationships with the finest institutions worldwide. In addition, at Los Alamos and other mission-oriented laboratories, an environment must be maintained in which the creativity of the staff is readily tapped in order to implement those missions. Sustaining that kind of environment is a formidable task. It has been made more complicated because emerging threats—global terrorism, nuclear proliferation, poverty, disease, diminishing fossil-fuel and water resources, stressed ecosystems—require that the Laboratory respond

from a broad base of scientific expertise and develop effective responses in a timely fashion.

Unlike many research environments, ours must balance the freedom to explore new ideas with a strong communal commitment to meeting national security challenges and to making sacrifices when necessary. This balance requires the presence of scientific leaders who can inspire others to contribute innovative ideas and who can lead the integration of those ideas into practical solutions. Not only must we divide our efforts between basic research and applied research and development activities, but we must also recognize and seek out synergistic research opportunities in which progress in one field yields greater understanding in another. Finally, there must be a deep recognition that the evolution of ideas is rarely predictable and that the Lab must position itself to encourage the creation and exploitation of new ideas to meet future challenges.

After World War II, key decision makers recognized the value of providing scientific organizations with the flexibility necessary to pursue innovative ideas. While serving as Army chief of staff, General Eisenhower wrote, "Scientists and industrialists must be given the greatest possible freedom to carry out their research.

The fullest utilization by the Army of the civilian resources of the nation cannot be procured by prescribing the military characteristics and requirements of certain types of equipment. Scientists and industrialists are more likely to make new and unsuspected contributions to the development of the Army if detailed directions are held to a minimum . . ." This kind of thinking is reflected again in the Atomic Energy Act of 1954: "The commission is directed to exercise its powers in such manner as to insure the continued conduct of research and development and training activities in the fields specified . . ."

Currently, within the Department of Energy (DOE) structure, the Laboratory-Directed Research and Development (LDRD) program provides resources for discretionary research. Although this particular formal structure is young compared with the 60-year history of the Lab, the origins of the program go back to the beginning of the weapons program. Indeed, some have argued that the initial work on spherical implosion represents the first LDRD-like effort at Los Alamos. In late April 1943, Seth Neddermeyer proposed that a three-dimensional implosion would be a potentially more effective means of assembling a supercritical mass than the one-dimensional "gun assembly,"

which was the baseline design concept. Neddermeyer's concept was initially considered unnecessary and outside the mainstream of work. Although the spontaneous rate of plutonium-239 fission is twice that of uranium-235, the difference did not appear to warrant a different design. Nevertheless, Oppenheimer decided to fund the work on spherical implosion to keep the option open. It was not until the data came in on the reactor-produced plutonium, which contained enough plutonium-240 to significantly increase the spontaneous fission rate, that the work on spherical implosion carried out by Neddermeyer and Kistiakowsky was recognized as essential. Today, the flexibility afforded through the LDRD program continues to provide the nation with science and technology critical to our defense and security.

The synergy between the nuclear weapons program and basic science is exemplified by the work of Nobel Laureate Fred Reines. During the war, Reines worked on many different aspects of nuclear weapons design. After the war, he became an expert on nuclear weapons effects and played a lead role in nuclear weapons testing. Toward 1948, Reines wanted to return to basic science. Carson Mark, his division director actively encouraged this transition and gave Reines the freedom to "sit and think." This he did for almost a year, during which time he decided to search for the elusive neutrino, a neutral particle with little or no mass, whose existence was postulated on the basis of the fundamental principle of energy conservation. If theory was correct, this particle should be produced in copious quantities during a nuclear explosion.

Under the nurturing eye of Carson Mark, Fred pulled together his vast experience with detection of different forms of radiation, his abilities to do big science, and the technical capabilities of Los Alamos to build a detector

*Continued on page 204*

## Resolving the Solar Neutrino Problem

*Andrew Hime*

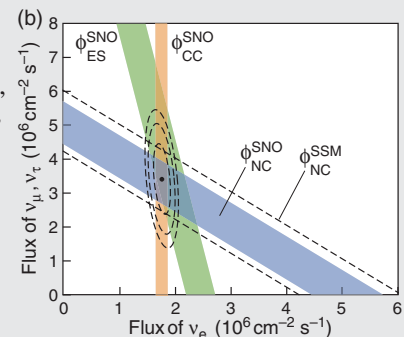
Since the seminal discovery of the neutrino by Los Alamos researchers Clyde Cowan and Fred Reines in the late 1950s, there has been a continuous effort by Laboratory scientists to study neutrinos and their interactions. Central to that study is the question of whether the different neutrino flavors—electron, muon, and tau—have mass, a question that has been answered decisively with recent data from the Sudbury Neutrino Observatory (SNO).

SNO was built to resolve the long-standing solar neutrino problem, wherein the measured flux of solar neutrinos reaching Earth is significantly lower than predicted. But all previous experiments had been sensitive to only electron neutrinos. Whereas those are the only neutrinos that can be created by the nuclear fusion reactions powering the Sun, the other flavors can be "produced" if neutrinos have mass. Through the process of flavor mixing, massive electron neutrinos from the Sun can "transform" into muon and tau neutrinos as they travel to Earth. This process would explain why the measured solar electron neutrino flux is lower than predicted by the standard solar model.



A multinational decade-long effort, SNO was designed to detect all neutrino flavors. Los Alamos was involved in all aspects of the project, including detector construction, commissioning, simulation, and calibration, as well as data analysis and scientific management. The primary detector contains 1000 tonnes of ultrapure heavy water in a large "bottle," shown in (a). The deuterium in the water can interact with high-energy electron neutrinos through charged-current (CC) weak interactions and with all neutrino flavors through neutral-current (NC) weak interactions. All neutrino flavors also undergo elastic scattering (ES) with electrons, but the reaction is sensitive mostly to electron neutrinos.

By enabling a direct comparison of the CC, NC, and ES reaction rates, SNO could determine if the neutrinos from the Sun are a mix of electron, muon, and tau neutrinos. As seen in (b), the best fit to the data (dotted circles are confidence limits) indicates that two-thirds of the electron neutrinos born in the Sun transform into muon and/or tau neutrinos. ( $\Phi_{NC}^{SSM}$  is the NC flux predicted by the standard solar model.) Together with data from other experiments, these results demonstrate that flavor mixing almost certainly resolves the solar neutrino problem and that neutrinos have mass. Moreover, the total flux of neutrinos measured at SNO agrees with predictions and our basic knowledge of how the Sun shines.



CC:  $\nu_e + d \rightarrow p + p + e^- - 1.44 \text{ MeV}$   
 NC:  $\nu_x + d \rightarrow p + n + \nu_x - 2.22 \text{ MeV}$   
 ES:  $\nu_x + e^- \rightarrow \nu_x + e^-$

## A New Source of Ultracold Neutrons

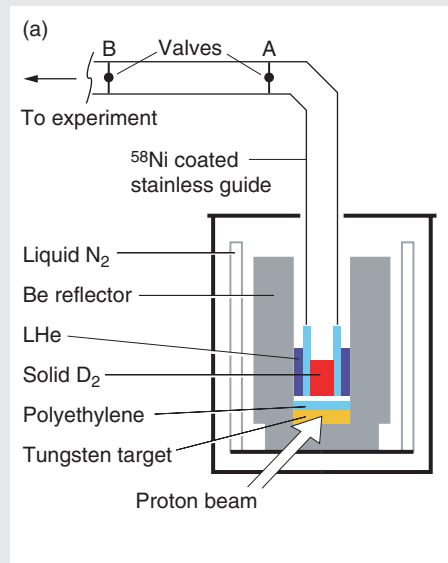
*Chen-Yu Liu, Steve K. Lamoreaux, Thomas J. Bowles, and Christopher Morris*

A neutron will normally diffuse right through materials such as steel or lead, but if the neutron's energy is exceedingly low, it will instead be reflected by those (or other) materials. We can now produce neutrons with kinetic energies less than about 300 nano-electron volts. When placed in a "bottle" of the right material, these ultracold neutrons (UCNs) become trapped, enabling us to collect them and to make a high-density UCN source.

Such a source can offer orders-of-magnitude improvements to experiments that make precise measurements of neutron properties. For example, Los Alamos is vigorously pursuing an experiment to measure the angular correlation between a neutron's spin and the momentum of the electron that emerges when the neutron decays. The experiment should provide a precise measure of the ratio of the vector and axial-vector coupling constants in the electroweak interaction in the Standard Model, and so can be used to test the unitarity of the CKM matrix (the matrix that rotates the complete set of quark mass states into the complete set of quark weak states). Another experiment aims at improving the limit set on the as-yet-unobserved neutron electric dipole moment (EDM), the existence of which would violate parity and time reversal symmetry. The neutron EDM, therefore, serves as a critical test of fundamental particle theories that are extensions to the Standard Model.

We have recently conducted experiments to understand and characterize the performance of solid deuterium as a so-called superthermal UCN source. As seen in graphic (a), a pulse of high-energy protons from the Los Alamos Neutron Science Center is

directed to a tungsten target that sits inside a liquid nitrogen dewar. High-energy spallation neutrons exit the target, lose most of their energy in a layer of polyethylene, and enter the liquid helium dewar that holds the solid deuterium. A neutron can collide with a deuterium molecule and resonantly transfer essentially all of its kinetic energy to the deuterium lattice. By this superthermal process, a relatively large fraction of the incoming neutrons become ultracold. They then diffuse to the UCN bottle region, where the sequenced closing of valves B and A trap them.



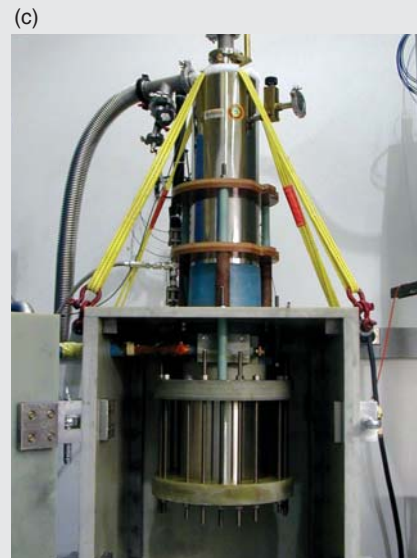
We have demonstrated a prototype, solid deuterium source, shown in (b), that produced about 120 UCNs per cubic centimeter inside the bottle region, a much higher value than can be provided by other facilities. A source with potentially six to seven times that UCN density is being engineered for the angular correlation experiment mentioned above. We are also investigating solid oxygen as a candidate for the next generation superthermal UCN source. Our preliminary calculations indicate that this new material might yield a UCN flux output 250 times greater than that achieved with solid deuterium.

## The National High Magnetic Field Laboratory—A User Facility

*Alex H. Lacerda and Gregory S. Boebinger*

The National High Magnetic Field Laboratory (NHMFL) at Los Alamos has emerged as one of the leading pulsed-magnetic-field research centers in the world, hosting an international user program that attracts around 150 visiting scientists a year to Los Alamos. Part of the NHMFL consortium (which has other facilities at Florida State University and the University of Florida and is jointly funded by the National Science Foundation, the State of Florida, and the Department of Energy), the NHMFL—Los Alamos is a multipurpose facility, powered not only by a 1.4-billion volt-amperes power generator (the country's largest) shown in (a) but also by the intellectual energy of its personnel and user community.

In (b), researchers are shown examining data that come from magnets in cells 3 and 4 of the facility, each of which contains a 60-tesla “short-pulse” magnet that provides a 25-millisecond magnetic pulse. These magnets are the workhorses of the NHMFL user program and are used for magnetotransport, magnetization, and radio-frequency conductivity experiments. Researchers in the background are clustered around cell 2, which contains a 50-tesla “midpulse” magnet, shown in (c), that is primarily used for optics experiments, such as photoluminescence. Another frequently used magnet is the “large-bore” 50-tesla magnet that is used for measurements requiring some “elbow room,” such as angle-



dependent magnetization and magnetoresistance measurements, or pulsed-echo ultrasound and gigahertz spectroscopies.

When energized at peak fields, the different magnets have magnetic energies between 0.5 and 100 megajoules. As a megajoule is roughly the energy equivalent of two sticks of dynamite, research involving high magnetic fields is, in part, an exercise in applied metal fatigue. With pressures in the magnets approaching 1.4 gigapascals (200,000 pounds per square inch), not only does copper wire stretch and rupture, but even garden-variety steels fracture catastrophically. To combat the metal fatigue, the Los Alamos engineers and metallurgists working with the NHMFL have helped to develop a variety of nanostructured conductors and exotic reinforcing materials with greatly enhanced mechanical strength.

The technical expertise needed to maintain the NHMFL magnets is an asset, and it is frequently applied to problems involving pulsed-power, electromagnetic modeling, and other projects that serve the Laboratory's broader mission. The pulsed magnets themselves provide a unique scientific platform for developing and testing experimental techniques in a readily reproducible transient environment. Finally, NHMFL scientists actively collaborate with other Los Alamos researchers on fundamentally important missions ranging from nanoscience to plutonium.



*Continued from page 201*

that would “see” the neutrino. In 1956, Reines and his team discovered the neutrino using the Savannah River nuclear reactor as the neutrino source. That work opened up the field of experimental neutrino physics, which is still actively pursued at Los Alamos. Recently, Laboratory scientists played a leading role in an international collaboration that presented the first evidence that neutrinos oscillate from one type to another. Such conversion of electron-type to muon-type neutrinos is now believed to solve the solar neutrino riddle—explaining why the number of electron neutrinos detected at the earth’s surface is far smaller than the number that solar models predict (see the box “Resolving the Solar Neutrino Problem” on the previous page).

To highlight how ideas in science and technology merge into major programs, we look back to biological research in the 1960s. The principle of flow cytometry—a flow method in which cells are rapidly interrogated one cell at a time—was invented through Laboratory-directed research, and the first device was reported by Mack Fulwyler in 1965 in *Science* (Volume 150). A few years later, Los Alamos flow cytometers could sort cells by DNA content, enabling scientists to study the effects of exposure to radiation or to carcinogenic chemicals, issues relevant to worker safety within the DOE. By 1985, the Laboratory and Lawrence Livermore National Laboratory were sorting the different chromosomes in the human genome with 95 percent accuracy and supplying libraries of cloned fragments for each chromosome to molecular biologists around the world. In the meantime, George Bell, a theoretical physicist formerly in the nuclear weapons program, had founded a unique group in theoretical immunology using LDRD funds. George encouraged his longtime collaborator

*Continued on page 206*

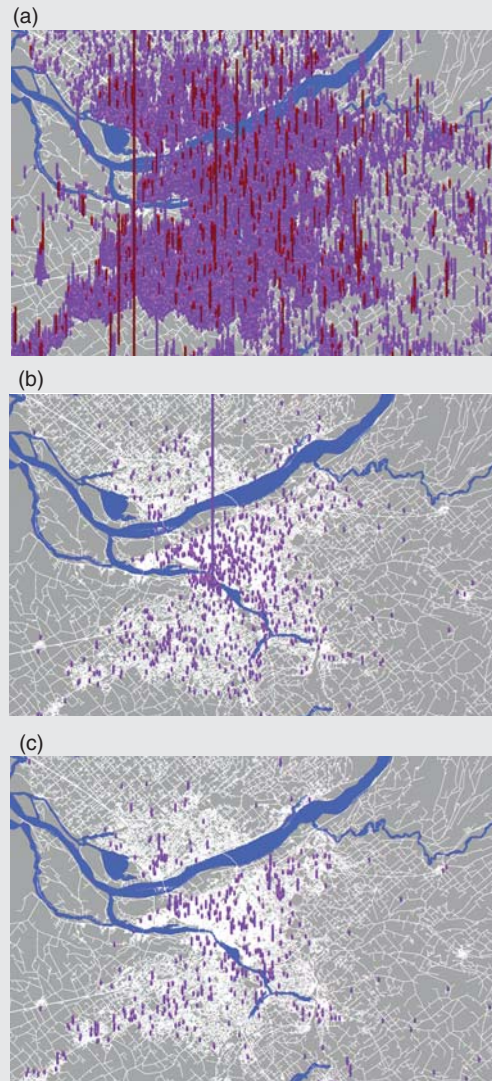
## EpiSims

*Stephen Eubank*

EpiSims is a novel software system for simulating the spread of disease in a large urban population. It is built upon the very detailed models of population mobility provided by UPMoST (Urban Population Mobility Simulation Technology), a simulation system that was also developed at Los Alamos.

UPMoST fuses information from different types of data, such as local census data, household activity data, transportation data, and others and uses them to build a synthetic population of individuals who move about an urban region in a realistic network of contact patterns. EpiSims takes this model of population mobility and adds to it models for how disease is transmitted from person to person or contracted from contaminated environments.

These graphics show the progression of a smallpox epidemic as estimated by an EpiSims simulation. The virus was released among students at a university in downtown Portland, Oregon. Six hours after having been exposed, the infected people—see graphic (a)—have quickly spread through the downtown area. The height of each bar indicates the number of infected people at each location. Graphic (b) shows that, after 40 days, and without a mitigation strategy, the virus has spread throughout the city. The simulation shows that the contagion has touched many demographic groups, so any strategy designed to mitigate the effects of the contagion must accurately take this demographic mixing into account. Graphic (c) shows the viral progression 40 days after the initial infection, but it includes a mitigation strategy that targeted those people who came into contact with contagious people. People who showed symptoms of smallpox were isolated, and their contacts were vaccinated starting 14 days after the attack.



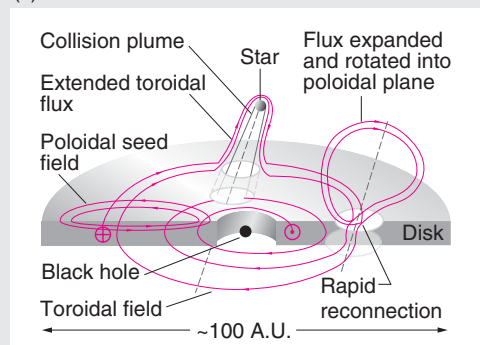
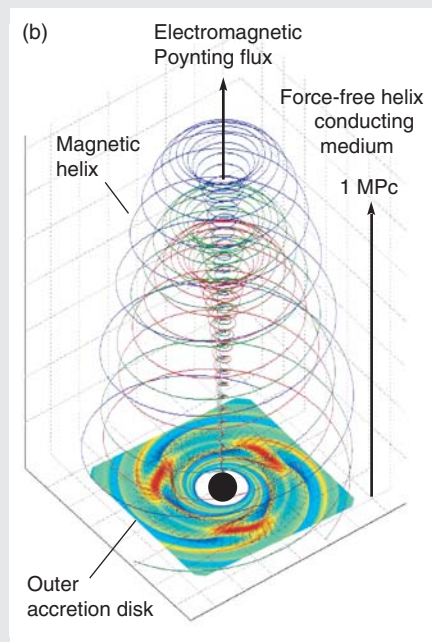
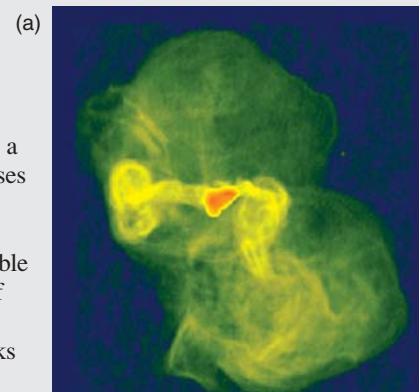
## The Magnetized Universe

Hui Li and Stirling A. Colgate

Observations of the cosmos reveal that in the center of most massive galaxies lies a supermassive black hole. These enigmatic objects form by packing  $10^8$  solar masses into a region the size of our solar system, and roughly  $10^{62}$  ergs of gravitational energy are released during the formation process. A part of this energy is ejected from the region of the black hole as a giant jet or outflow. Such an outflow is visible as the collimated luminous emission in (a), which is a color-coded “radio map” of the giant galaxy M87 in the constellation Virgo. It has been a great challenge to understand the formation of these outflows. Furthermore, where the outflow breaks up, we see radio lobes, which are identified as large regions of space filled with magnetic-field energy and high-energy particles. By studying approximately 100 galaxies with radio lobes, we found that the total energy of the lobes can be up to 10 percent of the energy released when the black hole formed. The identification of this enormous amount of magnetic energy has, among other things, led us to a model of black-hole formation in which gravitational energy is transformed into magnetic energy by a galactic-scale dynamo.

In the early stages of galaxy formation, circulating matter typically forms a disk because of its finite amount of angular momentum. Transporting angular momentum away from the central disk region is essential if the matter is to move inward and collapse to form a supermassive black hole. We have identified a global angular-momentum transport mechanism and performed hydrodynamic simulations to show that large-scale vortices, indicated by colors in (b), can efficiently transport angular momentum outwards. The inner portion of the disk is a conducting plasma that completes roughly  $10^{11}$  revolutions during the  $10^8$  years it takes for the black hole to form. We propose that a dynamo process will convert that rotational kinetic energy into magnetic-field energy. A cut-away view of the inner region is shown in (c). Because of the differential rotation, the disk twists and amplifies a poloidal “seed” magnetic field (with a radial component in the disk) into a toroidal field. Simultaneously, a large number of stars are swarming in and around the disk, and we expect an exceedingly large number of star/disk collisions to occur during the period of black-hole formation. Each collision makes a plume that transforms a fraction of the enhanced toroidal field back into the poloidal field. Calculations indicate that this process can lead to a positive dynamo gain, allowing the seed field to be exponentially amplified. We are conducting a laboratory experiment using liquid sodium to model this dynamo process.

Our calculations also show that the fraction of the poloidal field that extends outside the disk is wrapped up by the disk into a force-free helix, carrying away the energy released by the black-hole formation as a Poynting flux—see graphic (b). After tens of millions of years, the helix extends well beyond the galaxy to nearby galaxies. This fact, along with the observation that supermassive black holes are ubiquitous, led us to suggest that a significant fraction of the volume of the universe is filled with magnetic fields. We are still trying to understand how magnetic reconnection can reconfigure those fields and lead to the distorted shapes seen in (a). We are also trying to find out whether our model can lead to the acceleration of the high-energy particles (for example, cosmic rays) and whether, in the ultimate “global warming” scenario, the field heats up the wider intergalactic medium.



*Continued from page 204*

Walter Goad to follow his visionary ideas regarding molecular biology and start the first data bank (known as GenBank in 1983) for the storage and analysis of DNA sequences. Subsequently, one of George's younger collaborators, Charles Delisi, left Bell's group for an influential position at DOE, and Delisi leveraged those two LDRD-sponsored initiatives, the DNA libraries and data bank, to win DOE sponsorship of the nation's Human Genome Project. The efforts to understand, at a fundamental level, how genomes function have had a direct and continuing impact on our national security. Throughout the 1990s and continuing to this day, technologies from the Human Genome Project have helped combat the threat of biological warfare agents. Rapid DNA analysis tools and miniaturized instruments continue to be developed and fielded for either military or homeland security applications.

As new threats emerge, new areas of science crop up to help mitigate them. A particularly interesting example is the application of quantum mechanical principles to information technology. Richard Feynman recognized that processing information using quantum physics would result in significant differences compared with classical physics approaches. Today, we are witnessing the emergence of a technology based on quantum mechanical principles that can guarantee secure communications. Richard Hughes, originally a postdoctoral fellow in the elementary particles and field theory group, used LDRD funds to start an experimental effort in quantum cryptography, a technique that exploits the properties of single photons to distribute cryptographic keys with guaranteed security against unwanted interception. Hughes and his colleagues made rapid progress and are now supported by programmatic funds. This technology has immediate potential to improve our

national security. We do not yet know all the long-term implications of quantum information technology. However, we have a strong sense that continued research in quantum phenomena will play a major role in the technological edge and security of the nation.

These examples of innovation amply demonstrate that intelligent, technically gifted, tenacious people with broad interests create revolutions that change society. But how does an institution identify such creative people, and how long should they be supported to think and explore without constraint? Should a person whose interests do not match the immediate needs or mission of an institution be supported?

The diversity of past success stories tells us that there is no simple answer, nor is there a cookbook recipe. The best one can do is to assign the task to peers, assuming the maxim that it takes one to know one. The system works provided a sufficiently large community exists, whose members, in spite of large egos and a very high degree of competition, value truth and the quest for truth. For many years, the Laboratory has used technical peer review as its decision mechanism when confronted with funding discretionary research. That mechanism produced high standards in the past, and the Laboratory will best succeed in meeting future challenges by sustaining that tradition.

We will also continue to support the Laboratory's vigorous postdoctoral program. It is a superb mechanism for attracting the best young researchers to our programs. Each year, the Laboratory makes a special strategic investment by recruiting about 25 postdoctoral fellows and 6 distinguished postdoctoral fellows through a highly competitive process. The latter group is evenly divided among the J. Robert Oppenheimer, Feynman, and Reines categories.

Both these groups of fellows are selected on the basis of academic record, research activities and interests, and leadership potential. They are then supported fully for 2 to 3 years to pursue their own research interests, develop a familiarity with the Laboratory's programs, and investigate ways in which they can match their talents with the needs of the Laboratory. Over time, this program has proved to be extremely successful—a significant fraction of those fellows join programs early in their stay at Los Alamos and eventually become leaders, both as scientists and managers. Furthermore, many of the other outstanding candidates who participate in this competition but are not selected as "fellows" are hired as programmatically funded postdoctoral research associates. The competition is intense, and the combined pool of hired candidates, superb. This approach has ensured the highest possible quality of the technical staff recruited through the postdoctoral program. On average, there are 300 postdoctoral fellows and research associates working at the Laboratory at any one time. Many of them become part of the permanent staff. Those who leave Los Alamos for positions in academia or industry often continue to contribute to our intellectual vitality through collaborations with our staff.

Many of those recruits consider their postdoctoral experience as one of the most productive periods of their lives. As permanent staff, however, they often experience too many constraints on their productive time. Proposal writing, delivery on programmatic milestones, and managing business aspects can take too much time away from creative thinking and innovative problem solving. The Laboratory is working to change this situation. No doubt, from an institutional perspective, the Laboratory needs to hire from among the best

and brightest postdocs and hire those who can work effectively on teams, who see the value of multidisciplinary approaches to complex problems, and who can contribute to our diverse work—from fundamental science to pragmatic solutions, and from those to complicated national technology and security needs. But the Laboratory also needs to enable these young scientists to contribute creatively over many decades. The payoffs from basic science are generally long term, and basic science results are seldom accomplished without sustained effort. That effort is possible only through maintaining a good support infrastructure in all the business aspects of the Laboratory and securing stable funding for science.

Historically, very smart people willingly sacrificed their self-interests and worked collectively on projects when they believed in the urgency and importance of the mission. Two outstanding examples of such projects are the Manhattan Project, conducted at a time when eminent scientists understood the danger posed by Nazi Germany and its allies, and the call by President John F. Kennedy in 1961 to put an American on the moon by the end of the decade. Today, we face a similar challenge in the war on global terrorism and on poverty, disease, and diminishing resources. How does an institution retain the capability to respond to such new challenges after years of working on one mission, the stewardship of the nuclear stockpile? Once again, one comes to understand and appreciate the sound judgment and farsightedness of the Laboratory's founders, who put a premium on strategic investment in skilled and talented people and on creating an environment that rewards and nurtures them.

Many of the broad and interconnected challenges being addressed by Laboratory scientists are evident in the topics covered in this section on

strategic investments. In the article following this one, for example, Los Alamos scientists discuss how they are using sophisticated mathematical tools to combat HIV, how they determined when the virus entered the human population, and how they came to understand the complex battle that rages between the virus and the immune system. In another article, scientists describe their latest high-resolution techniques for modeling Earth's oceans. The research promises to increase our ability to predict climate change. For the emerging field of nanotechnology, we have developed a new type of laser based on nanometer-size bits of matter, paving the way for ultraefficient optical devices. Other researchers discuss the Laboratory's seminal role in fuel cell research and development as well as our efforts to simulate the flow of water into and out of a regional basin. The Laboratory is strongly committed to supporting an array of activities in basic science to sustain its vitality and meet the long-term challenges of its missions. The boxes accompanying this overview highlight but a few such projects undertaken by the Los Alamos staff.

At this critical juncture in history, we are rededicating ourselves to the pursuit of excellence in science and technology at Los Alamos. We are also finding new ways to communicate to society at large the essential nature of our basic and applied research and give the nation confidence in the integrity of our science and technology. ■

*For further information, contact  
Rajan Gupta (505) 667-7664  
(rg@lanl.gov).*

# Computational Tools to Battle HIV

*Bette T. M. Korber and Alan S. Perelson*



In 1981, AIDS, the acquired immune deficiency syndrome, was initially detected in a handful of gay men and reported in the Mortality and Morbidity Report of the Centers for Disease Control. Once the disease had been defined and described, its presence was rapidly recognized not only in the homosexual community in the United States but in other populations and countries. By 1983, the etiological, or causative, agent was isolated: a retrovirus<sup>1</sup> called human immunodeficiency virus, type I, or HIV-1. With that discovery, diagnostic tests soon became available, and by the mid-1980s, it was clear that there was a new pathogen on the loose in the world. It was soon recognized that a global epidemic of terrible magnitude was coming, although the extent of the devastation now experienced in sub-Saharan Africa was unimagined in those early days. The World Health Organization established an international network ([www.unaids.org](http://www.unaids.org)) to track this disease as it moved through populations, and through the years, it has maintained a grim record of HIV-1's death toll and spread. Current estimates indicate that more than 65 million people have contracted HIV-1, and 25 million of those are already dead. Africa has been the most brutally affected. Life expectancies in some of the worst afflicted

nations are projected to drop as much as 20 years by 2010. Eurasia sits at the brink of the rapidly advancing storm. Russia, China, and India are facing burgeoning epidemics with no cure and no vaccine yet in hand.

HIV has three peculiarities that make it a particularly challenging foe: its latent period, its devastation of the immune system, and its variability. It has, on average, a decade-long latent period during which infected individuals carry the virus but are not overtly ill; in fact, unless they are tested, they are likely to be unaware of their infection. During this period, the immune system responds to the virus but is unable to clear it, and the ability to make immune responses to any infection slowly deteriorates. Combinations of antiretroviral therapies have been developed that can control viral replication and prolong life and good health, but so far, these drugs have not been able to clear the infection. Therapy is very expensive and, because of side effects, can be difficult to take for years on end.

During the infection, HIV infects and decimates CD4<sup>+</sup> T lymphocytes, the very cells that are central to the immune response needed to counter a viral infection. But the most serious impediment to defeating the virus is its extraordinary variability. This virus evolves during the course of every infection. The immune system responds and inhibits one form of the virus, but a new form inevitably escapes from that response. This cyclical response and evasion continues throughout the infec-

tion, and the human host ultimately loses the race. The virus can quickly acquire drug-resistant mutations, particularly when therapy is only partially successful, and the drug-resistant forms of the virus can be transmitted. This rapid within-host evolution results in extraordinary variation at the epidemic level, and viruses from any two individuals are quite distinct.

The Theoretical Biology and Biophysics Group at Los Alamos has played an integral part in understanding and defining the nature of both the host-viral dynamics and the evolution of HIV. This article covers the history of some of the ideas that were developed by Los Alamos scientists in collaboration with experimental and theoretical colleagues worldwide.

## **The HIV-1 Databases: An International Resource**

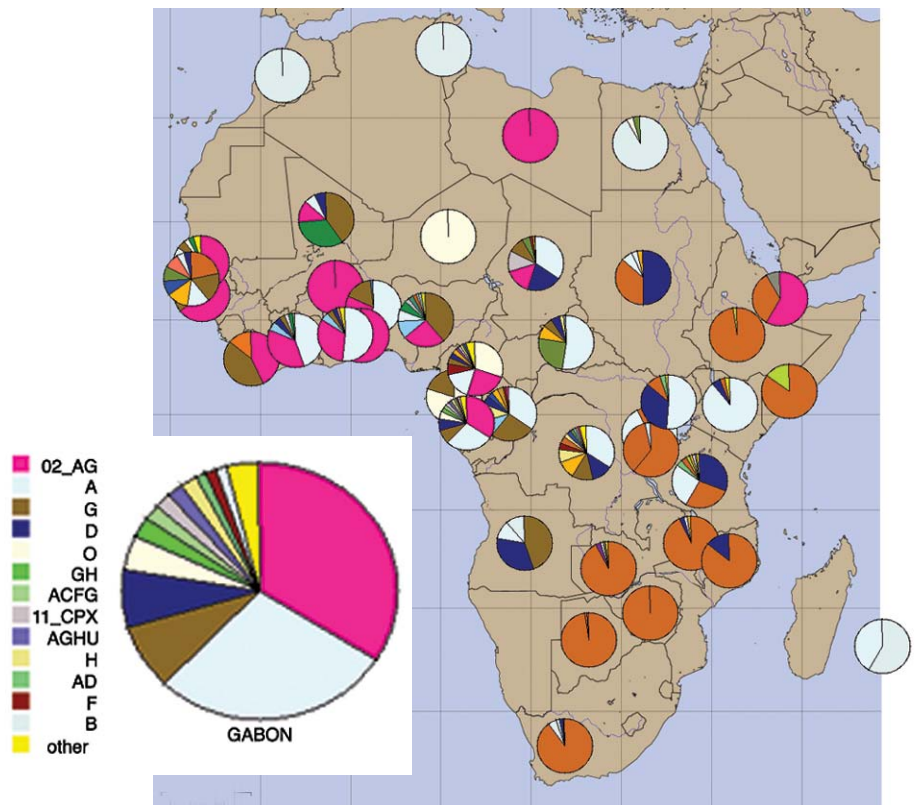
Los Alamos was the original home of GenBank, a database of all publicly available genetic sequences from virtually all organisms. GenBank is now housed at the National Library of Medicine at the National Institutes of Health (NIH). During the mid-1980s, HIV sequences began to be archived in GenBank, arriving not only from the United States and Europe, but from Africa as well. Gerald Myers, who was working at Los Alamos at the time, was struck by the extraordinary diversity of the incoming HIV sequences. He realized, with great prescience, that given this level of diversity, creating a vaccine

<sup>1</sup> A retrovirus is a virus that stores its genetic code in RNA rather than DNA, and once in the host, copies its RNA into DNA. The viral DNA is then incorporated into the host DNA and read to make new virus particles.

was going to be a tremendous challenge and that the HIV research community would benefit from additional curatorial effort. He proposed to generate viral sequence alignments to enable comparisons of similarities between the incoming and archived sequences, to correlate the sequences with any information about their gene products, to link sequences to additional information (for example, the health status of patients and geographic and phenotypic information), and to publish an annual compendium. Researchers with new HIV sequences could then quickly put them into the context of a global framework without having to extract and organize the information each time. The NIH, persuaded by Myers, subsequently funded a program to create the HIV sequence database.

In the mid-1990s, Myers initiated databases for other pathogens, leaving the HIV project to Los Alamos researchers Bette Korber and Carla Kuiken. Sequencing technologies were rapidly improving, and HIV sequences were being generated at an accelerating pace. With a dedicated staff contributing both computer and biological expertise, the HIV sequence database developed into an easily used, Web-accessible, relational database, which now houses roughly 80,000 HIV sequences. Web search tools were added to retrieve data in new ways. For example, a researcher working on a vaccine for South Africa can pull up an automated alignment of all regional sequences, organized by year of isolation, by subtype, or by phenotype (see Figure 1.) Pioneering computational tools were developed for handling such basic problems as detecting viral recombination, hypermutation, and polymerase-chain-reaction contamination, tools that served to improve vastly the overall quality of the HIV scientific literature.

In 1995, an immunology database that contains information concerning HIV immune responses was created



**Figure 1. Tools to Combat HIV**

Los Alamos is home to four databases that archive most of what is known about HIV and its associated immune responses, drug resistance, and vaccine trials. The database website at [www.hiv.lanl.gov](http://www.hiv.lanl.gov) also contains numerous analysis tools. This figure is a screenshot from a geography tool that provides a bird's-eye view of viral diversity throughout a region, in this case the African continent. The pie chart indicates the distribution of viral strains for different subregions. Clicking on an individual pie chart on the Web brings up more information about the local viral strains. Clicking on a strain accesses the sequence information contained in the database.

and added to the sequence database. It allowed a researcher to integrate immune-response data with sequence variability data. In 1997, an HIV drug-resistance mutation database was added, and in 2002, a vaccine-trial database became part of the collection. The latter was developed by Los Alamos postdoctoral fellow John Mokili. It summarizes and allows direct comparisons of data sets from hundreds of vaccine studies conducted in primates. The HIV database Web pages, with access to all four databases, received more than a million hits per month during the peak use period in 2002, averaging 35,137 pages accessed

per day. The Los Alamos databases became an integral part of global HIV research efforts, providing a foundation for continuing scientific work.

## HIV, the Shape Changer

The database was developed to be an international resource, but the integrated data it provides also serves as a basis for our own research efforts, and we have used it to study the evolution of the virus from many different perspectives. The exposed envelope protein of HIV (shown in Figure 2) is the most antibody-vulnerable part of the

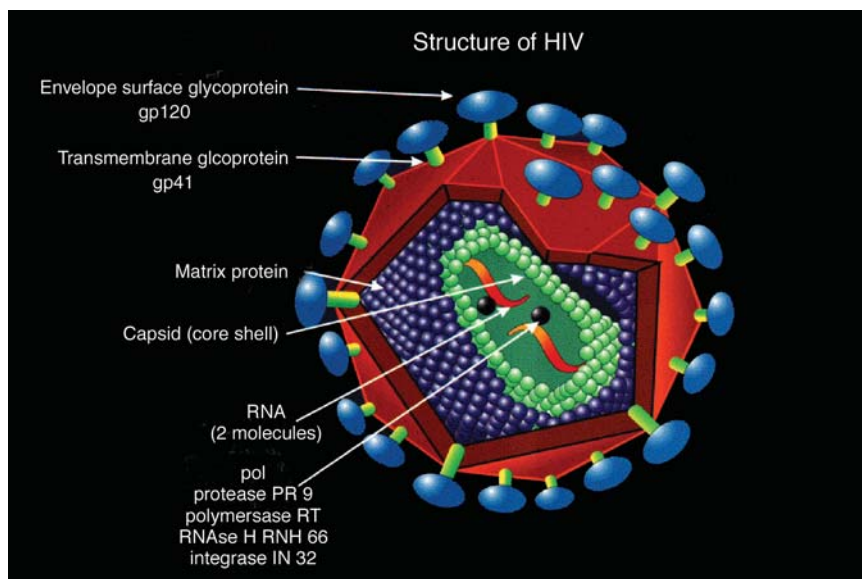
virus. The envelope has two different protein components: gp41, which is locked into the outer membrane of the virus, and gp120, which is bound to gp41. The component gp120 projects from the surface of the virus, allowing HIV to recognize and bind to receptors on the human cell it infects. Gp120 is remarkably mutable, and sequences from different lineages, or subtypes, of HIV can differ in more than 30 percent of their amino acids.

But even the high fraction of amino acid changes does not reflect HIV's true capacity to generate structural diversity. The virus has at least five mechanisms to alter its antigenic conformation—the molecular shape that is relevant to an immune response—and thus exhibits an extraordinary (and unique) capacity to avoid triggering the immune system. Following are the five mechanisms:

*Mutational change through base substitution.* This process typically comes to mind when one thinks about mutations—one nucleotide is replaced by another because the polymerase molecule made a mistake in copying genetic information. This process is the common mechanism—typically the only mechanism—by which viruses diversify and change. Substitutions occur at a high rate in HIV. The process is readily modeled and provides the basis for phylogenetic analysis of the evolutionary history of HIV.

*Insertions and deletions.* Hypervariable domains in HIV's genome lead to frequent alterations in the number of amino acids in the HIV-1 envelope proteins so that, within a few generations, the virus presents a new face to the immune system. The high frequency of insertion and deletion mutational events makes HIV-1 envelope proteins difficult to align and analyze, and the evolution of such regions cannot be modeled with currently available tools.

*Shifts in glycosylation site pattern.*



**Figure 2. Human Immunodeficiency Virus, Type 1**

This illustration depicts the essential features of HIV-1. The virus stores its genetic information in two strands of RNA. These lie within a protein layer known as the capsid, which is surrounded by the matrix protein. Also lying within the capsid is a metaprotein complex called pol, which contains four protein subunits: a reverse transcriptase, which will copy the RNA into DNA once the RNA has entered a host cell; an integrase, which inserts the viral DNA into the host cell's genome, an RNase, and a protease that cleaves pol into the four subunits. Thus, upon entering a cell, pol literally cuts itself apart. The entire virus is encased in a lipid membrane that is actually a piece of the host's cell membrane. The two proteins gp41 and gp120 are the envelope protein of the virus. The highly exposed gp120 protein binds to CD4 receptors on the host cell as the first step toward infecting the cell.

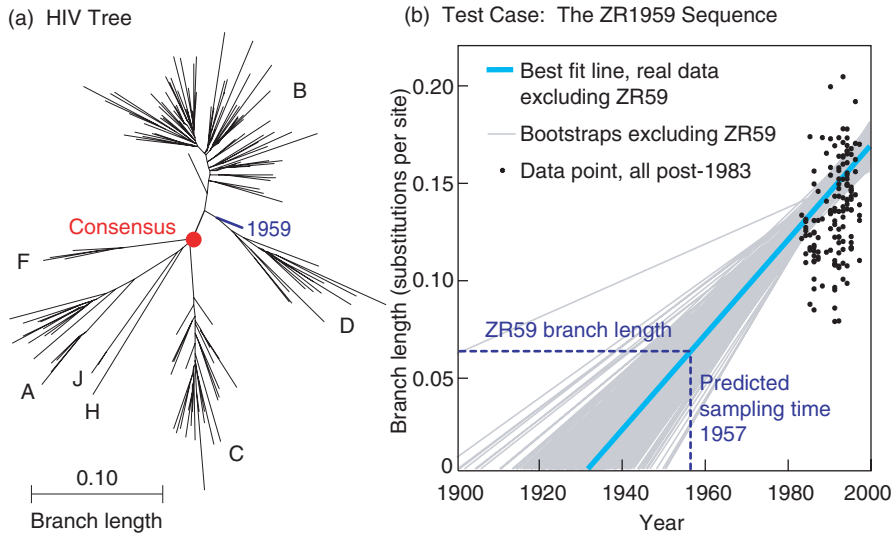
The outer face of the HIV envelope protein is essentially covered by carbohydrates—that is, it is heavily glycosylated. (It is one of the most heavily glycosylated proteins in nature.) The carbohydrates are thought to partially shield HIV from antibody responses. The average gp120 molecule has 25 glycosylation sites, but the number varies between 18 and 33. The gain or loss of a carbohydrate moiety can alter the conformation of a protein and abrogate the ability of certain antibodies to bind.

*Recombination.* HIV carries two strands of RNA into newly infected cells. Because of frequent substitutions, insertions, and deletions, the two strands may be different. When they are copied, they can recombine, parts of one strand intermixing with parts of another.

Researchers have found many examples of recombinant viruses that resulted from splicing together dramatically different strains of HIV. Undetected recombination can cause phylogenetic reconstruction to be inaccurate and evolutionary inferences to be incorrect.

*Change at a distance.* Mutational changes in gp41 can result in conformational changes in gp120 that can alter gp120 antibody binding sites.

Each of these mechanisms for change has been studied and tracked by the HIV database group, but it was base substitutions, the first kind of change, that allowed us to model the phylogenetic (or evolutionary) history of the virus. By ascertaining when HIV-1 entered the human population, we



### Figure 3. The Origin of HIV

(a) By aligning more than 150 HIV sequences, we constructed a phylogenetic tree that showed how HIV evolved. Each fanlike cluster of branches corresponds to an HIV subtype, and each terminating branch represents an HIV strain. The “root” of the tree (red dot) corresponds to the unobserved viral matriarch, the ancestral virus responsible for the strains now driving the AIDS pandemic. (b) The distance between the root and each branch (the branch length) is a measure of how many mutations occurred between the matriarch and each strain. All our HIV samples had a known sampling date, so given the tree and assuming a uniform rate of evolution, we could plot the branch length against the year of sampling and fit a line (turquoise) through the data points. Extrapolating the line backward indicates that the virus began to spread through the human population in 1930, give or take a decade. The gray lines are fits from a “bootstrap” method that was used to estimate the uncertainty. Our 1930 date was given support when we considered an “ancient” HIV sequence obtained in 1959. According to our tree, the sequence had a branch length of about 0.6, and according to our best fit, would have been spawned in 1957.

could estimate how rapidly the virus was diversifying and thus better understand the progression of the disease.

The oldest human blood sample documented to contain HIV-1 was taken in 1959 and came from an individual living in central Africa. It was sequenced in the late 1990s, and the sequence was analyzed at Los Alamos. This sample provided an “ancient” HIV sequence (relatively speaking), which allowed us to calibrate our evolutionary models. Drawing on a diverse set of Los Alamos scientists with interdisciplinary skills in computation, modeling, and statistics, and making use of the Los Alamos supercomputing facility to create optimized phylogenetic trees, we were able to estimate that the spread of

HIV through the human population began in 1930, plus or minus a decade (see Figure 3).

Further modeling suggested a very slow initial spread of the virus, possibly indicative of a time when it was confined to rural areas with limited transmission possibilities. These estimates enabled better understanding of the rate of diversification of HIV and of how long HIV took to get to the present level of diversity. They helped rule out some controversial theories about the origins of HIV, and they moved us closer to understanding the history of this virus in its human host, a vital topic given how this epidemic has changed the landscape of human history in the 20 years since its discovery.

Another area of study has been an ongoing effort to understand how the human immune response influences viral variation in populations. Our cells routinely chop up internal proteins into short amino-acid segments and ferry those segments to the surface, where they are “examined” by a prowling immune-system cell. The immune system is triggered when the segment is deemed foreign to the body, but it is well known that certain amino acids are less likely to trigger an immune-system response. Mining the immunology literature, we were able to identify the parts of the virus that get chopped up—so-called antigenic regions—that are the focus of the immune responses in many individuals. An algorithm then allowed us to predict successfully where antigenic regions would be concentrated in less well studied HIV proteins. In the most variable regions of these viral proteins, we found a significant enrichment of certain amino acids, indicating that HIV had evolved to make itself less vulnerable to attack. Thus, human immune responses have left a clear imprint on the evolution of the virus.

The virus also may be leaving its evolutionary imprint on us. Certain human genes can influence our susceptibility to infection and our ability to live with HIV. In collaboration with the Santa Fe Institute and members of our theoretical group, we have been working to understand and define these genes. In populations where the virus is highly prevalent, such advantages may shift the human population in favor of those carrying such genes.

Finally, we have designed artificial consensus sequences (or reconstructed ancestor sequences) that are more similar to circulating strains of the virus than the various strains are to each other. Proteins from these artificial sequences are now widely used by experimentalists to probe and study the T cell immune responses of HIV-infected individuals. Our hope for these artificial proteins is that they will be more likely than



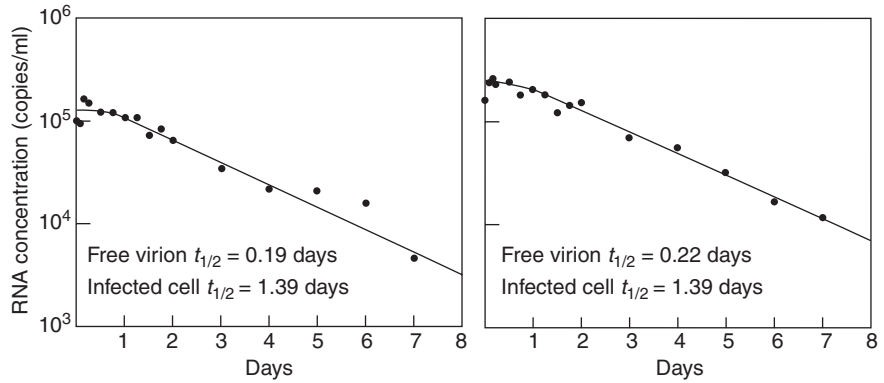
natural strains to elicit cross-reactive immune responses if used as a vaccine. They effectively reduce by half the number of amino-acid differences between a vaccine candidate strain and circulating viruses. Our colleagues at the University of Alabama and Duke University currently have experiments under way to test the immunological cross-reactivity of these proteins, and the initial results are encouraging. This kind of vaccine design strategy could be used in conjunction with other strategies—for example, those that deliberately expose immunologically vulnerable parts of the envelope that are usually hidden. Eventually, the result could be production of vaccines with better potential to protect individuals against infection from the extraordinarily diverse pool of circulating viruses.

### A Model of HIV Dynamics

Developing a successful HIV vaccine is our ultimate goal, but in the meantime, we have made enormous progress in learning how to fight the virus with drugs. In pioneering studies conducted by Alan Perelson, in collaboration with David Ho's group at the Aaron Diamond AIDS Research Foundation, Rockefeller University, we addressed the question of whether the average 10-year time from HIV infection to AIDS reflects the fact that HIV grew slowly, and therefore did not need aggressive treatment.

By studying chronically HIV-infected individuals whose HIV concentrations in blood were relatively constant, we showed that giving a drug called Ritonavir<sup>2</sup>, a potent inhibitor of HIV-1 protease, caused the HIV concentration in blood to

<sup>2</sup>Ritonavir goes by the tradename Norvir and is manufactured by Abbot Laboratories.



**Figure 4. HIV Dynamics in the Presence of an Inhibitory Drug**  
 This figure shows a fit of Equation (3) in the text to data from two representative patients. The circles show the concentration of HIV-1 RNA after Ritonavir treatment has begun on day 0. The theoretical curve was obtained by a nonlinear least-squares fit. From the fit, we could estimate the rate of clearance of the virus, the rate of loss of infected cells, and the initial viral load, that is, parameters  $c$ ,  $\delta$ , and  $V_0$ , respectively, in Equation (3).

drop about 100-fold in two weeks. We then showed that this rapid decline implied that the body was rapidly clearing HIV. With this and other experiments, we were able to estimate that the time to clear half of the HIV in blood (the half-life) was about one hour or less. But we also knew that infected T cells were constantly producing HIV. The virus' short half-life therefore implied that the T cells also had to die rapidly. By more detailed experiments and analysis, we were able to estimate that the CD4<sup>+</sup> T cells that were the major producers of HIV lived only about one day while producing HIV. The question of whether HIV kills these cells directly or the immune response plays some significant role is still being debated. Nevertheless, this work showed that HIV was being maintained in the body by a vicious cycle: the virus was being rapidly produced and cleared, and while it was present, it was infecting cells and killing them.

The analysis involved developing models of viral infection and the effect of treatment. To illustrate the approach, consider the following simple model of viral infection:

$$\begin{aligned} \frac{dT}{dt} &= s - \alpha T - kVT \\ \frac{dI}{dt} &= kVT - \delta I \\ \frac{dV}{dt} &= N\delta I - cV \end{aligned} \tag{1}$$

where  $T$  is the concentration of uninfected cells,  $I$  the concentration of infected cells, and  $V$  is the concentration of virus particles or virions. Here, we assume uninfected cells are created from a source at rate  $s$  and die at a per capita rate  $\alpha$ . In addition, we assume they are infected by the virus with a rate constant  $k$ . Infected cells are assumed to die at the *per capita* rate  $\delta$  and to release a total of  $N$  viral particles during their lives. We also assume that the body clears the virus by a first-order process, with a constant rate of clearance per virion given by  $c$ . For simplicity, the loss of virus upon infecting cells is neglected, although it can be included. When drug therapy with a protease inhibitor is given, a fraction  $\epsilon$  of newly formed virus particles is noninfectious. Thus, in the presence of the drug, the model equations become:

$$\begin{aligned}
\frac{dT}{dt} &= s - \alpha T - kV_I T \\
\frac{dI}{dt} &= kV_I T - \delta I \\
\frac{dV_I}{dt} &= (1 - \varepsilon)N\delta I - cV_I \\
\frac{dV_{NI}}{dt} &= \varepsilon N\delta I - cV_{NI}
\end{aligned}
\tag{2}$$

where  $V_I$  and  $V_{NI}$  denote infectious and noninfectious virus, respectively. If one analyzes patient data obtained over the first week of therapy,  $T$  does not change greatly and can be assumed to be constant. Under this approximation, the system of equations becomes linear and can be solved exactly.

Two other approximations were made to allow us to gain insight into the solutions and to make comparisons with patient data. First, measurements of the total amount of virus in blood showed that, in most chronically infected patients,  $V$  was approximately constant over periods of about weeks or months. Thus, before drug therapy, a steady state was assumed, which implied that  $c = NkT_0$ , where  $T_0$  was the measured level of T cells before therapy. Second, the drug efficacy was assumed to be 100 percent, that is,  $\varepsilon = 1$ . Under these considerations, one could show that

$$\begin{aligned}
V(t) &= V_0 e^{-ct} + \frac{cV_0}{c - \delta} \times \\
&\left[ \frac{c}{c - \delta} \left\{ e^{-\delta t} - e^{-ct} \right\} - \delta t e^{-ct} \right]
\end{aligned}
\tag{3}$$

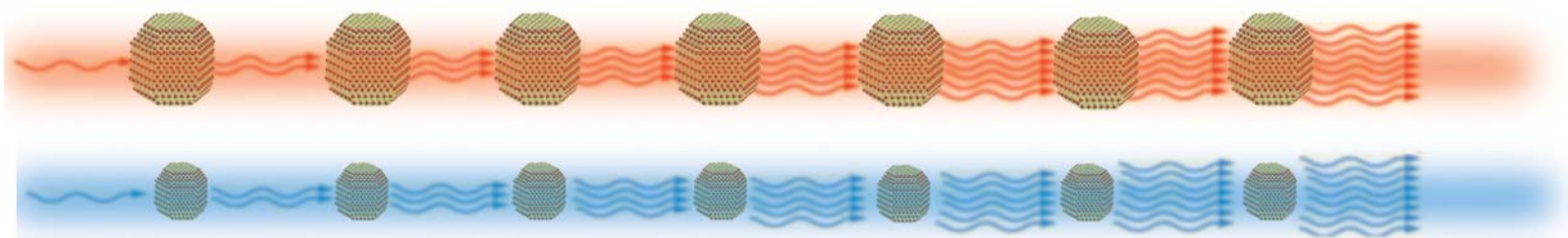
where  $T$  is time on therapy and  $V(0) = V_0$ . Using nonlinear regression techniques, we tried to fit this model to measured values of  $V$  and obtained estimates of the parameters  $c$  and  $\delta$  (see Figure 4). Further, we could show that if  $\varepsilon < 1$ , then these estimates of the clearance rate of HIV and the *per capita* death rate of T cells were minimal estimates (and thus infected cells and virus might be cleared even faster than estimated

by this method). We could also show numerically that, if  $T$  changed by the amounts observed in patients, then the estimates would vary by only a few percent. Hence, by this method, the first estimates of the *in vivo* half-lives of HIV and infected cells were obtained, and they showed in a stunning manner that HIV infection was highly dynamic. From estimates of the rate of growth of the virus needed to maintain the observed constant levels of virus in the face of the estimated clearance, we showed that HIV would mutate sufficiently to become resistant to any single drug—an even more important finding. This determination, along with the observation that drug therapy could rapidly decrease the viral load, helped usher in the age of combination drug therapy for the treatment of AIDS. ■

## Further Reading

- Eberstadt, N. 2002. The Future of AIDS. *Foreign Aff.* **81** (6): 22.
- Gaschen, B., J. Taylor, K. Yusim, B. Foley, F. Gao, D. Lang et al. 2002. Diversity Considerations in HIV-1 Vaccine Selection. *Science* **296** (5577): 2354.
- Korber, B., M. Muldoon, J. Theiler, F. Gao, R. Gupta, A. Lapedes et al. 2000. Timing the Ancestor of the HIV-1 Pandemic Strains. *Science* **288** (5476): 1789.
- The Los Alamos HIV sequence, immunology, vaccine, and drug-resistance databases. <http://www.hiv.lanl.gov>.
- Perelson, A. S. 2002. Modelling Viral and Immune System Dynamics. *Nat. Rev. Immunol.* **2**: 28.
- Perelson, A. S., A. U. Neumann, M. Markowitz, J. M. Leonard, and D. D. Ho. 1996. HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time. *Science* **271** (5255): 1582.
- Perelson, A. S., P. Essunger, Y. Cao, M. Vesanen, A. Hurley, K. Saksela et al. 1997. Decay Characteristics of HIV-1-Infected Compartments During Combination Therapy. *Nature* **387** (6629): 188.
- The UNAIDS WHO site concerning the global status of HIV/AIDS. <http://www.unaids.org>.
- Walker, B. D., and B. T. Korber. 2001. Immune Control of HIV: The Obstacles of HLA and Viral Diversity. *Nat. Immunol.* **2**: 473.

For further information, contact Bette Korber (505) 665-4453 ([btk@lanl.gov](mailto:btk@lanl.gov)).



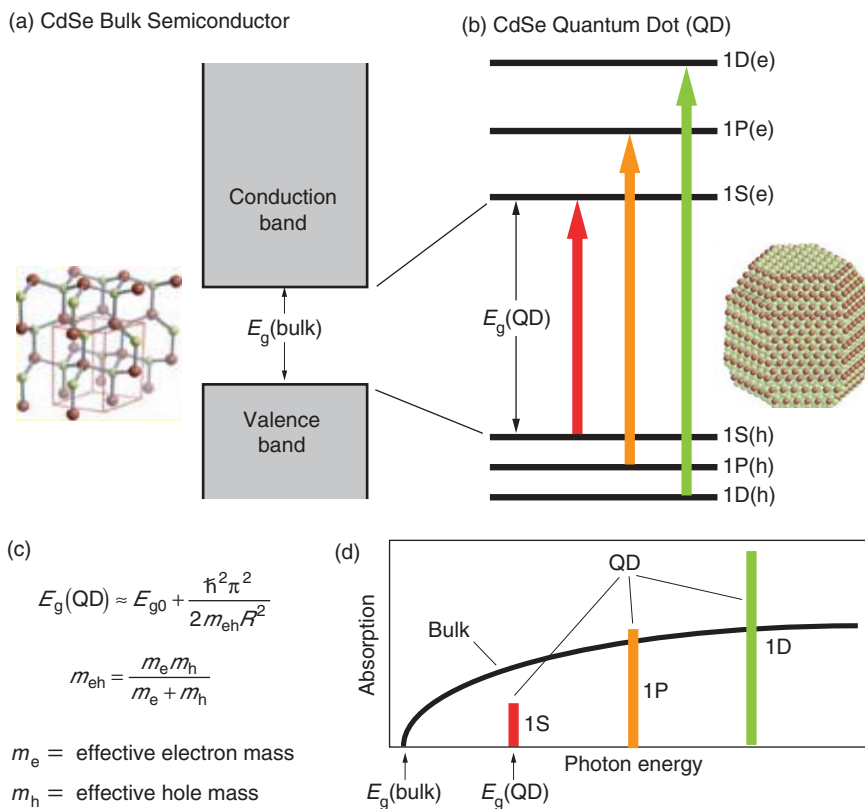
# Nanocrystal Quantum Dots

## *From fundamental photophysics to multicolor lasing*

Victor I. Klimov

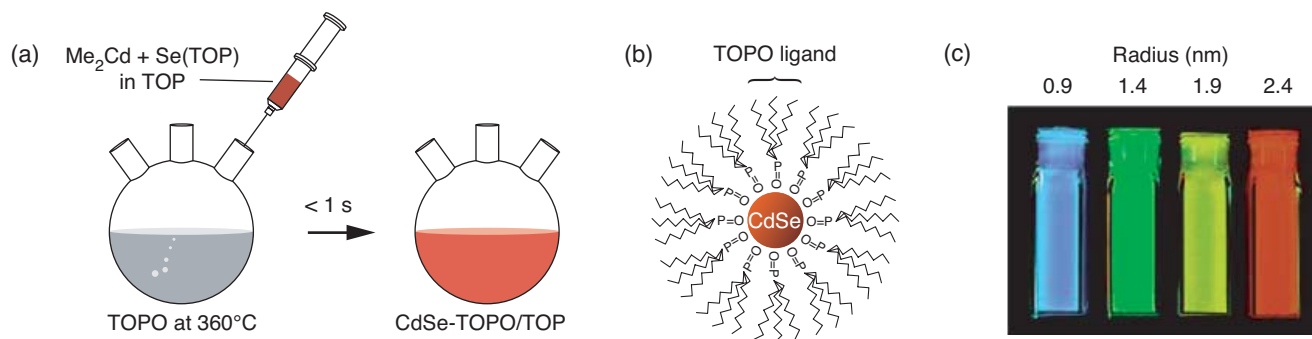
Semiconductor lasers are ubiquitous in modern society and play a key role in technologies ranging from CD players to optical telecommunications. Current-generation lasers have high power output and low lasing thresholds, are stable over a wide range of temperatures, and are cheap and easy to produce. Still, there is room for improvement. We are developing a new type of laser based on ultrasmall bits of semiconductor material called quantum dots (QDs). Consisting of only a few hundred to a few hundred thousand atoms, QDs bridge the gap between the solid state and single atoms, and hence these specks of matter exhibit a mix of solid-state and atomic properties. In our work, we concentrate on nanoparticles that are synthesized by colloidal chemistry, and therefore, they are often called colloidal or nanocrystal QDs (NQDs). Interestingly, the emission wavelength (that is, the emission color) of QDs depends on the dot size, and in the case of semiconductor nanocrystals, color can be controlled precisely through simple chemistry. We are therefore developing an altogether new type of color-selectable lasing medium.

Although this paper focuses on our NQD laser work, quantum dots are “bigger” than lasers. Because of their small dimensions and size-controlled electronic spectra, NQDs can be viewed as tunable artificial atoms



**Figure 1. Quantum Dots (QDs)**

(a) A bulk semiconductor such as CdSe has continuous conduction and valence energy bands separated by a “fixed” energy gap,  $E_g(\text{bulk})$ . Electrons normally occupy all states up to the edge of the valence band, whereas states in the conduction band are empty. (b) A QD is characterized by discrete atomic-like states with energies that are determined by the QD radius  $R$ . These well-separated QD states can be labeled with atomic-like notations, such as 1S, 1P, and 1D. (c) The expression for the size-dependence separation between the lowest electron and hole QD states— $E_g(\text{QD})$ , the QD energy gap—was obtained with the spherical “quantum box” model. (d) This schematic represents the continuous absorption spectrum of a bulk semiconductor (black line) compared with a discrete absorption spectrum of a QD (colored bars).



**Figure 2. Nanocrystal Quantum Dots (NQDs)**

(a) An organometallic method is used for the fabrication of highly monodisperse CdSe NQDs. Nucleation and subsequent growth of NQDs occurs after a quick injection of metal and chalcogenide precursors into the hot, strongly coordinating solvent—a mixture of trioctylphosphine (TOP) and trioctylphosphine oxide (TOPO) in the case shown. After a fixed period, removing the heat source stops the reaction. As a result, NQDs of a particular size form. (b) The colloidal NQDs obtained by the method illustrated in (a) consist of an inorganic CdSe core capped with a layer of TOPO/TOP molecules. (c) Solutions of CdSe NQDs of different radii, under ultraviolet illumination, emit different colors because of the quantum size effect. A 2.4-nm-radius dot has an energy gap of about 2 eV and emits in the orange, whereas a dot of radius 0.9 nm has a gap of about 2.7 eV and emits a blue color.

with properties that can be engineered to suit either the needs of a certain experiment or a specific technological application. When coated with a suitable, chemically active surface layer, NQDs can be coupled to each other or to different inorganic or organic entities and thus serve as useful optical tags. We can now chemically manipulate NQDs almost as well as standard molecules, and can assemble them into close-packed ordered or disordered arrays that mimic naturally occurring solids. Furthermore, because their dimensions, shapes, and surface properties can be manipulated with ease, NQDs are ideally suited to serve as nanoscale laboratories for studies of fundamental quantum mechanical effects.

### The Quantum Size Effect and QDs

One of the defining features of a semiconductor is the energy gap separating the conduction and valence energy bands. The color of light emitted by the semiconductor material is determined by the width of the gap. In semiconductors of macroscopic sizes—bulk semiconductors—the gap

width is a fixed parameter determined by the material's identity.

The situation changes, however, in the case of nanoscale semiconductor particles with sizes smaller than about 10 nanometers. This size range corresponds to the regime of quantum confinement, for which the spatial extent of the electronic wave function is comparable with the dot size. As a result of these “geometrical” constraints, electrons “feel” the presence of the particle boundaries and respond to changes in particle size by adjusting their energy. This phenomenon is known as the quantum-size effect, and it plays a very important role in QDs.

In the first approximation, the quantum-size effect can be described by a simple “quantum box” model (Efros and Efros 1982), in which the electron motion is restricted in all three dimensions by impenetrable walls. For a spherical QD with radius  $R$ , this model predicts that a size-dependent contribution to the energy gap is simply proportional to  $1/R^2$ , implying that the gap increases as the QD size decreases. In addition, quantum confinement leads to a collapse of the continuous energy bands of a bulk material into discrete, atomic-like energy levels. The discrete struc-

ture of energy states leads to a discrete absorption spectrum of QDs, which is in contrast to the continuous absorption spectrum of a bulk semiconductor (see Figure 1).

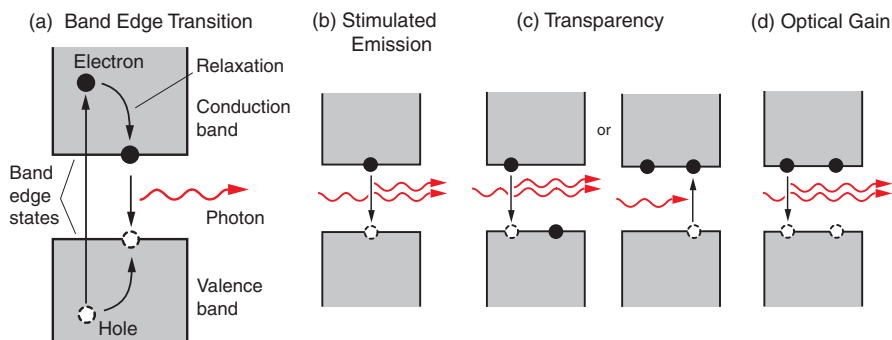
The NQDs discussed earlier are small quantum dots that are made by organometallic chemical methods and are composed of a semiconductor core capped with a layer of organic molecules (Murray et al. 1993). (See Figure 2.) The organic capping prevents uncontrolled growth and agglomeration of the nanoparticles. It also allows NQDs to be chemically manipulated as if they were large molecules, with solubility and chemical reactivity determined by the identity of the organic molecules. The capping also provides “electronic” passivation of NQDs; that is, it terminates dangling bonds that remain on the semiconductor's surface. As discussed below, the unterminated dangling bonds can affect the NQD's emission efficiency because they lead to a loss mechanism wherein electrons are rapidly trapped at the surface before they have a chance to emit a photon. Using colloidal chemical syntheses, one can prepare NQDs with nearly atomic precision; their diameters range from nanometers to tens of

nanometers and size dispersions as narrow as 5 percent. Because of the quantum-size effect, this ability to tune the NQD size translates into a means of controlling various NQD properties, such as emission and absorption wavelengths. The emission of cadmium-selenium (CdSe) NQDs, for example, can be tuned from deep red to blue by a reduction in the dot radius from 5 nanometers to 0.7 nanometer.

### Nanocrystal Lasers: Advantages and Problems

Lasers made from bulk semiconductor materials have been used for several decades. (Laser fundamentals are described in Figure 3.) Although numerous advances were made throughout those years, laser performance improved dramatically with the introduction of so-called quantum well lasers, in which charge carriers—electrons and holes—were confined to move in a plane—that is, they were free to move in a two-dimensional (2-D) quantum well. Compared with bulk semiconductors, the quantum well has a higher density of electronic states near the edges of the conduction and valence bands, and therefore a higher concentration of carriers can contribute to the band-edge emission. Consequently, it takes less intense “pumping” of energy into a quantum-well laser to get it to lase (the lasing threshold is lower). Additionally, quantum-well lasers show improved temperature stability and a narrower emission line.

In QDs, the charge carriers are confined in all three dimensions, with the result that the electrons exhibit a discrete atomic-like energy spectrum. In very small QDs, the spacing between these atomic-like states is greater than the available thermal energy, so thermal depopulation of the lowest electronic states is inhib-



**Figure 3. Laser Basics**

(a) “Pumping” energy into a semiconductor can excite an electron,  $e$ , into the conduction band. That electron leaves behind a hole,  $h$ , in the normally filled valence band, and thus an  $e$ - $h$  pair is created. The electron and hole each relax to the respective band-edge states by nonradiative processes. During the band-edge transition, a photon is emitted as the excited electron spontaneously recombines with the hole. (b) Stimulated emission occurs when a photon induces the excited electron to decay. The emitted photon has the exact frequency, phase, and polarization of the initial photon. (c) For a ground state that contains two electrons (population equality) can lead to two equally probable outcomes: The incoming photon stimulates the excited electron to decay, producing an extra photon (left), or the photon excites the ground-state electron and is absorbed (right). There is no net gain or loss of photons. In this case, the medium is in the transparency regime. (d) Optical gain can occur if there are more electrons in the excited state than in the ground state (population inversion) because photon absorptions are inhibited. If a population inversion is established in a bulk system and if the gain from stimulated emission is larger than losses that absorb or scatter photons, the system will exhibit amplified spontaneous emission (ASE). In a laser, an ASE-capable medium is placed in a reflecting cavity, and thus the photon field builds on itself.

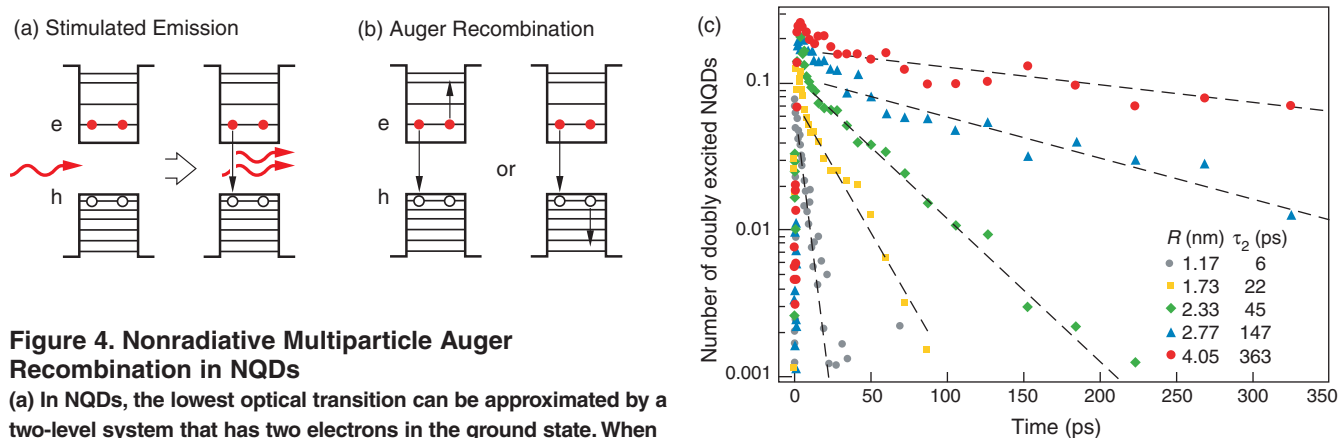
ed. It was therefore anticipated that a QD laser would have a temperature-insensitive lasing threshold at an excitation level of only one electron-hole ( $e$ - $h$ ) pair per dot.

Lasing in QDs was first reported in 1991 (Vandyshv et al. 1991) and was achieved in an optically pumped device with relatively large (approximately 10-nanometer) CdSe nanoparticles. The QDs were fabricated by high-temperature precipitation in molten glass. Later, lasing was also observed for QDs grown by epitaxial techniques (Ledentsov et al. 1994). As expected, the QD lasers showed an improved performance and featured a lower lasing threshold and enhanced temperature stability by comparison with quantum-well lasers.

These successes provided us with strong motivation for the development

of lasers based on NQDs less than 10 nanometers in diameter. In this size range, spacing between electronic levels can exceed hundreds of milli-electron-volts (meV), a much larger value than the room temperature energy scale of about 24 meV. Size-controlled spectral tunability over an energy range of 1 electron volt was expected. However, after a decade of research that provided some tantalizing hints of optical gain, NQDs failed to demonstrate lasing action.

The failures were often attributed to material defects or dangling bonds on the surface of the NQDs, which were a natural consequence of the large surface-to-volume ratio of the sub-10-nanometer particles. The defects lead to electronic states that lie within the material’s energy gap. Electrons can relax into those states,



**Figure 4. Nonradiative Multiparticle Auger Recombination in NQDs**

(a) In NQDs, the lowest optical transition can be approximated by a two-level system that has two electrons in the ground state. When both electrons are excited, a population inversion occurs, and the NQD can exhibit optical gain. An incoming photon stimulates one electron to decay, producing an extra photon. (b) The two-electron excited state also allows for a loss mechanism called nonradiative Auger recombination, whereby the energy from e-h recombination is not released as a photon but is transferred to either an electron or a hole. (c) Experiments show that the smaller the dot, the shorter the Auger recombination time ( $\tau_2$ ). Even the largest dot has a significantly shorter  $\tau_2$  than the radiative decay time.

whereupon they typically undergo either nonradiative or radiative (in-gap “deep-trap” emission) decay to the ground state. Thus the surface defects introduce carrier losses that inhibit the optical gain. Another concern raised in several theoretical papers was the reduced efficiency of electron-phonon interactions that results from the discrete, atomic-like energy structures, an effect that reduces the ability of carriers to enter into the band-edge states and hence reduces luminescence efficiencies. However, our research team eventually realized that the main difficulty in getting our ultrasmall NQDs to lase stemmed from a largely unforeseen problem known as multiparticle Auger recombination (Klimov et al. 2000).

### Multiparticle Auger Recombination vs Optical Gain

As in the case of other lasing media, QDs require a population inversion in order to produce optical gain (refer to Figure 3). The population inversion corresponds to the situation in which the number of electrons in a high-energy excited state is

greater than that in the low-energy ground state. In small dots, the lowest “emitting” transition can be treated as a two-level system that contains two electrons in its ground state. To invert such a system, one has to promote both electrons from the ground to the excited state, meaning that optical gain in QDs originate from nanoparticles that contain two e-h pairs (doubly excited nanoparticles).

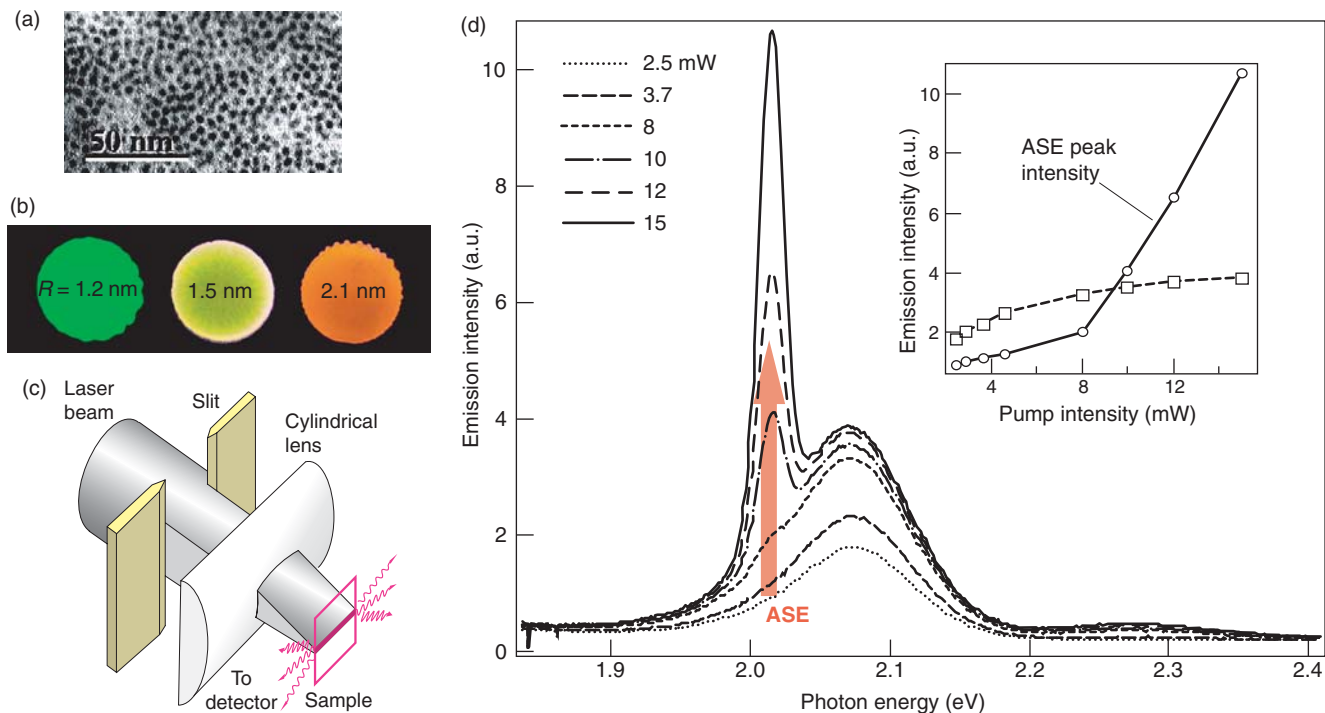
Paradoxically, whereas the intrinsic decay of singly excited QDs is due to the e-h recombination and the emission of a photon, the deactivation of two e-h pair states is dominated by nonradiative Auger recombination (Klimov et al. 2000a). In the latter case, the e-h recombination energy is not released as a photon but is transferred to a third particle (an electron or a hole) that is re-excited to a higher energy state (see Figure 4). Auger recombination has a relatively low efficiency in bulk semiconductors because of restrictions imposed by energy and momentum conservation. But linear, or translational, momentum conservation is a consequence of the translation symmetry of bulk crystals, and this symmetry is broken in QDs (the electrons feel the dot’s boundaries). Therefore, translational-

momentum conservation does not apply to QDs, so the probability of Auger effects is greatly enhanced.

Since Auger recombination and optical gain develop from the same initial state (that is, two e-h pairs in a dot), the Auger decay is unavoidable in the regime of optical amplification and will always impose an intrinsic limit on optical gain lifetimes. In CdSe NQDs, for example, Auger recombination leads to the deactivation of doubly excited nanoparticles on time scales from approximately 400 picoseconds to approximately 10 picoseconds, depending on the dot size (the smaller the dot, the faster the recombination). These time scales are significantly shorter than the time of the radiative decay (approximately 20 to 30 nanoseconds), which obviously should hinder the development of lasing.

### QD Solids: A New Type of Lasing Medium

We realized the hindering role of Auger recombination only toward the end of 1999, after we had conducted detailed studies of multiparticle dynamics in CdSe NQDs (Klimov et al. 2000a). Soon after, we also realized



**Figure 5. Observation of Amplified Spontaneous Emission**

(a) This is a typical transmission electron microscopy (TEM) image of a matrix-free NQD solid film. The black dots are the semiconductor cores, whereas the space between the dots is taken up by the capping molecules. (b) The figure shows images of three CdSe NQD solid films taken under ultraviolet illumination. The films are fabricated from dots whose radii are 1.2, 1.5, and 2.1 nm. If the TOPO has a 1.1-nm length, these films have filling factors ranging from ~17% to ~26%. (c) This illustration shows our experimental setup. The cylindrical lens focuses the pump beam into a stripe on the NQD film. The ASE was detected at the edge of the film, which acted as an optical waveguide. (d) As the intensity of the pump beam increased, a sharp ASE band developed in the emission spectra of the NQD film. (Inset) The intensity of the ASE peak (circles) rose sharply once a pump laser intensity of 8 mW was reached, indicating the start of stimulated emission and optical gain (the NQD radius was 2.1 nm, and the sample temperature,  $T = 80$  K). The open squares show the sublinear dependence of the emission intensity outside the sharp ASE peak.

how to overcome this problem. Optical gain relies on the effect of stimulated emission, the rate of which can be enhanced by simply increasing the concentration of NQDs in the sample.

We estimated that the stimulated emission rate would exceed the Auger decay rate in a medium with NQD filling factors of 0.2 to 1 percent (Klimov et al. 2000b). Such densities are readily achieved in close-packed NQD films (also known as NQD solids). For example, NQDs capped with trioctylphosphine oxide (TOPO) will self-assemble into a thin film that can have filling factors as high as 20 percent, well above the estimated critical loading required for the development of stimulated emission.<sup>1</sup>

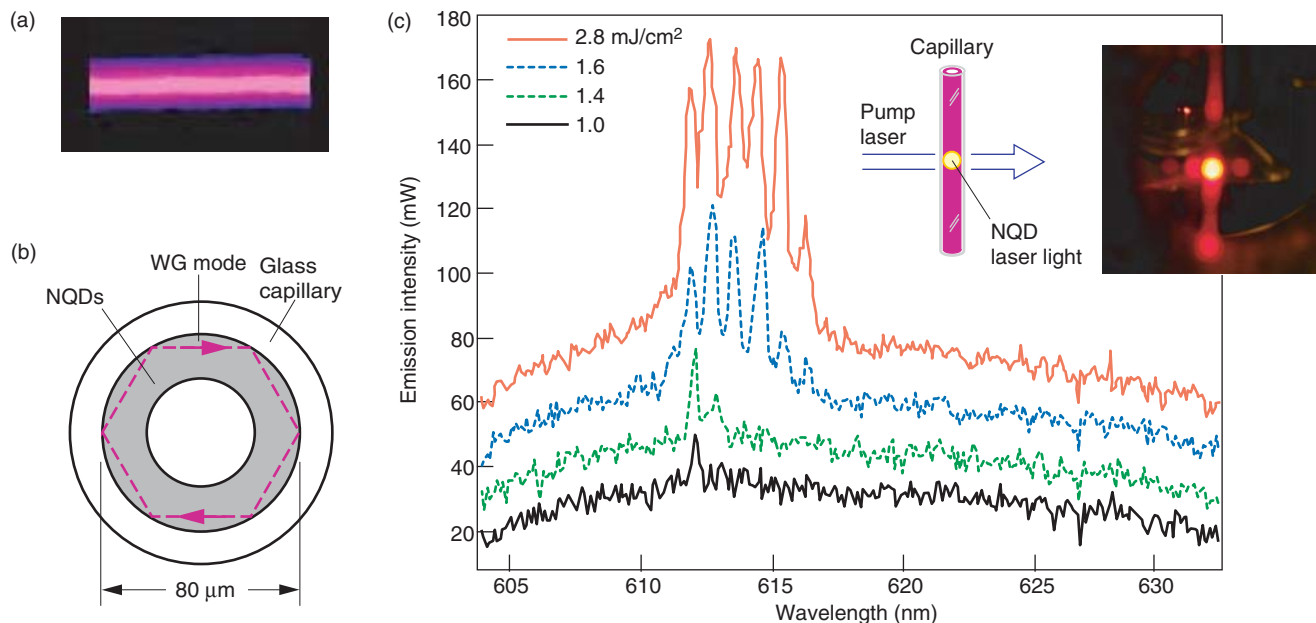
We demonstrated optical gain for

the first time by using close-packed, matrix-free films of CdSe NQDs (Klimov et al. 2000b). In these experiments, the NQD samples were optically excited by the output of an amplified titanium:sapphire pump laser (see Figure 5). At pump intensities of approximately 8 milliwatts, we observed the development of a sharp, amplified spontaneous emission (ASE) peak situated on the low-energy side of the spontaneous emission band. The dependence of this peak on the pump-laser intensity showed a threshold behavior that was a clear signature of the transition to the optical-amplification regime. We also confirmed that the frequency of the ASE peak changed with the size of

the dot. Because of strong quantum confinement, the peak in the smallest dots was blue-shifted with respect to that in bulk CdSe by more than 0.5 electron volt.

In order to slow down the NQD degradation that results when the sample heats up, we performed initial experiments at cryogenic temperatures. More recently, after improving the optical quality of our NQD solids, we were also able to demonstrate optical gain in NQDs at room temperature (Mikhailovsky et al. 2002).

<sup>1</sup>Another approach to achieving high-density NQD materials is to incorporate the NQDs into transparent sol-gel matrices. See Sundar et al. 2002 and Petruska et al. 2003 for details.



**Figure 6. NQD Lasing**

(a) This microphotograph of an NQD microcavity fabricated by incorporation of an NQD solid layer into a microcapillary tube was taken under ultraviolet illumination (the NQD layer on the inner side of the tube appears pink). (b) A cross-sectional view of the NQD microcavity illustrates an optical path of a WG mode. (c) This plot shows the development of lasing into sharp WG modes. The spectra are taken at higher and higher pump fluences. Lasing into a single, sharp WG mode develops at  $\sim 1$  mJ cm $^{-2}$ . The position of this mode (612.0 nm) corresponds to the optical gain maximum (Malko et al. 2002). As the pump fluence is further increased, additional WG modes develop on the low-energy side of the 612.0-nm mode. The insets show a schematic and photo of the laser setup.

Interestingly, the same pump fluences (number of photons per pulse per centimeter squared) that were used to excite room temperature ASE in CdSe NQDs were not sufficient to produce light amplification in bulk CdSe samples. The reason is that light amplification in bulk CdSe can be due to both low-threshold excitonic and high-threshold e-h plasma mechanisms. Excitons are bound states of e-h pairs that are “naturally” confined in space because of the Coulomb attraction between opposite charges. The e-h interaction energy in the exciton (approximately 16 meV in bulk CdSe) provides a barrier for the re-excitation of electrons and holes into the “dense” continuum of unbound e-h pair states. The existence of this “natural” barrier reduces the threshold for the “excitonic” optical gain compared to that for unbound charge carriers. However, at room temperature, excitons dissociate because of large electron thermal ener-

gies (approximately 24 meV). This process quenches the exciton-related gain and results in a significantly increased ASE threshold. Because of the large interlevel spacing in NQDs, “quantum-confined” excitons are more robust than bulk excitons, allowing one to excite room temperature ASE at pump levels comparable to those at cryogenic temperatures. This is an illustrative example of enhanced temperature stability in lasing applications expected for strongly confined dots.

In order to demonstrate true lasing action, the NQD gain medium must be combined with an optical cavity that provides efficient positive feedback. Figure 6 shows one example of a “laser fabricated in a beaker” that we made by incorporating NQD solids (Klimov et al. 2001, Malko et al. 2002) into a microcapillary tube. The cylindrical microcavity can support two types of optical modes: planar waveguide-like modes that develop along

the tube length, and whispering gallery (WG) modes that develop (because of total internal reflection) around the inner circumference of the tube. The modes propagating along the tube can only achieve the ASE regime because no optical feedback is present. The WG modes can support a true lasing action (microring lasing). After several attempts, we were able to uniformly fill the interior of the tube with the NQDs and achieved the first occurrence of NQD lasing (Klimov et al. 2001, Malko et al. 2002). Several types of cavities have since been utilized to demonstrate NQD lasing, including polystyrene microspheres (Klimov and Bawendi 2001), and distributed-feedback resonators (Eisler et al. 2002).

## Outlook

The first demonstrations of NQD lasing devices indicate a high poten-



tial for NQD materials to be new types of lasing media, characterized by wide-ranging color tunability, high temperature stability, and chemical flexibility. Thus far, we have only achieved lasing action by using a pump laser to create the population inversion in NQDs. An important conceptual challenge, however, awaits us in the area of electrical injection pumping. Currently, our lasing media consist of NQDs suspended in a non-conducting matrix, and it is not possible to excite the dots electrically.

One possible strategy to achieve electrical injection is by combining “soft” colloidal fabrication methods with traditional, epitaxial crystal-growing techniques and incorporate dots into high-quality injection layers of wide gap semiconductors. A possible technique that is “gentle” enough to be compatible with colloidal dots is energetic neutral-atom-beam epitaxy. This method utilizes a beam of neutral atoms carrying significant kinetic energy of several electron volts. The beam energy is sufficient for the activation of nonthermal surface chemical reactions, eliminating the need to heat the substrate in order to grow high-quality films for NQD encapsulation.

Because of Auger recombination, however, electrical pumping of NQD lasing devices would still be significantly more difficult than pumping of simple, “nonlasing” light emitters. Interestingly, there is a possible approach to completely eliminate Auger recombination from NQDs. It stems from the realization that the optical-gain requirement of two e-h pairs (the same initial state that allows Auger recombination to occur) is a consequence of the electron-spin degeneracy of the lowest emitting transition. Two electrons occupy the same ground state; therefore, both must be excited to achieve a population inversion. If the ground-state degeneracy could be broken (perhaps through interactions with magnetic

impurities) the gain can, in principle, be realized with a single e-h pair, and Auger decay would no longer be a problem for either optically or electrically pumped NQDs. ■

## Acknowledgments

I would like to acknowledge contributions of Alexandre A. Mikhailovsky, Jennifer A. Hollingsworth, Melissa A. Petruska, Anton V. Malko, and Han Htoon to the work reviewed here. I also thank Jay Schecker and Necia Cooper for a thorough editorial work on the manuscript. This work was supported by Los Alamos Directed Research and Development Funds and the U. S. Department of Energy, Office of Sciences, Division of Chemical Sciences.

## Further Reading

- Arakawa, Y., and H. Sakaki. 1982. Multidimensional Quantum Well Laser and Temperature Dependence of its Threshold Current. *Appl. Phys. Lett.* **40** (11): 939.
- Efros, Al. L., and A. L. Efros. 1982. Pioneering Effort I. *Sov. Phys. Semicond.* **16**: 772.
- Eisler, H.-J., V. C. Sundar, M. G. Bawendi, M. Walsh, H. I. Smith, and V. I. Klimov. 2002. Color-Selective Semiconductor Nanocrystal Laser. *Appl. Phys. Lett.* **80** (24): 4614.
- Klimov, V. I., and M. G. Bawendi. 2001. Ultrafast Carrier Dynamics, Optical Amplification, and Lasing in Nanocrystal Quantum Dots. *MRS Bulletin* **26** (12): 998.
- Klimov, V. I., A. A. Mikhailovsky, J. A. Hollingsworth, A. Malko, C. A. Leatherdale, H.-J. Eisler, and M. G. Bawendi. 2001. “Stimulated Emission and Lasing in Nanocrystal Quantum Dots.” In *Quantum Confinement: Nanostructured Materials and Devices. Proceedings of the Electrochemical Society 19*, 321. Edited by M. Cahay, J. P. Leburton, D. J. Lockwood, S. Bandyopadhyay, and J. S. Harris.
- Klimov, V. I., A. A. Mikhailovsky, D. W. McBranch, C. A. Leatherdale, and M. G. Bawendi. 2000. Quantization of Multiparticle Auger Rates in Semiconductor Quantum Dots. *Science* **287**: 1011.
- Klimov, V. I., A. A. Mikhailovsky, Su Xu, A. Malko, J. A. Hollingsworth, C. A. Leatherdale et al. 2000. Optical Gain and

- Stimulated Emission in Nanocrystal Quantum Dots. *Science* **290**: 314.
- Ledentsov, N. N., V. M. Ustinov, A. Yu. Egorov, A. E. Zhukov, M. V. Maksimov, I. G. Tabatadze, and P. S. Koplev. 1994. Optical Properties of Heterostructures with InGaAs–GaAs Quantum Clusters. *Semicond.* **28** (8): 832.
- Malko, A. V., A. A. Mikhailovsky, M. A. Petruska, J. A. Hollingsworth, H. Htoon, M. G. Bawendi, and V. I. Klimov. 2002. From Amplified Spontaneous Emission to Microring Lasing Using Nanocrystal Quantum Dot Solids. *Appl. Phys. Lett.* **81** (7): 1303.
- Mikhailovsky, A. A., A. V. Malko, J. A. Hollingsworth, M. G. Bawendi, and V. I. Klimov. 2002. Multiparticle Interactions and Stimulated Emission in Chemically Synthesized Quantum Dots. *Appl. Phys. Lett.* **80** (13): 2380.
- Murray, C. B., D. J. Norris, and M. G. Bawendi. 1993. Synthesis and Characterization of Nearly Monodisperse CdE (E = S, Se, Te) Semiconductor Nanocrystallites. *J. Am. Chem. Soc.* **115**: 8706.
- Petruska, M. A., A. V. Malko, P. M. Voyles, and V. I. Klimov. 2003. High-Performance, Quantum Dot Nanocomposites for Nonlinear Optical and Optical Gain Applications. *Adv. Mater.* **15** (7-8): 610.
- Sundar, V. C., H.-J. Eisler, and M. Bawendi. 2002. Room-Temperature, Tunable Gain Media from Novel II-VI Nanocrystal-Titania Composite Matrices. *Adv. Mater.* **14** (10): 739.
- Vandyshv, Yu. V., V. S. Dneprovskii, V. I. Klimov, and D. K. Okorokov. 1991. Lasing on a Transition Between Quantum-Well Levels in a Quantum Dot. *JETP Lett.* **54** (8): 442.

For further information, contact Victor Klimov (505) 665-8284 ([klimov@lanl.gov](mailto:klimov@lanl.gov)).



# Nanoscience and the Center for Integrated Nanotechnologies

*Donald M. Parkin*

Science and technology are on the verge of a revolution, fueled by what Dr. John Marburger, President Bush's science advisor, calls "the atom-by-atom understanding of functional matter." The revolution goes by the name of "nanotechnology," and it offers a dazzling range of possibilities for observing the functioning of living systems, modifying the functional properties of materials, and designing atomic-scale structures with entirely new properties.

In broad terms, nanotechnology researchers seek to understand and exploit systems of atoms and molecules that are partly governed by structure on length scales of 1 to 100 nanometers. For perspective, a row of 10 hydrogen atoms would span about 1 nanometer. At the nanometer scale, the boundaries between traditional scientific disciplines and realms of expertise begin to fade. Thus, the key to fulfilling the promise of nanotechnology is an integration of the traditionally separate science disciplines—physics, chemistry, materials science, and biology—coupled to a robust program in computation and engineering.

At Los Alamos, we believe that nanotechnology will be a critical component of our efforts to meet mission responsibilities in national security, threat reduction, energy, and fundamental science. It will enable improvements in chemical and nuclear sensing, high-performance military platforms, and nuclear defense systems and lead to the creation of biosensor systems that can detect emerging diseases or biothreats. Nanotechnology promises a

similar revolution in medical diagnostics and therapeutics through the development of new drug formulations and delivery methods.

The National Nanotechnology Initiative was launched in 2000 as a coordinated government program to address national nanotechnology objectives. In response to that initiative, Los Alamos and Sandia National Laboratories jointly created the Center for Integrated Nanotechnologies (CINT). Formally a Nanoscale Science Research Center of the Department of Energy Office of Science, CINT is devoted to establishing the scientific principles that govern the design, performance, and integration of nanoscale materials. It is one of five such centers throughout the country.

CINT operates as a national user facility and provides the external user community—university faculty, students, other national laboratory scientists, and industrial researchers—with no-cost, open, and peer-reviewed access to the center's capabilities and expertise. Los Alamos operates two additional national user facilities, the National High Magnetic Field Laboratory and the neutron scattering facility at the Los Alamos Neutron Science Center. (See the article "The LANSCE National User Facility" on page 138.) Both facilities will provide critical capabilities to the CINT activity.

The science capabilities that CINT will nurture and develop derive from the combined capabilities at Los Alamos and Sandia. We have focused on five such capabilities, the first involving the intersection of microbi-

ology with nanoscale materials, or what we call the "nano-bio-micro-interface." This is an especially challenging area of research.

One goal is to use nanoscale biomolecular and bio-inspired assemblies to create functional microscale or larger devices. An example might be a pathogen detection system, wherein an organic receptor (which would recognize and bind to the pathogen) couples directly to an inorganic sensor platform. Another far-reaching goal is to develop new materials whose structure, function, and assembly are inspired by natural systems. In either case, the central challenge of this capability is to gain control of the physical interface between the biomolecular and synthetic materials. We hope to develop biofunctional and biocompatible surfaces and understand how to assemble biomolecular components at interfaces. We also hope to develop new approaches to the study of biological systems, approaches that will be based on new nanoscale materials and material characterization tools.

Within the nanophotonics and nanoelectronics capability, we seek to understand and control fundamental electronic and photonic interactions in nanostructured materials. The properties of electrons in tiny bits of semiconductor material, for example, become significantly modified as the bits shrink in size to 10 nanometers or less. Appropriately termed "quantum dots," these nanoscale pieces of matter behave in some ways like bulk materials, and in others, like single

atoms. By gaining an understanding of their electronic properties, we have been able to turn quantum dots into a new type of optical-gain media, culminating in our demonstration of a color-selectable, quantum dot laser. (See the article "Nanocrystal Quantum Dots" on page 214.) In addition, we hope to develop nanostructures that are significantly more complex than existing materials, incorporating multiple constituents, finer length scales, and new three-dimensional architectures. Like the nano-bio-micro capability, the research will be strongly supported by instrument development.

In the area of complex functional materials, we will explore new materials and their routes to synthesis, materials that promote complex and collective interactions between individual nanoscale components. The approach is to investigate self-assembly processes, relevant interfacial phenomena, approaches to hierarchical organization of materials, and integration strategies to access phenomena not available in individual components. Many of these activities are realized in Los Alamos work that is devoted to synthesizing field-effect transistors from molecular crystals. Here, the basic semiconductor can be self-assembled from molecules with specific functionality, and the major challenge is to control the interface between conventional electronic materials and the molecular crystal. This work is directed toward a new generation of electronic materials that are flexible and can cover large areas.

Increasing our understanding of the mechanisms underlying the mechanical behavior of nanoscale materials and structures is the direction of the nanomechanics capability. The scientific challenges in this area are to synthesize new materials with novel mechanical properties based on tailored nanostructures. We hope to understand how structuring at the nanometer length scale influences mechanical responses

(such as energy dissipation), coupling, and nonlinearities.

Theory and simulation fit well with the experimentally focused capabilities previously described. State-of-the-art computational resources will help us address the complex, multiple length-scale problems. A key opportunity arises from the fact that oftentimes, nanostructured materials have responses that are dictated by structure length scales, which are readily accessible to computer modeling methods. Using simulations, we can explore the structure of biomolecular assemblies, nanostructured interfaces, and self-assembled nanoscale materials, as well as optical and electronic structures. For example, we want to model and simulate the mechanical properties of organic crystals in a polymer matrix, whose individual constituents have dimensions in the nanoscale. We are also performing calculations of the phosphorylation reactions catalyzed by kinase enzymes.

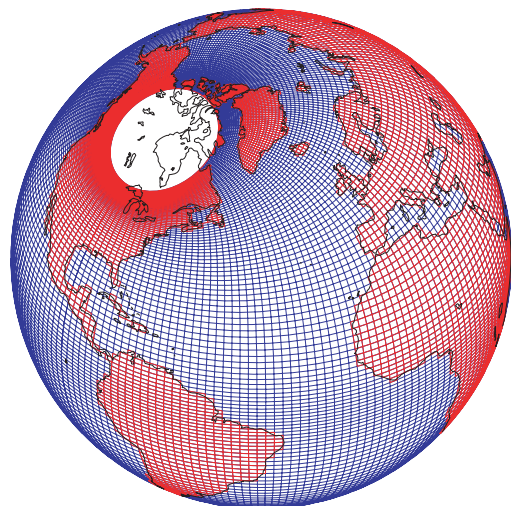
## A New Center

To provide state-of-the-art equipment and a cooperative research environment, the CINT program is currently overseeing \$76 million worth of construction that will culminate in a 96,000 square-foot Core Facility in Albuquerque, New Mexico, and a 34,000 square-foot Gateway Facility in Los Alamos, both to be in operation by 2006. Together with an existing Gateway at Sandia, these three facilities will provide space for researchers to synthesize and characterize nanostructured materials, theoretically model and simulate their performance, and integrate nanoscale materials into larger-scale systems. CINT will jumpstart its user program in 2003 before the new facilities are complete. It will leverage existing resources at Los Alamos and Sandia and carefully

build its user program in collaboration with the national and international science communities, so that by 2006 it will have both a fully operating user program and state-of-the-art facilities. The inherently multidisciplinary nature of these CINT building blocks will foster interdisciplinary research teams. It is from the integration of CINT capabilities with Laboratory and external scientist teams, new state-of-the-art facilities, and existing nanoscale science resources with grand challenge problems that we derive our name, the Center for Integrated Nanotechnologies. ■

# Eddy-Resolving Ocean Modeling

*Robert C. Malone, Richard D. Smith,  
Mathew E. Maltrud, and Matthew W. Hecht*



**W**eather forecasting has been developed into a fine art, with elaborate data collection systems feeding present conditions into detailed computer models. Despite this great effort, it appears to be impossible to predict weather for more than about two weeks in advance. Yet we hope to predict the effects of greenhouse gas emissions and other human-induced effects on climate decades to centuries into the future. This goal may, in fact, be possible because what we call climate is really the statistical “envelope” of weather events, and thus we are asking for much less detailed information than would be necessary to forecast weather.

Earth’s climate is controlled by the complex interaction of many physical systems, including the atmosphere, the ocean, the land surface, the biosphere, and in the polar regions, sea ice. To be able to predict future climate change, or at least to determine what can and cannot be predicted, we have to understand both the natural variability of the climate system and the extent to which human activities affect it. The ocean is of key importance in understanding climate, because changes in ocean circulation patterns are believed to be of primary importance in controlling climate variability on time scales of decades to centuries.

Unfortunately, realistic global ocean simulations pose a severe computational problem because the ocean contains both very small spatial scales and very long time scales compared with the atmosphere. Most of the kinetic energy in the ocean is contained in the so-called “mesoscale eddies,” whose sizes range from 10 to 300 kilometers. These eddies constitute the “weather” of the ocean. They are the oceanic equivalent of high- and low-pressure systems in the atmosphere, where the spatial scales are much larger. Weather fronts typically extend over distances of 1000 to 3000 kilometers.

Whereas the spatial scales are smaller, the time scales in the ocean are much longer than in the atmosphere. Temperature anomalies in the atmosphere persist for at most a few months (unless they are associated with longer-term anomalies in the ocean surface temperature, as occur in an El Niño event). The ocean, because of its inertia and large heat capacity, has a much longer memory. Water mass properties in the deep ocean can reflect conditions that existed at the surface hundreds of years in the past. Residence times of deep-water masses are typically several hundred years and more than a thousand years in the deep Pacific Ocean. Because of this phenomenon, the integration time

required to “spin up” an ocean model from an initial state of rest to a near-equilibrium state is several thousand simulated years.

Using the computer resources available today, it is not possible to integrate a basin- or global-scale ocean model with a resolution of about 10 kilometers (or about  $0.1^\circ$  resolution in longitude) for 1000 years or more in a reasonable amount of time. On the machines available in the United States, the global  $0.1^\circ$  simulations discussed below typically require about one week of computer time per simulated year, so a 1000-year simulation would take nearly 20 years to execute. On the Japanese Earth Simulator, currently the world’s fastest supercomputer, the same model runs more than 10 times faster. But we still need another factor of 50-to-100 increase in computing power before multicentury, eddy-resolving climate simulations become feasible, and it will likely be at least a decade before such resources become available.

Another major issue is data storage. Typical model output from a  $0.1^\circ$  global model is about 1 terabyte per simulation year, so archiving, analysis, and long-term data storage pose severe problems. Because of these limitations, ocean models that are now being used in multicentury global climate simulations have spatial resolutions ranging between  $1^\circ$  and  $4^\circ$  (or

about 100 to 400 kilometers), whereas models with resolutions of  $0.1^\circ$  (or about 10 kilometers) can run simulations of decades only.

During the past 12 years, Los Alamos has built the Climate, Ocean, and Sea-Ice Modeling (COSIM) project, with support from the Department of Energy (DOE). Our emphasis has been on the development and application of ocean and sea-ice models, but research is shifting toward fully coupled global climate modeling. In collaboration with the National Center for Atmospheric Research (NCAR), we are developing coupled climate models using low- to moderate-resolution ocean components. The NCAR community climate system model, which is the most widely used, fully coupled climate model in the United States, uses the Parallel Ocean Program (POP) model and the sea-ice model CICE, both developed at Los Alamos. These models were designed to run efficiently on parallel computer architectures and employ novel numerical algorithms that improve both the numerical efficiency and physical accuracy of the simulations. Los Alamos is also the home of the isopycnal ocean model HYCOM, which uses density instead of depth as the vertical coordinate (except in the surface mixed layer). More information on climate, ocean, and sea-ice modeling at Los Alamos is available on our web server:

<http://www.acl.lanl.gov/climate>.

A major emphasis of our research over the last decade has been to make use of the supercomputing resources at Los Alamos for very high resolution, eddy-resolving ocean simulations, albeit of relatively short duration, using the POP model. This approach is the main focus of this article. Ten to 20 simulation years is sufficient time for the model to reach a quasi-equilibrium state, where the velocity field has

### What Drives the Ocean Circulation?

The ocean circulation is driven by fluxes of momentum, heat, and fresh water at the air-sea interface. Fluxes of momentum are due to stress from the surface winds and from the movement of sea ice in polar regions. The surface wind stress is the primary driver of the upper-ocean circulation and is responsible for the major current systems, such as the midlatitude gyres with their associated strong western boundary currents (that is, the Gulf Stream off the east coast of North America and the Kuroshio Current in the Pacific off the east coast of Japan). Surface wind stress also drives the complex system of equatorial currents in the tropics, as well as the Antarctic Circumpolar Current in the Southern Ocean.

The other drivers of circulation—fluxes of heat and fresh water—are collectively known as buoyancy fluxes because they produce changes in the density of seawater, which depends on its temperature, salinity, and pressure. The surface heat flux is caused by incoming solar radiation, outgoing long-wave radiation, latent heat associated with evaporative cooling, and direct thermal transfer, also known as “sensible heat flux,” which is due to differences in air and sea-surface temperatures. Fluxes of fresh water are primarily associated with precipitation and evaporation in the open ocean but are also due to melting or freezing of sea ice in polar regions and river runoff in coastal regions. The heat flux modifies the density of seawater by altering its temperature, whereas the fresh-water flux modifies the density by changing the salinity of seawater.

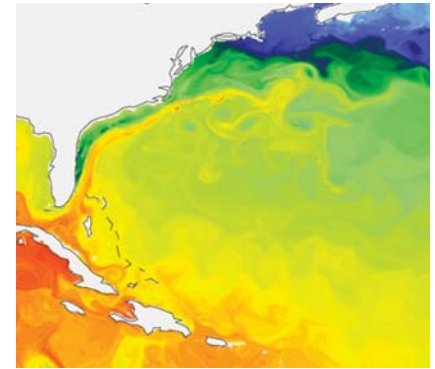
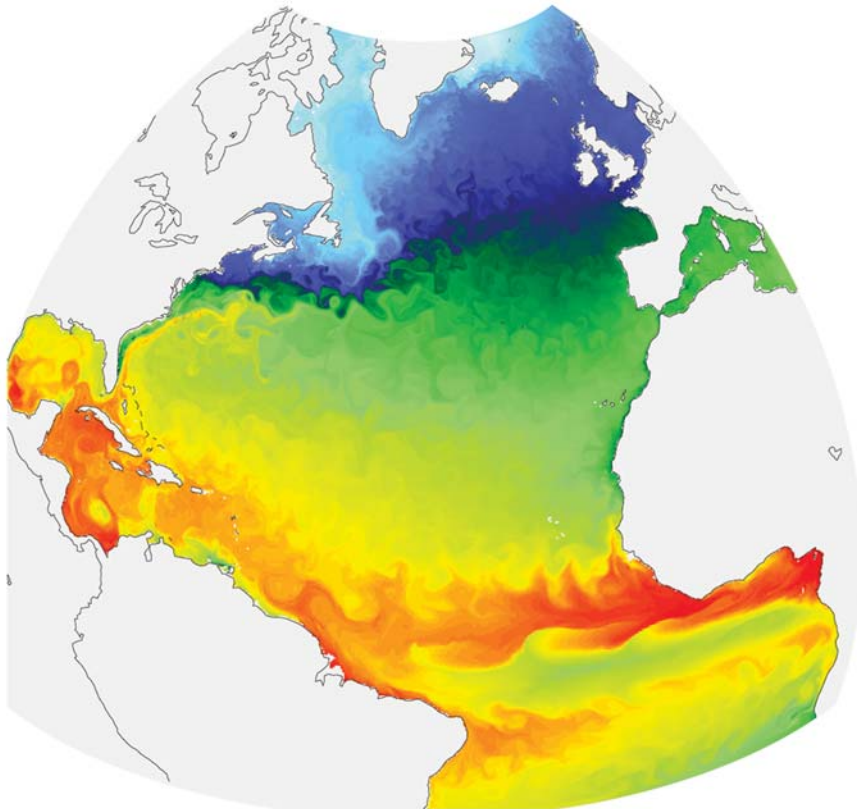
The buoyancy fluxes are the primary drivers of a circulation known as the thermohaline circulation, which is characterized by very localized sinking of dense water in subpolar regions and broad upwelling at low and mid latitudes. The thermohaline circulation is a very important factor in the earth’s climate, because it controls the transport of heat by the ocean from the tropics to high latitudes, as well as the rate of formation of deep water in subpolar regions.

adjusted to the initial density field. These short simulations are therefore appropriate for studying the dynamics of the ocean circulation on time scales of a decade or less, but they are not appropriate for studying the long-term evolution of deep-water masses or climate variability on time scales of decades and longer.

Nevertheless, the high-resolution simulations are very important for climate research since the model output provides realistic fields of turbulent statistics (such as eddy fluxes of mass and heat) that can be used to guide the

development of subgrid-scale (SGS) parameterizations for use in coarse-resolution climate simulations.

Understanding the behavior of these models will also pave the way for future eddy-resolving climate simulations. Furthermore, the model provides comprehensive three-dimensional datasets that can aid in the interpretation of the extensive observations taken over the last decade, such as high-quality satellite altimetry measurements and the variety of in situ measurements collected as part of the World Ocean Circulation Experiment (WOCE).



**Figure 1. Ocean Heat Transport**  
In Earth's climate system, the ocean and the atmosphere each contribute about half the total transport of heat from the tropics to high latitudes. The figure is a snapshot of sea-surface temperature from a  $0.1^\circ$  simulation of the North Atlantic Ocean. Red colors indicate warm water, and blue colors, cold water. The Gulf Stream, which follows the coastline of the southeastern United States, carries warm water from the tropics to high latitudes. (Inset) This magnified view focuses on the Gulf Stream. In this simulation, it correctly separates from the coast at Cape Hatteras.

### The North Atlantic Ocean at $0.1^\circ$ Resolution

Our first major simulation performed with the POP model was a global ocean simulation driven by observed surface winds for the decade 1985 to 1995 (Maltrud et al. 1998). (The ocean circulation is driven primarily by surface winds, but surface fluxes of heat and fresh water are also important. See the box on the opposite page.) This model had a horizontal resolution of  $0.28^\circ$ , corresponding to a grid spacing ranging from about 30 kilometers at the equator to about 10 kilometers at high latitudes. (The variation in grid spacing occurs because this model uses a Mercator grid, in which the grid resolution in both the north-south and east-west directions varies as the cosine of latitude. The grid spacing is shown as a function of latitude in Figure A on the next page). The

$0.28^\circ$  resolution was sufficient to allow the development of a weak eddy field, but the eddy energy was much too low compared with observations. Although it was able to reproduce many aspects of the wind-driven circulation, this simulation, like other “eddy-permitting” simulations conducted by different researchers, was unable to reproduce basic features of the mean circulation, such as the points at which western boundary currents (for example, the Gulf Stream) separate from the coastlines or the observed paths of major current systems such as the North Atlantic Current, which flows northeast along the Grand Banks east of Newfoundland. Such errors can lead to huge mismatches between modeled and observed air-sea heat fluxes and can lead to incorrect feedback in coupled models.

The reasons for the deficiencies in this and other eddy-permitting simula-

tions are still not completely understood, but detailed analysis of the global simulation compared with satellite observations (Fu and Smith 1996) clearly demonstrated the need for even higher spatial resolution, and theoretical arguments suggested a horizontal resolution of  $0.1^\circ$  or higher would be needed to capture the bulk of the energy in the turbulent mesoscale eddy field. At that time (1997), a global simulation was not feasible at this resolution, so we opted to conduct a limited-domain simulation of the North Atlantic Ocean at  $0.1^\circ$ , using a grid containing about 50 million ocean grid points. This model, also driven by observed winds, covers the period 1985 through 2000 (Smith et al. 2000). The model domain extends from  $20^\circ\text{S}$  in the South Atlantic to  $72^\circ\text{N}$ , and includes the Gulf of Mexico and the western half of the Mediterranean Sea.

Figure 1 shows a snapshot of the

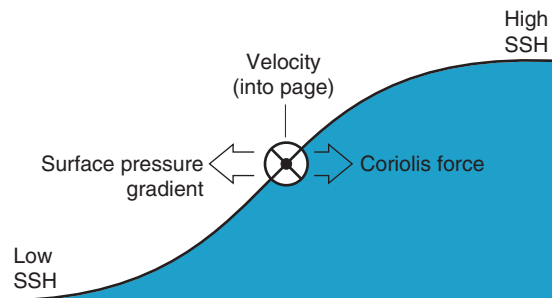
sea-surface temperature from a  $0.1^\circ$  North Atlantic simulation. The path of the Gulf Stream, which carries warm water from the tropics to high latitudes, can be clearly seen. The current follows the coastline of the southeastern United States until it separates from the coast at Cape Hatteras; from that point on, it begins to meander and pinch off warm and cold core eddies. In lower-resolution simulations, the Gulf Stream does not separate at Cape Hatteras as observed. This discrepancy has been a long-standing problem with ocean circulation models.

**Eddy Variability.** A remarkable feature of the  $0.1^\circ$  simulation is the emergence of a ubiquitous mesoscale eddy field that is substantially stronger than had been seen in previous simulations and which is, by many measures, in good agreement with observations. The eddy kinetic energy constitutes about 70 percent of the total basin-averaged kinetic energy in the North Atlantic. The model results agree well with observations of the magnitude and geographical distribution of near-surface eddy kinetic energy and sea-surface-height (SSH) variability. (Regions of strong SSH variability correspond to regions of strong, highly variable currents and turbulent flow. See the box on this page.) The model results also agree with the wave number versus frequency spectrum of surface height variations in the Gulf Stream, as well as with measurements of the eddy kinetic energy as a function of depth in the more quiescent eastern basin. The model appears to be simulating realistic values of kinetic energy over a broad range of space and time scales.

Figure 2 shows the root-mean-square (rms) SSH variability from the model, averaged over a 4-year period, as well as a recent high-quality blend of altimeter data from the TOPEX/Poseidon satellite and the satellites

### Currents, Sea-Surface Height, and Satellite Altimetry

The leading-order balance of forces in both the atmosphere and the ocean is between the Coriolis force, which is due to the earth's rotation, and horizontal pressure gradients. This state is known as geostrophic balance. The Coriolis force is proportional to the earth's rotational frequency and to the magnitude of the local current velocity, but it is directed perpendicular to the velocity (to the right in the Northern Hemisphere).



#### Figure A. Geostrophically Balanced Near-Surface Current

The pressure at a given depth is, to leading order, given by the weight of the overlying water column, which varies with the SSH. A drop in the SSH produces a horizontal pressure gradient that is balanced by the Coriolis force, which is proportional to the current velocity.

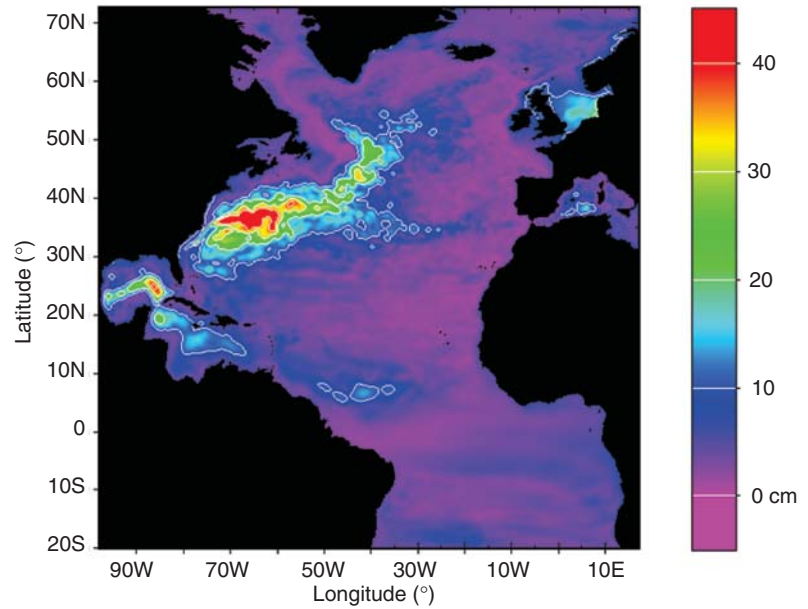
In the ocean, the Coriolis force associated with a near-surface current is in geostrophic balance with the horizontal pressure gradient because of changes in sea-surface height (SSH), as shown in the figure in this box. Thus, the near-surface pressure gradients are proportional to gradients of SSH. As a result, contours of constant SSH approximate streamlines of the near-surface flow, just as, in the atmosphere, contours of constant pressure (isobars) approximate streamlines of the winds.

In principle, an accurate map of the SSH would allow us to determine the near-surface currents. In practice, absolute measurements of SSH are difficult because the location of the sea surface in the absence of any flow is poorly known. If the ocean were at rest, the sea surface would coincide with a gravitational equipotential surface known as the geoid. Existing measurements of the geoid are not accurate enough to allow precise measurements of absolute surface height. However deviations of the SSH from the geoid can be made with much greater accuracy. Typical vertical fluctuations in the SSH associated with strong currents and eddies are about 1 to 3 meters, whereas modern satellite altimeters can measure vertical changes in SSH relative to the geoid with an accuracy of about 1 to 2 centimeters. Thus, the noise in the measurements is an order of magnitude smaller than the signal, and this situation allows very accurate measurements of the SSH variability, such as those shown in Figures 2 and 4.

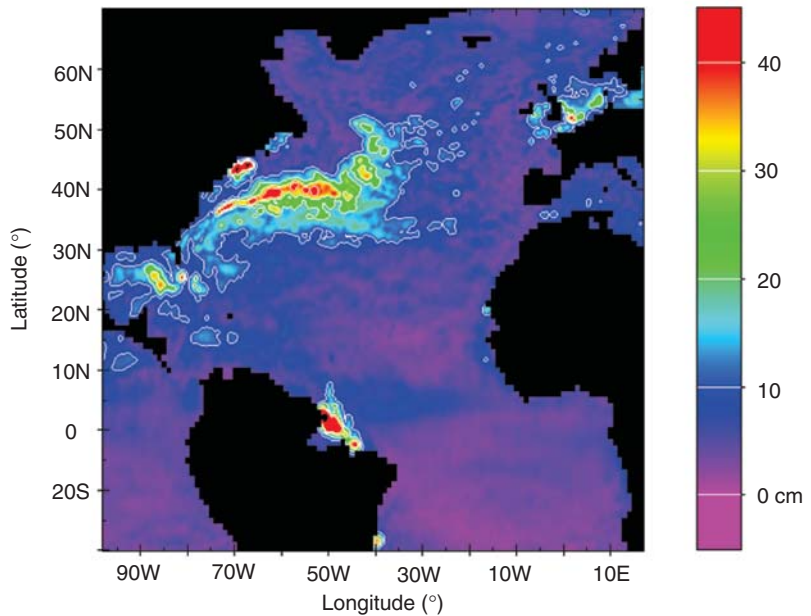
sent by the European Remote-Sensing Satellite (ERS) Programme (Le Traon and Ogor 1998, Le Traon et al. 1998). This type of satellite data has revolutionized our understanding of the world ocean, because it provides a time series of surface properties with near-global coverage (instead of, for example, a snapshot of a limited section of the ocean resulting from a series of instrument casts obtained along a research vessel cruise track). The level of agreement between model and observations evident in Figure 2 is unprecedented. It represents a milestone for both numerical ocean modeling and satellite altimetry. In fact, time series of two-dimensional fields of surface height from the model are now being used by scientists in the United States and in France to help interpret the existing satellite altimetry measurements and to aid in the development of the next generation of satellite altimeter experiments.

**Time-Mean Circulation.** Although the agreement between the model and observations in eddy variability is impressive, what is most remarkable about the  $0.1^\circ$  simulation is that the time-averaged, or time-mean, circulation exhibits several significant improvements relative to previous simulations. Figure 3 shows the time-mean SSH from the model. As discussed in the box on the opposite page, contours of constant SSH approximate streamlines of the near-surface flow, and strong currents are associated with sharp drops in SSH across these contours (that is, in stronger currents, the streamlines are “crowded together”). Major current systems such as the Gulf Stream and the North Atlantic Current are clearly visible in the figure. The Gulf Stream separates at Cape Hatteras, and its peak velocities, transports, spatial scales, and the cross-stream structure of the current are in good agreement with current-meter data. South of the

(a) SSH Variability (POP, 1998–2000)



(b) TOPEX/ERS SSH Variability (4/95–4/97)



### Figure 2. SSH Variability in the North Atlantic Ocean

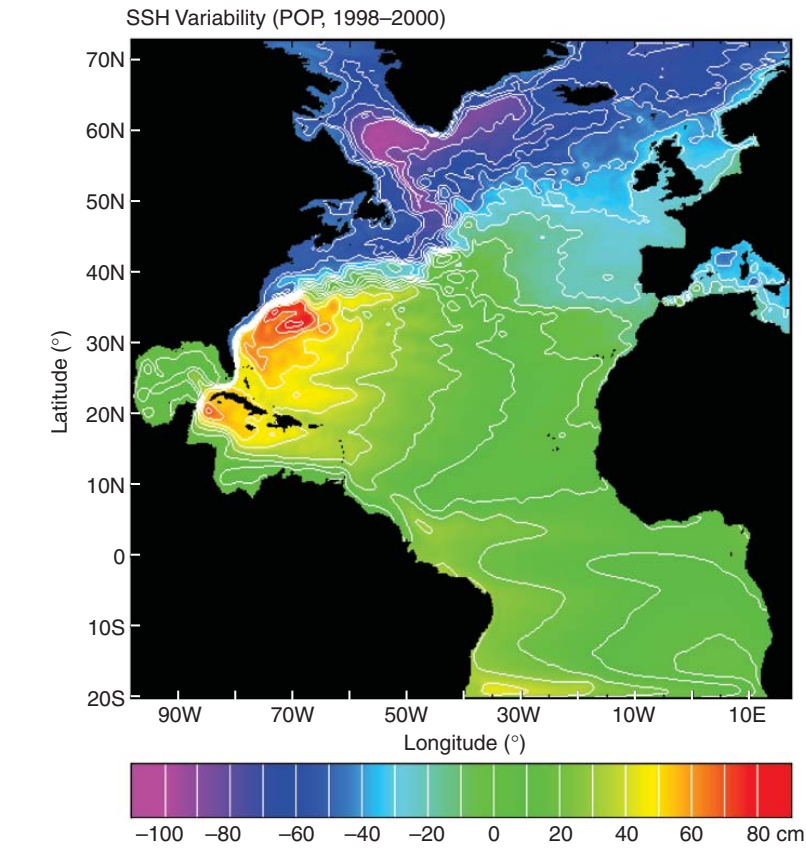
Panel (a) shows the  $0.1^\circ$  POP model simulation of SSH variability, whereas (b) shows altimeter observations derived from data from the TOPEX/Poseidon, ERS-1, and ERS-2 satellites. The SSH variability is related to mesoscale turbulence, which is generated by instabilities of the mean flow, and hence the eddy field is most intense in regions of strong western boundary currents. The height variability is most intense in the region of the Gulf Stream extension (around  $30^\circ\text{N}$  to  $45^\circ\text{N}$  latitude and  $75^\circ\text{W}$  to  $50^\circ\text{W}$  longitude) and in the vicinity of the North Atlantic Current ( $40^\circ\text{N}$  to  $50^\circ\text{N}$  and  $50^\circ\text{W}$  to  $35^\circ\text{W}$ ). Some regions of high variability that appear in the observations but not in the model (such as off the west coast of South America near the equator and off the North American coast southwest of Nova Scotia) are residual errors associated with the removal of tides from the altimetry measurements.



Grand Banks, the Gulf Stream splits into the northeast-flowing North Atlantic Current and a southward flow that feeds the Azores Current. The time-mean path of the North Atlantic Current is in good agreement with observations from float data, including the detailed positions of troughs and meanders. This is the first realistic simulation that correctly simulates the Azores Current, which flows eastward at about 35N in the central and eastern basin. Its position, total transport, and eddy variability are consistent with observational estimates. (The surface height variability for this current can be seen in Figure 2 as a tongue of high variability between 30N to 35N and 40W to 20W that appears in both model and observations.)

This simulation is by no means perfect; there are notable discrepancies with observations in some areas. For example, the Gulf Stream separates at Cape Hatteras, but its eastward path after separation is about 1.5° too far south. Nevertheless, the overall improvement in the time-mean flow relative to previous simulations indicates that we have crossed a threshold in resolution and entered a new regime of the flow that is much closer to the real circulation of the North Atlantic.

What is responsible for this regime shift? We do not yet know the complete answer to this question. It is likely that the increased resolution alone is responsible for much of the improvement. The resolution is high enough that we are able to resolve the typical length scale of the eddies (the Rossby radius) and hence capture the bulk of the energy in the eddy spectrum. (See the box on the opposite page.) The improvements in the mean circulation strongly suggest that the turbulent eddy field plays a crucial role in determining the character of the mean flow. Another contributing factor is undoubtedly the improvement in the representation of the bot-



**Figure 3. Mean SSH in the North Atlantic Ocean**

Mean SSH from a 0.1° POP model simulation. Contours of constant SSH approximate streamlines of the near-surface flow. Sharp drops in SSH across these streamlines indicate the presence of strong geostrophic currents.

tom topography. Unlike atmospheric circulation, ocean circulation is very strongly constrained by the bottom and coastal boundaries, and using the latest high-resolution data sets for ocean depth, we are much better able to represent the coastal and sea-floor topography in this high-resolution model. Another feature that changes dramatically at high resolution is that currents like the Gulf Stream become much stronger, narrower, and deeper than in the lower-resolution simulations. These deep currents in many areas reach the ocean floor (in agreement with observations) and are therefore much more strongly influenced and steered by the bottom topography. In contrast, in coarse- and moderate-

resolution models, currents like the Gulf Stream are unrealistically broad and shallow, and are not as strongly influenced by the bottom topography.

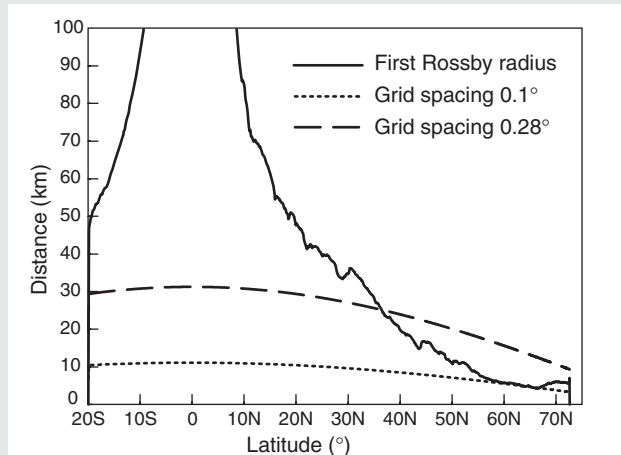
It should be emphasized that going to 0.1° or higher resolution is not in itself a guarantee that the simulation will show the same improvements we have seen. Several other modeling groups have now begun to carry out very high resolution simulations, and not all of these have had the same success. An example is the global 0.1° model discussed in the next section. Although simulations with this model do show improvements in many areas, they have so far been unable to reproduce the correct path of the North Atlantic Current, which,

### Geostrophic Turbulence: The Weather of the Ocean

Weather maps at midlatitudes show wavelike horizontal excursions of temperature and pressure contours superposed on eastward mean flows such as the jet stream. These disturbances can “pinch off” and evolve into large-scale eddies that encompass the familiar high- and low-pressure centers. Similar excursions of the mean flow are found in the ocean in eastward-flowing currents such as the Gulf Stream. These disturbances are due to an inherent instability of the midlatitude jets known as “baroclinic instability,” which occurs in the presence of strong horizontal density gradients. It is believed that baroclinic instability is the dominant mechanism for generating turbulent motion in the midlatitude jets. (Another type of instability, known as “barotropic instability,” can also generate large-scale turbulent flow in the atmosphere and ocean. This instability occurs in the presence of strong horizontal shear and is more dominant in the tropics.)

Baroclinic instability occurs in rotating, stratified fluids, with strong geostrophically balanced currents, which are associated with steeply sloping density surfaces. Turbulent energy is extracted from the potential energy of the mean flow that is stored in the sloping density surfaces of geostrophic currents. The net effect of pinching off an eddy from an eastward jet is to flatten the slope of the density surface, thus releasing potential energy. This instability is very different from the more familiar shear-flow instabilities such as the Kelvin-Helmholtz instability, in which perturbations grow by extracting energy from the mean shear flow.

A key feature of baroclinically unstable flow, which distinguishes it from most other types of turbulence, is that it has an inherent length scale, known as the “deformation radius” or “Rossby radius.” This radius is the horizontal length scale associated with unstable modes having the largest growth rate. Perturbations with wavelengths much smaller than the Rossby radius do not grow, whereas those with wavelengths much larger than the Rossby radius grow very slowly. The Rossby radius depends on the degree of stratification (or vertical density gradient) and on the local vertical component of planetary rotation. The figure shows the Rossby radius in the ocean as a function of latitude averaged over the east-west direction, computed by using a mean density field from the 0.1° North Atlantic simulation. Also shown is the horizontal grid resolution in the 0.28° and 0.1° models discussed in the text. A key feature of the 0.1° simulation is that the grid resolution is less than or equal to the Rossby radius at all latitudes. Typical mesoscale eddies have horizontal diameters that are three to 10 times larger than the Rossby radius, so the 0.1° grid is expected to allow at least marginally good resolution of the eddies at all latitudes. This fact is undoubtedly a key reason that this simulation shows substantial improvements in eddy variability.

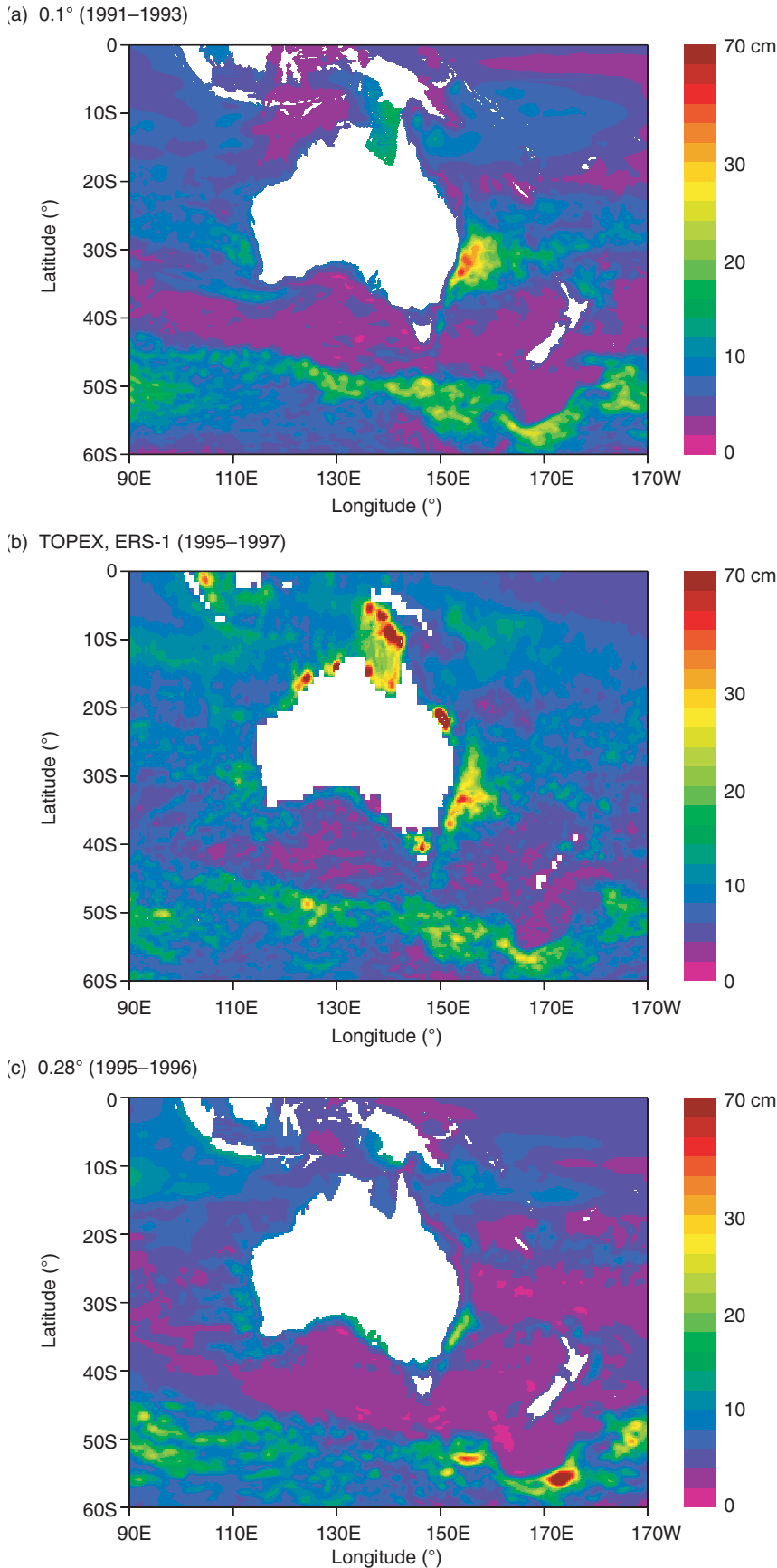


**Figure A. Zonally Averaged Rossby Radius**  
The zonally averaged Rossby radius is computed from the time-mean density of the 0.1° North Atlantic POP simulation. This radius is compared with the grid spacing of the 0.1° and 0.28° POP models.

instead of turning northeast at the Grand Banks, continues eastward across the Atlantic, as it does in lower-resolution models. This error leads to large mismatches between the modeled and observed surface

heat fluxes. We are in the process of investigating the reasons for this difference in the global and North Atlantic models.

**Sensitivity Experiments.** One thing we have discovered is that the solutions are very sensitive to the choice of SGS parameterizations of horizontal viscosity and diffusion. Initially, we had hoped that, at this



**Figure 4. SSH Variability in a 0.1° Global Simulation**

The rms SSH variability in the Southern Ocean near Australia is (a) from the global 0.1° POP model, (b) from the blended analysis of data from the TOPEX/Poseidon, ERS-1, and ERS-2 satellites, and (c) from the global 0.28° POP simulation. The agreement with observations is much better in the 0.1° model, especially in regions of strong currents such as the East Australia Current (near 30S, 155E) and the Antarctic Circumpolar Current (across the domain between 45S and 60S). The localized regions of high variability along the northern coast of Australia and south of New Guinea in the observations are residual tidal errors.

high resolution, we would simply be able to pick the coefficients of viscosity and diffusivity to be as small as possible to control numerical noise that appears on the grid scale, but that was not the case. Using the smallest possible mixing coefficients leads to unrealistic features, and the best solutions are obtained with larger values. This fact suggests that even at 0.1° resolution, we need to parameterize the effects of unresolved physical processes. We are investigating the sensitivity of the solution to different values of the mixing coefficients—as well as to different formulations of the SGS parameterizations—with a suite of new 0.1° North Atlantic simulations. We have developed novel SGS parameterizations that use horizontally anisotropic forms for viscosity and diffusivity, and we have shown that these lead to improvements in the solutions compared with the more standard isotropic forms. What we learn from these sensitivity studies in the North Atlantic model is being transferred to the more expensive global 0.1° simulations.

## Global Simulations

Spurred by the success of the 0.1° North Atlantic simulations, we have configured a 0.1° global ocean model. It uses a “displaced-pole” grid developed at Los Alamos (Smith et al. 1995), similar to the one shown in the opening graphic. Standard grids that use lines of constant latitude and longitude as coordinates have a singularity that is due to the convergence of meridians at the North Pole. The displaced-pole grid eliminates this singularity by displacing the northern grid pole into the North American continent. This grid includes the entire global ocean except for ocean points within the circle surrounding Hudson Bay. This model, containing more than 300 million grid points, is expensive to run. Both the Department of Defense (Navy) and the DOE provided computational resources that allowed the completion of a 15-year simulation. More recently, several 15-year simulations have been run on the Japanese Earth Simulator. Figure 4 shows the rms SSH variability in a section of the Southern Ocean surrounding Australia from both the 0.1° and 0.28° global models and satellite observations. As in the North Atlantic simulation (Figure 2), the agreement with observations is much better in the 0.1° model.

The immense computational resources required to run these simulations make sensitivity experiments extremely difficult, not only because of the amount of computer time involved but also because of the severe problem of archiving and analyzing the immense amount of data produced by each run. Each simulation must be carefully planned and designed. The next generation of supercomputers will make this task more tractable and allow us to move closer to the goal of a fully coupled, global climate model with an eddy-resolving ocean component. The experience we are gaining today in

our basin- and global-scale ocean simulations will pave the way for these future climate models. ■

## Acknowledgments

The 0.28° global simulations were performed in collaboration with Professor Albert Semtner, Naval Postgraduate School, Monterey, California. The 0.1° North Atlantic simulations were performed in collaboration with Dr. Frank Bryan, National Center for Atmospheric Research, Boulder, Colorado. The 0.1° global simulations are being performed in collaboration with Dr. Julie McClean, Naval Postgraduate School, Monterey, California.

For further information, contact Robert Malone (505) 667-5925 ([rcm@lanl.gov](mailto:rcm@lanl.gov)).

## References

- Fu, L.-L., and R. D. Smith. 1996. Global Ocean Circulation from Satellite Altimetry and High-Resolution Computer Simulation. *Bull. Amer. Met. Soc.* **77**: 2625–2636.
- Le Traon, P. Y., and F. Ogor. 1998. ERS1/2 Orbit Error Improvement Using TOPEX/POSEIDON: The 2 cm Challenge. *J. Geophys. Res.* **103**: 8045–8057.
- Le Traon, P. Y., F. Nadal, and N. Ducet. 1998. An Improved Mapping Method of Multisatellite Altimeter Data. *J. Atmos. Ocean. Tech.* **15**: 522–534.
- Maltrud, M. E., R. D. Smith, A. J. Semtner, and R. C. Malone. 1998. Global Eddy-Resolving Ocean Simulations Driven by 1985–1994 Atmospheric Winds. *J. Geophys. Res.* **102**: 25203–25226.
- Smith, R. D., M. E. Maltrud, F. O. Bryan, and M. W. Hecht. 2000. Numerical Simulation of the North Atlantic Ocean at 1/10°. *J. Phys. Oceanogr.* **30**: 1532–1561.
- Smith, R. D., S. Kortas, and B. Meltz. 1995. Curvilinear Coordinates for Global Ocean Models. Los Alamos National Laboratory document LA-UR-95-1146.

For more information on POP, CICE, and climate modeling at Los Alamos, including references and documentation for these models, see <http://www.acl.lanl.gov/climate>.



# Virtual Watershed

## *Simulating the water balance of the Rio Grande Basin*

*Larry Winter and Everett P. Springer*

**R**eliable supplies of clean, fresh water are essential to life and economic growth. It is not surprising then that demands for water increased dramatically during the last century as human populations grew, and energy consumption and industry expanded. As demand approaches supply, societies will become vulnerable to even minor variations in the climate and use of the land. Ironically, we now need to critically manage a resource that had almost no value less than a generation ago.

Scarce water resources *can* be managed objectively if decisions are based on the best available science and realistic computational models of complex watersheds. Detailed physics-based models, running much faster than real time on high-performance computers, can be used to test hypotheses about the performance of watersheds facing inevitable land use changes, climate change, and increased climate variability. Decision makers can use such models to evaluate management alternatives or the effects of alternate climate regimes and to support decisions about allocations of water between agriculture, ecosystems, industry, and municipalities.

Los Alamos National Laboratory and the National Science Foundation Science and Technology Center for Sustainability of Semi-Arid Hydrology and Riparian Areas are developing a high-resolution, physics-based compu-

tational model, known as the Los Alamos Distributed Hydrology System (LADHS). The model can be used to assess water resources at scales that are relevant to science and to decision makers. It is composed of four interacting components: a regional atmospheric model that is driven by global climate data, a land surface hydrology model, a subsurface hydrology model, and a river-routing model. When coupled together, these four components represent the complete hydrosphere. Our scientific and engineering goals are to retain the essential physics of all the separate components and to include realistic feedback among them. Because several alternative application codes (legacy codes) exist for each of these components, two of our key software goals are to link existing applications together with minimal code rewriting and to provide a software environment that is flexible enough to accept different alternatives.

We describe our progress in using the LADHS by means of a concrete example: quantifying the water balance of the Rio Grande Basin.

### **The Rio Grande Watershed**

The Rio Grande is a major river system in the southwestern United States and northern Mexico. Our interest is in the upper Rio Grande, which extends from headwaters in the

San Juan and Sangre de Cristo Mountains of southern Colorado to Fort Quitman, Texas (about 40 miles downstream from El Paso and Juarez), where it runs dry (see Figure 1). The upper basin covers about 90,000 square kilometers and includes the cities of Santa Fe and Albuquerque, New Mexico, and the El Paso–Juarez metropolitan area. The Rio Grande system provides water for flora, fauna, agriculture, domestic consumption, recreation, business, and industry.

Water moves through the basin along multiple natural pathways, the most important of which are precipitation, surface runoff, infiltration, groundwater recharge and discharge, and evapotranspiration, as seen in Figure 2. Spring snowmelt and summer monsoon storms are the main sources of water in the basin (Costigan et al. 2000). The northern Rio Grande and its tributaries are dominated by snowmelt runoff, but streamflow in the southern tributaries is dominated by summer rain from the North American monsoon.

The atmosphere and river discharges are the main mechanisms for transporting water out of the basin—indeed, out of any basin. Annual river flows have averaged about a million acre-feet per year in the upper Rio Grande, but variability is quite high. The basin has also been subjected to lengthy drought periods, such as the

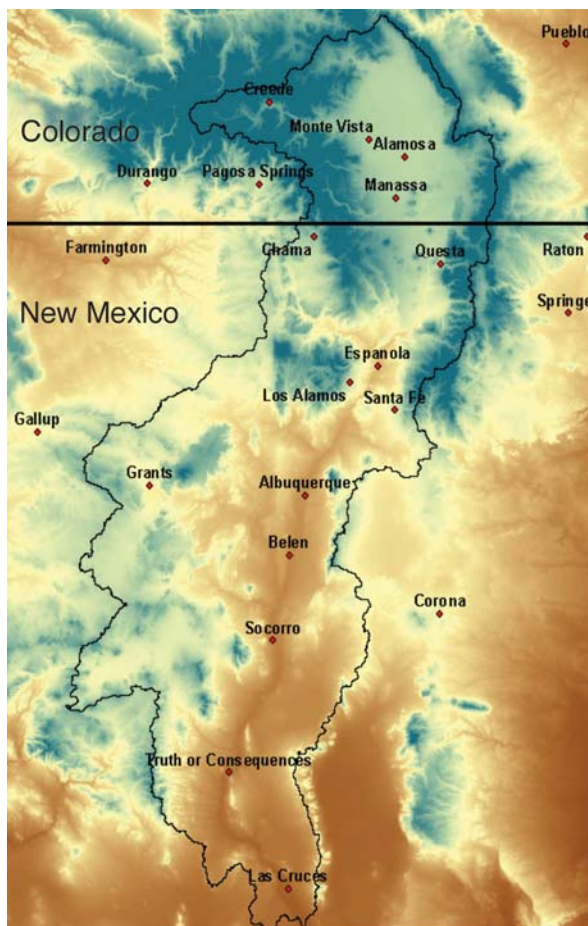
one in the 1950s that caused a rapid shift in forest and woodland zones on the Pajarito Plateau (Allen and Breshears 1998). We may be entering another such drought period now.

Apart from its land, sky, and rivers, the other major feature of the Rio Grande Basin is groundwater, which is the primary source of water for metropolitan areas. Losses from the river to the groundwater are localized, as are gains to the river from the groundwater. In some areas, streamflow is even supported by groundwater. Typically, the groundwater is recharged through mountain blocks and in streams along mountain fronts.

Increasing demands from competing uses may eventually deplete groundwater resources and affect surface-water resources. Indeed, water availability is already an important issue throughout the basin. Sustainability of water resources in the upper Rio Grande Basin requires an understanding of the conjunctive use of ground and surface water, especially groundwater recharge from different sources.

## The LADHS

Our computational approach is to link a regional atmospheric process with surface and subsurface hydrologic processes in a data flow that corresponds to regional water cycles. The detailed physics of the physical processes are summarized in Table I, along with the resolutions that we employ in our model. The flow of data through the model reflects mass and energy exchanges among the four domains in our representation of the hydrosphere.



**Figure 1. The Upper Rio Grande Basin**  
The upper Rio Grande runs from southern Colorado to the western-most tip of Texas. The black boundary defines the basin. All ground and surface water within the basin eventually flows towards the river.

Fluxes are basically driven by dissipative waves operating at different scales.

It should be noted that like every major river in the West, the Rio Grande is highly regulated; thus, the measured streamflow reflects the operation of diversion and storage dams as well as natural forces. Reservoirs and their operations are critical to determining regional effects of climate variation, because management of the water resource can alleviate or modify the impact of variability through storage and operation (Lins and Stakhiv 1998). At present, the LADHS emphasizes interactions among natural processes, although the system is mod-

ular enough to accept components representing human demands and resources.

**Regional Atmosphere.** The regional atmosphere component of our model is currently represented by the Regional Atmospheric Modeling System (RAMS). It provides precipitation, temperature, humidity, radiation, and wind data to the surface-water hydrology component. RAMS solves the Navier-Stokes equations with finite-differencing methods to estimate potential temperature, mixing ratio of water, atmospheric pressure, and horizontal and vertical components of wind (Pielke et al. 1992, Cram et al. 1992). The model consists of modules that allow for many possible configurations of parameterizations for processes such as radiation calculations and cloud microphysics. RAMS can use telescoping, interactive, nested grids to represent a large area with relatively coarse resolution and smaller areas within this domain with greater resolution. For each time step, the coarse-grid information is interpolated to the fine grid and the fine-grid variables are averaged back up to the coarse grid to provide the two-way interaction. We can enter nonstationary global climate effects into RAMS via global boundary conditions. These would be set by observed sea-surface temperatures and atmospheric fields or by output from a global climate model.

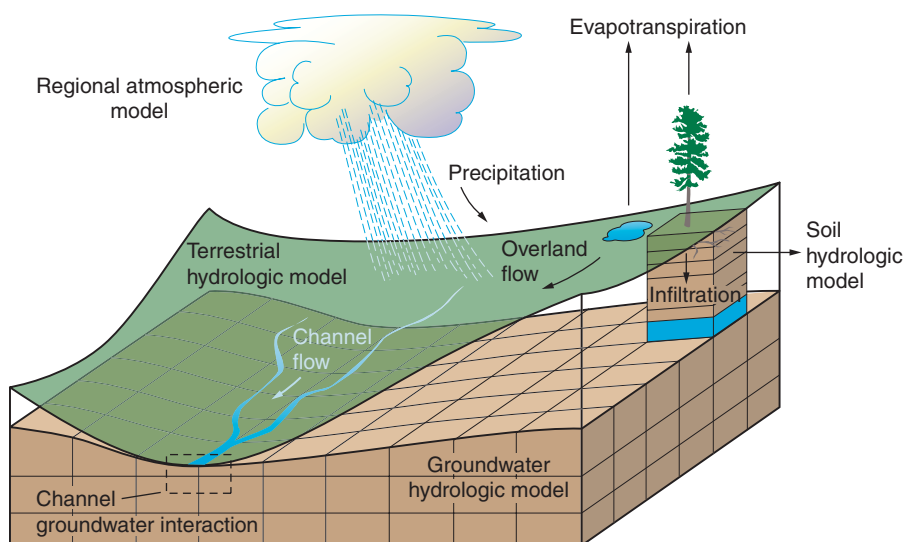
**Land Surface.** The Los Alamos Surface Hydrology (LASH) System is a grid-based water balance model (Xiao et al. 1996, Ustin et al. 1996) that represents land surface hydrology and, in particular, the hydrology of river basins. It also represents some

processes in high resolution to account for soil erosion, contaminant transport, and biogeochemical cycling. The model simulates surface and subsurface flows in two dimensions. Surface flows are routed using a diffusive wave approximation to the momentum equation with an explicit finite-difference scheme solution (Julien et al. 1995). Subsurface flow is routed using a finite-difference form of Darcy's law to determine the amount of flow between adjacent elements. The soil profile consists of two layers, plus a third if a saturated zone is present. Evapotranspiration, or the process by which plants extract water from a subsurface layer and "secrete" it through their leaves into the atmosphere, is based on the incomplete cover model presented by Ritchie (1972).

**River Routing.** Our initial approach was to use the National Weather Service's Dynamic Wave Operational Model (Fread 1988) to model how rivers and channels would flow, given our land contours, since we planned to simulate basins under natural (unregulated) flow conditions. However, those conditions do not provide the data needed by water resource managers. We are evaluating other codes for their ability to include reservoirs and dendritic drainage patterns.

**Subsurface Hydrology.** Groundwater represents a major water resource that is not included in current climate models. The Finite Element Heat and Mass (FEHM) code is a three-dimensional multiphase flow code that we use to model both the shallow subsurface aquifers and regional aquifers (Zyvoloski et al. 1997). FEHM solves mass- and energy-flow equations in a porous medium using control-volume finite elements.

So far, we have concentrated on coupling RAMS and LASH together,



**Figure 2. The Hydrologic Cycle**

A river basin is a dynamic region, with water entering and leaving along multiple natural pathways. Precipitation (primarily rain, hail, or snow) brings fresh water into the basin. The water can flow overland (surface runoff) and make its way to small channels, streams, and tributaries before becoming part of the river. Water also enters the ground, where it can flow beneath the land surface and eventually feed the river, or it can recharge (resupply) aquifers. The major process that returns water to the atmosphere is evapotranspiration, a dual process consisting of evaporation from surface areas, and transpiration, wherein plants absorb and subsequently evaporate groundwater. The LADHS couples these processes, providing a complete water balance for the river basin.

because the land surface-atmosphere interface controls most hydrologic exchanges on time scales of less than a few years. LASH requires meteorological data from RAMS, such as precipitation, temperature, wind speed, short- and long-wave radiation, and air pressure, whereas RAMS must receive evapotranspiration and related quantities from LASH. However, both RAMS and LASH are legacy codes that were not designed to be coupled to other codes. The scale and size of the data structures used by each code are different; two- and three-dimensional arrays must be exchanged; RAMS runs in a master/slave style and has a user-defined distribution of data that depends on the number of processors; and the two applications have different grid orientations.

The Parallel Applications Workspace (PAWS), developed at Los Alamos, provides a flexible software environment for connecting these separate parallel applications. PAWS can also

accept any alternate application codes we wish to incorporate into the model. A central PAWS controller coordinates communications between applications so that they can share parallel data structures, such as multidimensional arrays. Applications can have unequal numbers of processors, use different parallel data layout strategies, and be written in different languages. After the workspace is established before runtime, PAWS does not interfere with processing. The PAWS controller coordinates the creation of connections between components and data structures.

Originally developed through the DOE Accelerated Strategic Computational Initiative and Office of Science DOE 2000 Advanced Computational Testing and Simulation Toolkit, PAWS has been extended and generalized by the requirements of LADHS. New capabilities include handling multiple grid orientations and data with strides greater than 1,

**Table I. LADHS Physical Processes and Model Resolutions**

Component	Physics	Characteristic Scales	Model Resolution
Groundwater	Darcy's equation	mm-m/day	~100 m
Unsaturated subsurface	Multiphase flow	mm-cm/min	100 m
Atmosphere	Navier-Stokes equations	mm-m/s	1–5 km
Overland flow	Saint-Venant equations	cm-m/s	100 m
Snowmelt	Diffusion (heat and mass)	m/hr	100 m
Stream	Saint-Venant equations	m/s	By reach
Evapotranspiration	Diffusion	m/s	100 m

transmitting local data within guard-cell-bound memory, interacting with a master/slave component model, and using multiple communication strategies. These capabilities are also of interest to the Common Component Architecture Forum, of which the PAWS project is a member and which is working on defining standardized component interfaces for high-performance computing.

One of our next steps will be to implement in PAWS the entire LADHS—RAMS, LASH, FEHM, and river-routing applications.

### Initial Studies and Results

In our initial studies, we have been especially interested in how the spatial extent and timing of precipitation influences soil moisture, a metric that is of particular interest to farmers. We have chosen the 1992–1993 water year (October 1992–September 1993) as our test period and the northern half of the Rio Grande Basin (southwestern Colorado and northern New Mexico) as our test area. The 1992–1993 water year was an El Niño year with higher than normal precipitation in the Southwest, especially during the winter season.

Precipitation is notoriously difficult to simulate because it is highly localized. Nonetheless, its timing and extent are critical to regional and local

water budgets. Our precipitation estimates are based on high-resolution simulations using RAMS with three nested grids. The largest grid, 80 kilometers on a side, covers most of the western United States, along with parts of Canada, Mexico, and the Pacific Ocean. This grid is necessary to simulate the flow features in the region. A medium-scale grid contains the states of Utah, Arizona, Colorado, and New Mexico and has a horizontal grid spacing of 20 kilometers. Given that resolution, large terrain features, such as mountain ranges, are resolved well enough to be recognized by the model. A third grid, 5 kilometers on a side, is also used in many of the simulations to better resolve smaller terrain features.

Our initial results indicate that the RAMS model can reproduce the pronounced year-to-year variability observed in precipitation patterns across the western United States (Costigan et al. 2000). Simulated and observed monthly precipitation totals compare fairly well, although they are far from perfect (see Figure 3). In general, the 1992–1993 water year was wetter than normal, and our model had a tendency to overestimate precipitation at some high-elevation locations.

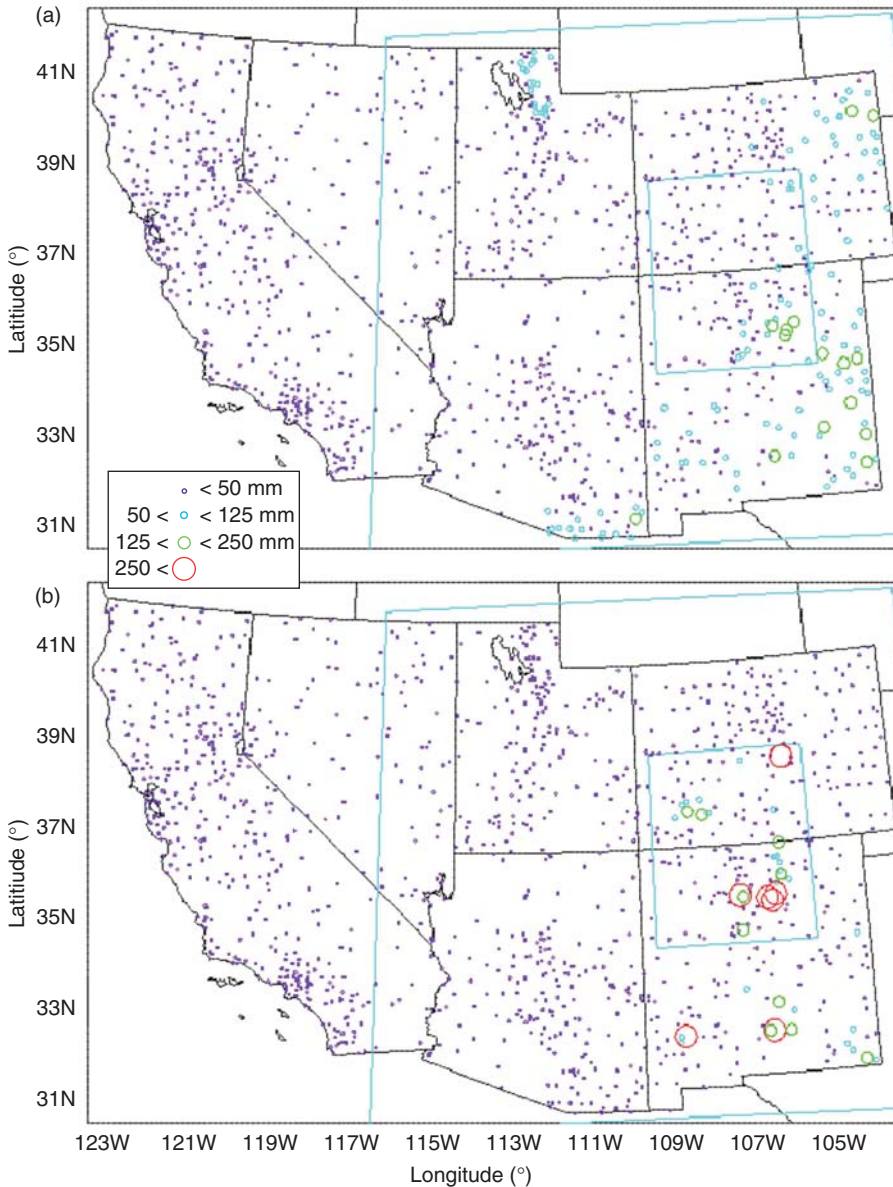
Figure 4 shows an example of output from the coupled land surface/atmosphere model, in which we simulated the effect of snow-water equivalent on soil moisture. Snow-water

equivalent is the amount of water contained in snow, and its extent is the same as the snowpack. Snow accumulation is based on the RAMS definition of snow, with snowmelt determined by temperature. It is produced by RAMS at 5-kilometer resolutions, and the blocky nature of the snow distribution in Figure 4(a) is evident. LASH operates at a much finer, 100-meter resolution. RAMS and LASH were coupled by a statistical down-scaling technique based on kriging, which is an estimation procedure used in geostatistics (Campbell 1999). The highly resolved land surface (modeled by 9,307,500 grid cells) results in a smooth, detailed map of soil moisture. That level of detail is important when simulating local processes such as soil erosion and contaminant transport.

### Conclusion

Although we cannot experiment with a system as large and valuable as the hydrosphere of the Rio Grande Basin, computer hardware and software have advanced until simulations of river basins can be highly realistic. Gaps in the data and inadequacies in coupling the components of the model are now the main limits on basin-scale simulations. In some cases, coupling is simply a matter of scaling one process to another while conserving mass and energy. In other cases, new





**Figure 3. Precipitation from RAMS**

The plots are a comparison of (a) observed data and (b) RAMS output for July 1993. The blue lines mark the approximate location of nested grid boundaries. Circles are centered on the observation sites with their size representing the accumulated precipitation (in millimeters) for the month. Model results were bilinearly interpolated to the observation sites in order to facilitate comparisons. While not perfect, the RAMS estimate of the seasonal precipitation agrees with the measured data.

science is required. This is especially true of “ecohydrology” and “agrohydrology,” where the effects of riparian areas and farming on processes like aquifer recharge and evapotranspiration must be quantified.

We also need new science to represent the impacts of municipalities and industry. Although large networks exist for observing some data, such as temperature and precipitation, they are the exception. Remote sensing, especially

satellite based, and new geological and geophysical characterization techniques may eventually fill many data gaps. However, the theory of coupled basin-scale modeling will need methods of quantifying uncertainty because no data set will ever be exact.

As human activity pushes against the margins of available water supplies, we may soon need a crystal ball to assess the effects of even small increases in demand or small variations in supply. What does a crystal ball look like? One version may be a large computer, a computational model, and a team of scientists that can apply the model and interpret the results. ■

### Further Reading

Allen, C. D., and D. D. Breshears. 1998. Drought-Induced Shift of a Forest-Woodland Ecotone: Rapid Landscape Response to Climate Variation. *Proc. Natl. Acad. Sci. U.S.A.* **95**: 14839.

Campbell, K. 1999. Linking Meso-Scale and Micro-Scale Models: Using BLUP for Downscaling. *1999 Proceedings of the Section on Statistics and the Environment*, American Statistical Association.

Costigan, K. R., J. E. Bossert, and D. L. Langley. 2000. Atmospheric/Hydrologic Models for the Rio Grande Basin: Simulations of Precipitation Variability. *Global Planet. Change* **25**: 83.

Cram, J. M., R. A. Pielke, and W. R. Cotton. 1992. Numerical-Simulation and Analysis of a Prefront Squall Line. 1. Observations and Basic Simulation Results. *J. Atmos. Sci.* **49** (3): 189.

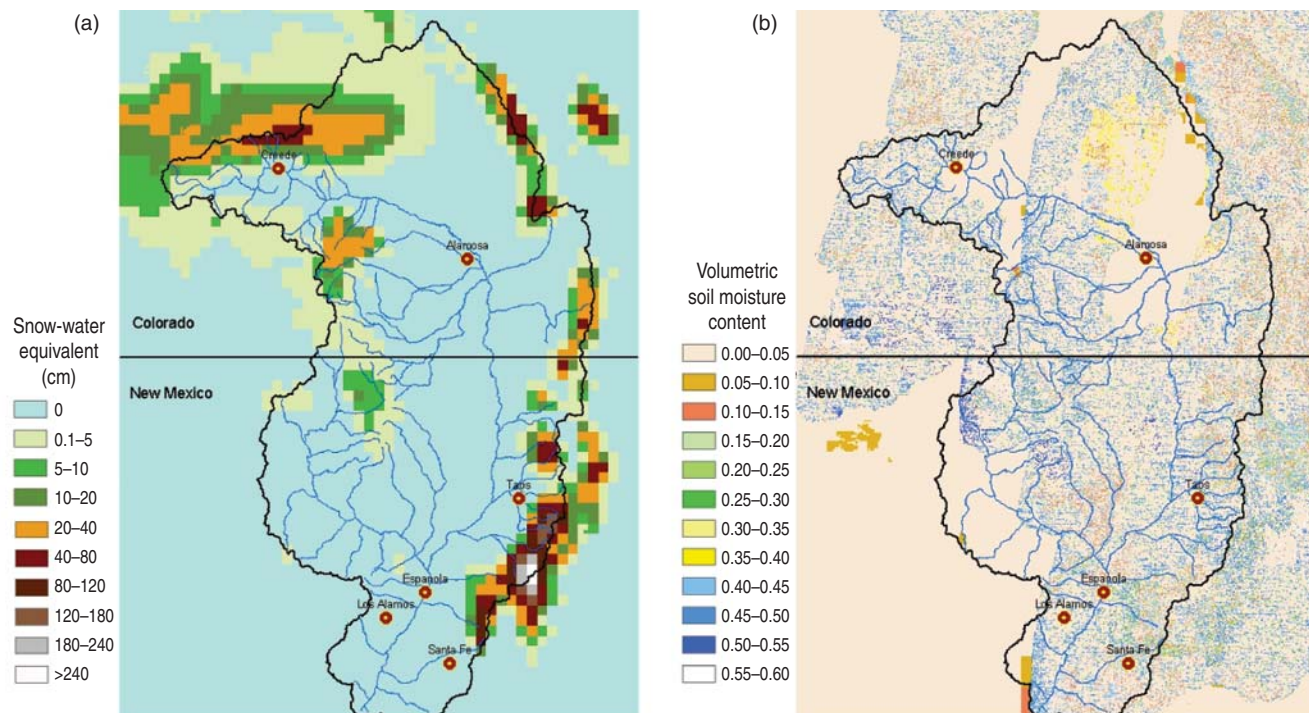
Fread, D. L. 1988. The NWS DAMBRK Model: Theoretical Background/User Documentation. National Weather Service, Office of Hydrology, Silver Spring, MD.

Julien, P. Y., B. Saghafian, and F. L. Ogden. 1995. Raster-Based Hydrologic Modeling of Spatially-Varied Surface Runoff. *Water Resour. Bull.* **31** (3): 523.

Lins, H. F.; and E. Z. Stakhiv. 1998. Managing the Nation’s Water in a Changing Climate. *J. Am. Water Resour. Assoc.* **34** (6): 1255.

Pielke, R. A., W. R. Cotton, R. L. Walko, C. J. Tremback, M. E. Nicholls, M. D. Moran, D. A. Wesley, T. J. Lee, and J. H. Copeland. 1992. A Comprehensive Meteorological Modeling System—RAMS. *Meteor. Atmos. Phys.* **49**: 69.

Ritchie, J. T. 1972. A Model for Predicting



**Figure 4. Examples of Snow Pack and Soil Moisture Results from LADHS**

Panel (a) shows the RAMS estimates of snow-water equivalent. Snow is mainly found in the San Juan and Sangre de Cristo Mountains during this October-November period. The snow distribution is not resolved very well because of the coarseness of the RAMS grid (5-km grid cells). (b) The plot shows the surface soil moisture estimates from LASH. Coupling between RAMS and LASH, which uses a finer grid (100-mm cells), smoothes the snow distribution. The distribution of soil moisture ranges from very dry in the San Luis Valley around Alamosa, Colorado, where there is little precipitation on an annual basis, to very wet conditions in higher-elevation zones where snow accumulation and melt usually occur.

- Evaporation from a Row Crop with Incomplete Cover. *Water Resour. Res.* **8**: 1204.
- Ustin, S. L., W. W. Wallender, L. Costick, R. Lobato, S. N. Martens, J. Pinzon, and Q. Xiao. 1996. Modeling Terrestrial and Aquatic Ecosystem Responses to Hydrologic Regime in a California Watershed. Chapter 6 in Status of the Sierra Nevada, Volume III Assessments, Commissioned Reports, and Background Information. Sierra Nevada Ecosystem Project Final Report to Congress, Wildland Resources Center Report No. 38, University of California, Davis.
- Xiao, Q.-F., S. L. Ustin, and W. W. Wallender. 1996. A Spatial and Temporal Continuous Surface-Subsurface Hydrologic Model. *J. Geophys. Res.* **101**: 29565.
- Zyvoloski, G. A., B. A. Robinson, Z. V. Dash, and L. L. Trease. 1997. User's Manual for the FEHM Application a Finite-Element Heat- and Mass-Transfer Code. LA-13306-M. Los Alamos, New Mexico: Los Alamos National Laboratory.

For further information, contact  
Everett Springer (505) 667-0569  
(everetts@lanl.gov).

# Water for Energy

## *A critical piece of the energy sustainability puzzle*

*Anthony Mancino and Charryl L. Berger*

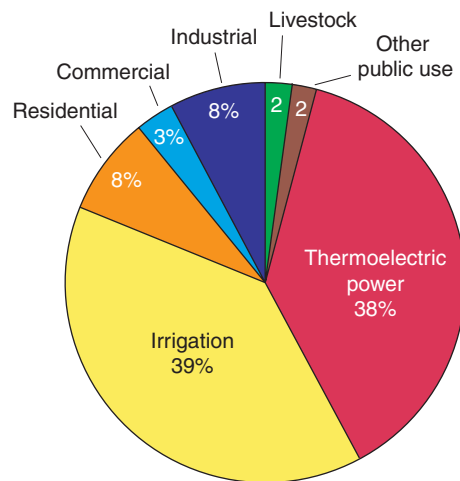
Few people need to be convinced that national security and the standard of living enjoyed in the United States depend on an abundant, reliable, and sustainable supply of electricity. Likewise, no one doubts the importance of clean and abundant water to our economy, health, and the environment. However, insufficient attention has been paid to the intimate connection between energy and water. Vast amounts of water are needed to support electricity production. Could future water shortages be an overlooked vulnerability that hinders our attempts to achieve national energy security and sustainability? And if so, what scientific and technical steps could be taken to address the problem?

Thermoelectric power production, which includes coal-fired and nuclear power plants, is second only to agricultural irrigation in fresh water withdrawals. As shown in Figure 1, irrigation and thermoelectric power generation are nearly tied in the amount of fresh water withdrawn annually (Solley et al. 1998): 134 billion gallons per day (Bgal/day) and 132 Bgal/day, respectively. Of the 132 Bgal/day withdrawn by power generation, 71 percent supports electricity generation from fossil fuels, and 29 percent supports electricity generation from nuclear power plants. These numbers reflect only the amount of cooling water withdrawn for condensing steam in steam-electric power generation. They do not include water used in any other phase of the energy cycle—such as fuel mining, refining, or

transport—nor do they include the enormous quantities of water that pass through hydroelectric plants.

The 132 Bgal/day seems alarmingly high until one accounts for the difference between withdrawal and consumption. Withdrawal is defined as the total amount of water extracted from a surface or groundwater body, whereas consumption represents the portion of withdrawal that evaporates, transpires, or becomes part of a product or crop. Irrigation and electricity generation are nearly equal in withdrawals, but irrigation consumes 81 Bgal/day, whereas power generation consumes only 3 Bgal/day.

Even though the quantity consumed in power production appears less troubling, 3 Bgal/day is not a trivial amount, and furthermore, the total amount must be available initially for U.S. power plants to continue operating as they do now. In addition, the 129 billion gallons that is returned to the source is typically 12°–30° Fahrenheit higher than the source body of water. Because the elevated temperature can harm aquatic organisms and alter the local ecosystem, strict thermal discharge limits and fish protection regulations have been imposed on power plants. Most plants already operate at the threshold of these limits. If water levels should drop, the heated discharge would raise the overall water temperature of the partially depleted lake or river beyond regulatory limits. The electric-power industry could find itself unable to keep up with electricity demand.



**Figure 1. Water Withdrawals**  
This pie chart shows U.S. fresh water withdrawals in 1995 by sector. Thermoelectric power requires nearly as much water as agriculture.

An increase in the use of renewables would greatly alleviate the reliance of electricity on water, but renewables currently account for only 2 percent of U.S. electricity whereas coal and nuclear provide 72 percent. It is unlikely that any alternative can rapidly usurp 72 percent of the current electricity infrastructure and market. In fact, coal use is projected to increase steadily over the next 20 years while nuclear generation continues at its current capacity (U.S. Department of Energy 2003). As a result, water will remain critical to meeting energy demands.

### Addressing the Problem

In searching for ways to address the issue of water for energy, we looked inward to a multidisciplinary group of scientists here, at Los Alamos National Laboratory, and outward to other national laboratories, the Electric Power Research Institute, industry representatives, and state water regulators. Based on our discussions, we believe a comprehensive solution should include the following three areas: (1) prediction and decision support, which would focus on creating a suite of decision tools that would help to identify trouble spots by analyzing “what if” scenarios, (2) tech-

nological solutions, which should focus on minimizing the effects of energy production on fresh water quantity and quality, and in particular, on investigating alternative cooling technologies, and (3) a concerted public/private partnership, because it is unlikely that the accelerated technology development and implementation suggested before will occur without it.

Decision tools would be based on coupled, high-performance computer models that link together the many complex systems and forces. (One such model is discussed in the article, "Virtual Watershed" on page 232.) The computational tools would help decision makers optimize the balance of water usage among stakeholders, guide technology investments, and aid economic development plans. The ultimate solution for thermoelectric power plants is condensing steam with a dry, air-cooled system, and such systems are already operating at a small percentage of U.S. plants. Although these systems can eliminate cooling water use by 95 percent, they are significantly more expensive to construct than wet systems and require four to six times the energy to operate. They are also much larger, taller, and louder than conventional systems, which may be of concern at certain locations (Electric Power Research Institute 2002a). Further development is necessary to decrease the cost and increase the efficiency of dry cooling systems.

Advanced drilling and pumping technology could help us access non-potable water from currently unused saline aquifers since thermoelectric power production does not require fresh water. Advanced sensing, filtration, and remediation are important as well because a large supply of contaminated water is the same as, or worse than, no water at all. By monitoring water conditions accurately and treating contamination rapidly and effectively, we can ensure that water resources remain usable and reusable.

It is also imperative that we accelerate the development and implementation of energy production methods that use less water or no water, including renewables such as solar and wind power. Hydrogen-powered fuel cells, for centralized and distributed power generation, hold great promise in the long term. They require only a small amount of water for fuel processing and no cooling water. They actually create water that can be recycled to the fuel-processing stage. The result is a net water consumption of approximately 30 gallons for every megawatt-hour (MWh) generated as opposed to the 300 and 400 gal/MWh consumed by coal-fired and nuclear plants respectively (Electric Power Research Institute 2002b). Los Alamos has been a leader in fuel-cell technology and will continue to develop robust and more efficient systems. (See the article, "Toward a Sustainable Energy Future" on page 240.)

Research and development focused on water for energy would involve long-term, high-risk investments with little near-term profit incentive, so it is unlikely that the private sector would pursue such a program aggressively. It is essential, therefore, that the federal government be involved. The complexity of the problem will require a multidisciplinary scientific and technical approach similar to the one typically employed at national laboratories.

Although the picture presented here is focused on the United States, the situation worldwide is very much the same and often worse. An increasing number of developing countries aspire toward the affluence of the United States and Western Europe, and that affluence correlates directly with the amount of energy consumed per person. As a result, global stability, which is crucial to our national security, will depend upon the same scientific and technological solutions required to achieve U.S. energy security and sustainability. Global stability will be dif-

ficult to achieve without a focused research and development effort to address the interdependencies between water and energy. ■

## Further Reading

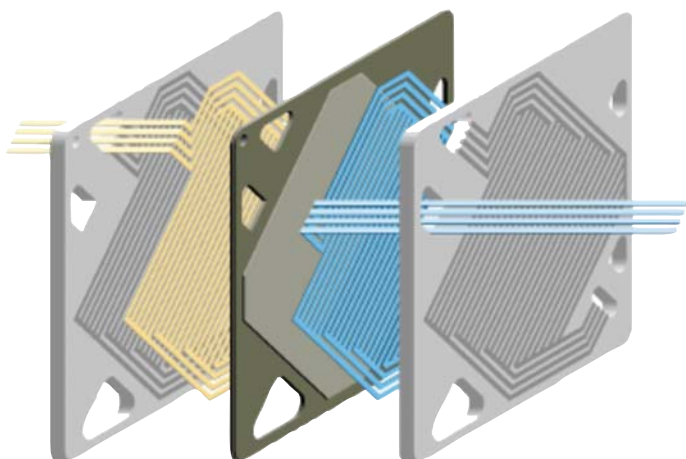
- Solley, W. B., R. R. Pierce, and H. A. Perlman. 1998. "Estimated Use of Water in the United States in 1995." U.S. Department of the Interior Geological Survey Circular 1200. Available on the Internet at <http://water.usgs.gov/watuse/pdf1995/html/>.
- U.S. Department of Energy. 2003. "Annual Energy Outlook 2003 Early Release." Energy Information Administration document AEI2003. Available on the Internet at <http://www.eia.doe.gov/oiaf/aeo/pdf/earlyrelease.pdf>.
- Electric Power Research Institute. 2002a. *Comparison of Alternate Cooling Technologies for California Power Plants: Economic, Environmental, and Other Tradeoffs*. Palo Alto, California: EPRI.
- Electric Power Research Institute. 2002b. *Water and Sustainability (Volume 3): U.S. Water Consumption for Power Production—The Next Half Century*. California Energy Commission document 1006786, Palo Alto, California.

*For further information, contact Charryl Berger (505) 667-3261 (cberger@lanl.gov).*

# Toward a Sustainable Energy Future

## *Fuel cell research at Los Alamos*

Kenneth R. Stroh



*“But in the end, my dear Cyrus, all this industrial and commercial development which you predict will continually grow, is it not in danger of coming to a halt sooner or later?... you can’t deny that one day all the coal will be used up... And what will they burn in the place of coal?”*

*“Water,” replied Cyrus Smith.*

*“Water!” exclaimed Pencroft. “Water to heat steamships and locomotives, water to heat water?”*

*“Yes, but water decomposed into its basic elements... Yes, my friends, I believe that water will one day serve as our fuel, that the hydrogen and oxygen which compose it, used alone or together, will supply an inexhaustible source of heat and light...”*

—Jules Verne, *The Mysterious Island* (1874)

About 25 years ago, and more than 100 years after Jules Verne wrote *The Mysterious Island*, scientists and engineers at Los Alamos began an R&D program into fuel cells, a program that today leads our nation and the world toward a modern version of Jules Verne’s prophetic vision. In the “hydrogen economy,” hydrogen provides the means to carry the energy trapped in coal, uranium, or wind, to our homes, cars, or offices. Fuel cells provide a clean, efficient energy-conversion technology. Together, hydrogen and fuel cells offer the promise of a sustainable energy future.

Hydrogen was first isolated as a separate element by English chemist Henry Cavendish in 1766. Called

“inflammable air” upon discovery, the colorless, odorless gas was later named for its propensity to form water on combustion in air (*hydro-gen*). Sixty-five years later, in 1839, Sir William Grove, a British jurist and amateur physicist, invented the hydrogen-oxygen fuel cell, although Christian Friedrich Schonbein (who discovered ozone and invented guncotton) had demonstrated the basic electrochemistry a year earlier. The overall reaction combines hydrogen gas ( $H_2$ ) with oxygen gas ( $O_2$ ) to form water and generate heat:



The reaction proceeds through two steps, or half reactions, each con-

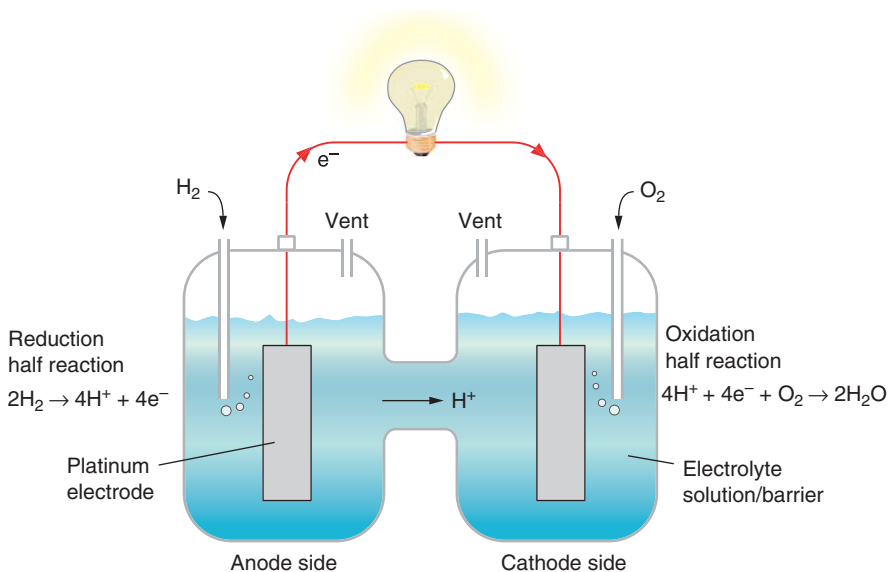
strained to take place on a different side of a reaction cell (the anode side and the cathode side). The two sides of the reaction cell are separated by an all-important electrolyte “barrier.” Hydrogen gas is fed to the anode, where a metal catalyst—typically, platinum—facilitates the breakdown of the hydrogen gas into hydrogen ions and electrons. The positively charged ions move through the electrolyte to the cathode, but the negatively charged electrons must take an external conducting path. Once at the cathode, the electrons combine with the ions and the oxygen gas—again with the help of a catalyst—to produce water (see Figure 1).

The flow of electrons through the external conduction path constitutes

an electric current that can do work. Thus, the fuel cell is a source of electric power that, like a battery, converts the chemical energy in the hydrogen fuel directly to electricity. Unlike a battery, the fuel cell has the reactants externally fed and will continue to provide full output as long as hydrogen and oxygen are supplied.

While the electrochemistry of the fuel cell was well understood years ago, enthusiasts discovered that engineering a practical device was difficult. The platinum catalyst is very expensive, and there are significant issues regarding the accessibility of gases to the electrodes, the purity of the gases, the electrolyte composition, the removal of water, and so on, all of which affect fuel-cell performance. Fuel-cell technology languished until World War II, when Francis Tom Bacon, an engineer at Cambridge University, refined the electrochemical cell and built a complete power system. Decades later, when the National Aeronautics and Space Administration (NASA) sought a compact, lightweight, reliable, and efficient power system for manned space flight, the technology really “took off.” What emerged from approximately 200 fuel-cell R&D contracts let by NASA during the Apollo program was the General Electric (GE) solid polymer electrolyte (SPE<sup>TM</sup>) fuel cell, which was used in the Gemini space capsules. GE’s technology subsequently became known as the polymer electrolyte membrane (PEM) fuel cell, and it has been the main focus of fuel-cell work at Los Alamos National Laboratory since 1977.

**The PEM Fuel Cell.** The centerpiece of the PEM fuel cell is the solid, ion-conducting polymer membrane, which replaces the liquid electrolyte. Looking much like a thick sheet of plastic food wrap, the membrane is typically made from a tough, Teflon-



**Figure 1. Fuel Cell Basics**

A fuel cell takes hydrogen gas and combines it with oxygen gas to produce electricity. The only “waste” products are water and heat. As shown in this conceptual liquid-electrolyte cell, hydrogen gas is fed to the anode side of the cell, and oxygen gas is fed to the cathode side. The two electrodes are identical, each consisting of an electrical conductor studded with platinum. At the anode, the platinum catalyzes the breakdown of hydrogen gas into hydrogen ions and electrons (reduction half reaction). The ions are conducted to the cathode by the electrolyte solution, but the electrons, which do not move through the electrolyte, flow to the cathode along the connecting wire. The platinum catalyzes the water-forming reaction involving ions, electrons, and oxygen (oxidation half reaction). Because of the separation of charge, a voltage potential of about half a volt is established between the anode and cathode.

like material called Nafion<sup>TM</sup>. This material is unusual in that, when saturated with water (hydrated), it conducts positive ions but not electrons—exactly the characteristics needed for an electrolyte barrier.

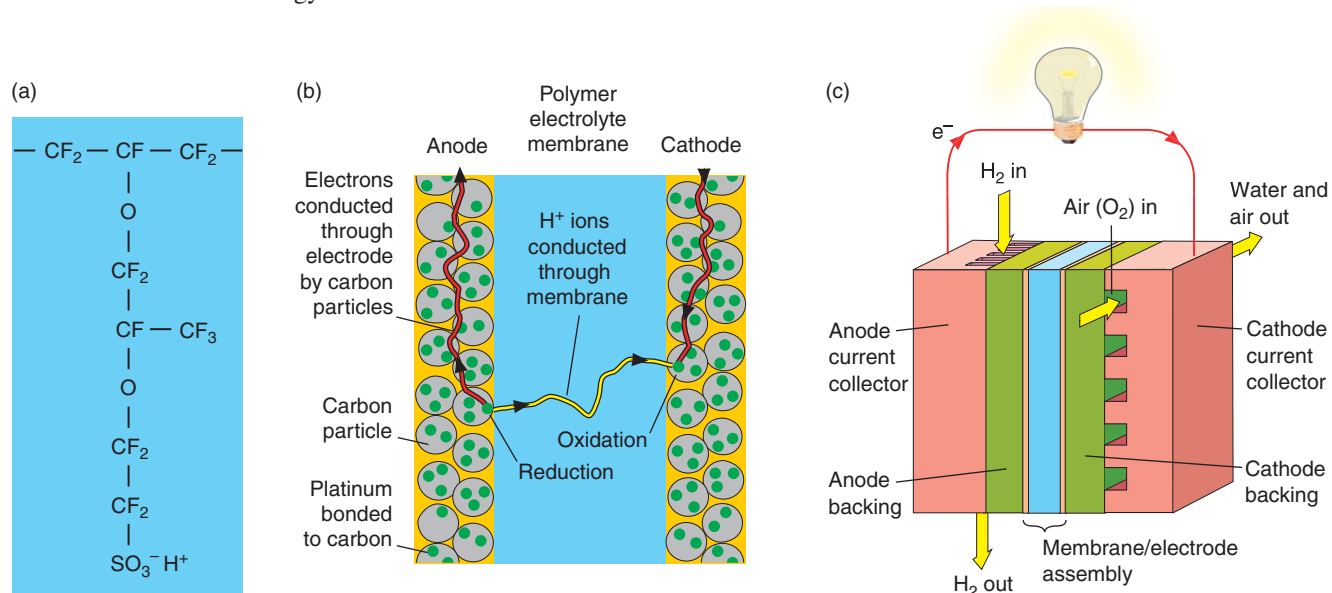
The membrane is sandwiched between the anode and cathode electrode structures—porous conducting films, about 50 micrometers thick, consisting of carbon particles that have nanometer-size-platinum particles bonded to them. Because the platinum particles have such a high surface area, the total catalytic activity of an electrode can be enormous. The electrodes are porous so that gases have ready access to the full surface area (see Figure 2).

In addition to having catalytic and electric conducting properties, the electrodes—and the backing material that supports them—are crucial to the

water management of the cell.

Ironically, even though water is a product of the fuel-cell reaction, both the hydrogen and oxygen gas streams must be humidified to keep the polymer membrane hydrated. If there is too little water, the membrane begins to lose the ability to conduct ions. However, if there is too much water, it floods the porous electrodes and prevents the gases from diffusing to the catalytically active sites. Thus, water produced at the cathode must continually be removed. Water management and the control of gas flows in and out of the cell are the keys to efficient cell operation.

Each PEM fuel cell develops a potential of about 0.4 to 0.8 volts between the anode and cathode, depending on temperature, gas pressure, flow, and other operating conditions. Higher voltage is obtained when



**Figure 2. The PEM Fuel Cell**

(a) The side chains of the Nafion™ polymer making up the membrane contain a sulfonic acid ion cluster,  $\text{SO}_3^-\text{H}^+$ . The negative  $\text{SO}_3^-$  ions are permanently attached to the side chain, but when the membrane becomes hydrated, the  $\text{H}^+$  ions can attach to water molecules and form  $\text{H}_3\text{O}^+$  ions. These ions are free to migrate from  $\text{SO}_3^-$  site to  $\text{SO}_3^-$  site, making the hydrated membrane an effective conductor of  $\text{H}^+$  ions. (b) The electrodes on each side of the membrane consist of sootlike carbon particles with nanometer-size platinum particles bonded to them. Running between the carbon particles are Nafion™ polymers to aid in ion

conductivity. The porous electrodes are pressed against the PEM. (c) The membrane and electrode assembly is sandwiched between backing layers and end plates. The porous, electrically conducting backing layers provide a way for incoming gases to diffuse uniformly to the electrodes and help in water management. The conducting end plates channel electrons to external circuits. Channels cut into the plates create a circuitous flow field for the gases and provide a way to remove water on the cathode side (see the opening graphic). This cell can be linked in series with other cells to make a higher-voltage fuel cell stack.

a number of cells are placed in series to create a “fuel-cell stack.” Because the electrochemically active cross-sectional area of each cell has a characteristic current density (typically several hundred to more than a thousand milliamperes per square centimeter, depending on operating conditions), the desired electrical current for a specific application can be achieved by changing the cross-sectional area of the stack. Through proper design and selection of operating conditions, stacks producing as much as 100 kilowatts have been achieved.

The PEM cells operate at relatively low temperatures (about 80° Celsius), a feature that makes them attractive for applications requiring multiple start-stop cycles, such as passenger vehicles. Because a fuel-cell-based electric propulsion system offers two to three times the energy efficiency of an internal combustion engine and associated drive train, the transporta-

tion sector is a driving force accelerating fuel-cell R&D. Besides energy efficiency, fuel cells offer low emissions (approaching zero if the hydrogen is made cleanly), and they help with energy and economic security because hydrogen can be made from a diverse array of domestic energy resources.

In the near term, hydrogen can be produced from natural gas, fossil fuels, biomass, or through electrolysis of water (separating  $\text{H}_2\text{O}$  into hydrogen and oxygen in a process akin to running a fuel cell in reverse). The electricity needed to run these processes would come from fossil-fuel and nuclear power plants, or from renewable resources such as the sun and the wind. In the longer term, hydrogen can be produced by use of heat from nuclear reactors and a thermochemical process to do the water electrolysis. In another concept, zero-emission coal systems could sequester

carbon in the process of separating hydrogen. Hydrogen, therefore, provides a means to carry the energy contained in coal or nuclear fuels to an end user. It is also possible to produce the hydrogen entirely from renewable resources and thus enable a truly sustainable energy future.

## Los Alamos R&D

Work at Los Alamos in the last 25 years has been enabling for the emerging fuel-cell industry, and the Laboratory holds several seminal patents required by would-be product developers. Arguably, the breakthrough that brought the PEM fuel cell out of the space program and made possible its consideration as a ubiquitous power-conversion technology was our development in the late 1980s and early 1990s of the low-platinum PEM electrodes described earlier.

Before these developments, state-of-the-art PEM “electrodes” had relatively large platinum particles embedded (basically, rammed) directly into the membrane. To maintain electron conductivity, those electrodes required very large amounts of platinum. The cost of platinum was not an issue for NASA, but it stifled any commercial applications.

Low-platinum porous electrodes had already been developed for liquid electrolyte systems, but their initial application to the dry PEM cell was unsuccessful. The sootlike carbon particles did not conduct ions very well, and in the absence of a liquid electrolyte, most ions had no conducting path to the membrane.

The Los Alamos breakthrough came when Ian Raistrick applied a solution that contained dissolved Nafion™ material to the surface of the porous electrode. Once the solution dried and the electrodes were pressed to the membrane, the Nafion™ material provided an ion-conducting path from the PEM to the platinum particles. Mahlon S. Wilson later invented methods for fabricating repeatable thin-film electrodes bonded to the PEM membrane—the so-called membrane electrode assembly (MEA). In combination, these techniques have dramatically lowered the required precious-metal catalyst loadings by a factor of more than 20 while simultaneously improving performance. They are now used by fuel-cell manufacturers and researchers worldwide.

Another Los Alamos innovation dramatically improves cell tolerance to hydrogen impurities and performance in the presence of impurities, enabling low-temperature PEM fuel cells to operate not only with pure hydrogen, but also with hydrogen-rich gas streams derived from hydrocarbon fuels (such as gasoline, methanol, propane, or natural gas). Such gas streams invariably contain trace amounts of carbon monoxide (CO), a



**Figure 3. Equipment and Testing** Brent Faulkner is pictured (photo taken around 1994) with the 10-kW electrochemical engine, a complete PEM fuel-cell power system, developed through a joint collaboration between the Laboratory and GM. The insulated components to the right are part of the methanol steam reformer that produced, on demand, hydrogen-rich gas from a liquid fuel. The two 5-kW PEM fuel cell stacks (made by Ballard) have blue endplates and can be seen at left. Additional equipment required to complete the power system is out of the photo frame.

species that poisons the catalyst by getting adsorbed and lowering the active surface area. The Los Alamos technique, invented by Shimshon Gottesfeld, injects small amounts of oxygen-containing air into the fuel stream before it enters the fuel cell. Even at PEM operating temperatures, the oxygen can oxidatively remove the CO from the catalyst surface, thus maintaining electrode surface area for the hydrogen oxidation reaction.

Other significant Los Alamos technology advances include development of processes to generate hydrogen-rich gas streams on demand from gaseous and liquid hydrocarbon fuels, significant improvement of direct methanol fuel cells (cells that run on methanol instead of hydrogen), development of fuel-cell test procedures and performance characterization methodologies, and fundamental data-supported modeling of fuel cell per-

formance. Technology transfer has been facilitated by publishing, licensing, student programs, direct training of industrial personnel, cooperative R&D agreements, and migration of our technical staff and students to industry.

## The Electrochemical Engine

Our researchers have worked closely with industry from the beginning of the Los Alamos fuel-cell program. One key example was the establishment of the Los Alamos–General Motors (GM) Joint Development Center (JDC) at the Laboratory in 1991, funded by GM and the Department of Energy (DOE). The JDC effort was focused on development of the electrochemical engine, a complete PEM fuel-cell power system fueled by methanol (shown in Figure 3), which was converted on demand to a hydrogen-rich gas by a steam-reformation process. At that time, liquid fuel was considered crucial to the acceptance of fuel-cell-powered vehicles because using liquid fuel would allow exploitation of the transportation sector’s mature fuel-distribution infrastructure.

In the steam-reforming process, methanol ( $\text{CH}_3\text{OH}$ ) reacts with water vapor in a series of controlled catalytic reactors to form a mixture of hydrogen and carbon dioxide. The carbon dioxide flows through the PEM cell with little effect, and the hydrogen is consumed in the fuel-cell reaction. The JDC team worked on all aspects of the power system: from optimizing the membrane electrode assembly to providing system integration, modeling, and testing. At the same time, a DOE-funded core research activity, in what is now the Electronic and Electrochemical Materials and Devices Group, was making enabling breakthroughs that



were soon incorporated into the electrochemical engine. Phase I of the project developed and demonstrated the complete 10-kilowatt (gross electric) electrochemical engine shown in Figure 3. Phase II extended this work to a stand-alone, 30-kilowatt (net electric) engine.

As Phase II neared its end, GM took the knowledge and expertise gained at the JDC and established a corporate fuel-cell R&D center in upstate New York. In a recent letter to the Laboratory director, the GM director of fuel-cell activities noted, "General Motors and Los Alamos have a long and successful history working together to research and develop fuel cells for automobiles. This collaboration, supported by the Department of Energy, serves as the technical foundation for the intensive development effort in fuel cells at General Motors today."

After the electrochemical engine project, the engineering research effort shifted away from methanol and toward making "stack-quality gas" from gasoline. This work was done in an effort to further reduce the fuel infrastructure barrier to commercialization of fuel-cell-powered vehicles. The fundamental fuel-cell research and development also put an increased emphasis on optimizing fuel-cell performance and on achieving durability in "gasoline reformat," the product of gasoline processing typically consisting of 40 percent hydrogen, 18 percent carbon dioxide, 30 percent nitrogen, 12 percent water, less than 10 parts per million carbon monoxide, and unspecified "other impurities." Gasoline reforming is typically accomplished by partial oxidation of gasoline to provide the endothermic heat of reaction required for subsequent process steps, including steam reforming of the remaining hydrocarbons and conversion of the residual carbon monoxide to carbon dioxide and hydrogen by reaction



**Figure 4. The Consumer Market** President George W. Bush tries out a cellular telephone powered by a DMFC, while Bill Acker of MTI MicroFuel Cells looks on. The fuel cell is in the President's hand and is connected to the phone by wire. The methanol fuel is stored in a small plastic capsule that is inserted into the cell. MTI MicroFuel Cells is commercializing Los Alamos technology under license. (White House photo by Paul Morse.)

with water (the "water-gas shift").

In 1997, a team from the Engineering Sciences and Applications' Energy and Process Engineering Group sent 2200 pounds of equipment to Cambridge, Massachusetts, by air freight to integrate a Los Alamos fuel-product cleanup system with a gasoline reformer developed by Arthur D. Little and a PEM fuel-cell stack developed by Plug Power. (Los Alamos also took a precommercial Ballard stack for testing.) To great acclaim, the "integrated" system generated the world's first electrical power from a low-temperature fuel cell operating on gasoline reformat. The Laboratory-industry team received the Partnership for a New Generation of Vehicles Medal in a 1998 ceremony at the White House.

Although technology development over the last two decades has been dramatic, PEM fuel cells are still too expensive and do not have the power density, durability, or reliability to be economically and functionally competitive with conventional power-conversion devices. Today's development

program is oriented toward reducing costs through materials substitutions, performance improvement, and system simplification and increasing durability through understanding performance degradation and life-limiting effects.

In 2003, the program direction shifted. There is now a significant focus on PEM fuel cells running on pure hydrogen stored onboard the vehicle. This change in emphasis, embodied in the president's Freedom Cooperative Automotive Research and Fuel initiatives, resulted from the desire to reduce system complexity, thereby reducing cost and improving reliability, and the desire to minimize the country's dependence on imported oil while maximizing environmental benefits. However, storing enough hydrogen onboard to enable a 350-mile driving range is challenging and is still the subject of research. Because of this change in direction, our fuel-processing effort is shifting to offboard stationary processors that would reform natural gas, propane, or liquid hydrocarbon fuels to gaseous hydrogen. There is also a growing emphasis on reforming "difficult" fuels such as diesel and Jet A (a common aviation fuel) for use in auxiliary power units for both civilian and military applications.

## The Future

Although the bulk of our funding has come from transportation programs, fuel cells are inherently scalable, and one of the earliest market introductions is likely to occur in distributed power systems. A fuel cell sitting beside a home, using reformed natural gas or propane, would provide not only electricity but also "waste" heat that could be captured and exploited for space heating and hot-water production. Such heat and power systems could convert fuel

chemical energy into useful products with close to 75 percent efficiency, far exceeding what a utility power plant can achieve. Furthermore, by placing the generating asset near the end use, we would avoid electrical transmission losses. The resulting distributed power system would be much more robust.

Still, the first large-scale commercialization of fuel cells is likely to occur in the portable electronics market, because fuel-cell power systems offer greater energy densities than batteries. The miniature fuel cells developed for this application rely on a Los Alamos development that adapted and optimized the basic PEM technology to use dilute methanol as the fuel. The methanol molecule (CH<sub>3</sub>OH) can be considered a high-energy-density hydrogen carrier. The methanol is directly oxidized at the anode in a multistep process, and protons are transported across the polymer electrolyte membrane just as they are in a hydrogen fuel cell. Although direct methanol fuel cells (DMFCs) are less efficient than hydrogen fuel cells and require more expensive catalysts, they seem a good match for hand-held electronics and small portable applications, in which a 1-watt, state-of-the-art battery can cost \$100 (or \$100,000 per kilowatt)—refer to Figure 4.

The Laboratory has also assembled an impressive intellectual-property portfolio in DMFC technology. Licensing and royalty revenues from hydrogen fuel-cell and DMFC portfolios generate about one quarter of the annual intellectual-property revenue for the Laboratory.

Citizen awareness of the concepts of energy security, economic security, and sustainability are growing. Because of increasing bipartisan political support and the continuing innovation and commitment of our world-class research scientists, engineers, and technicians, we expect that the Laboratory's contributions to solving

these complex national and global problems will only increase in the future. President Bush's budget request for fiscal year 2003 contained language to establish a fuel-cell National Resource Center at Los Alamos to address ". . . technical barriers to polymer electrolyte membrane fuel-cell commercialization."

While the designation of this National Resource Center and details of the center's work scope, operations, and funding requirements are subject to further discussion, we believe the center, if established, will focus on close collaboration with industry, universities, and other national laboratories, and will perform fundamental research to enable development of the next generation of fuel cells and related technologies, which will feature reduced cost, higher performance, and increased durability. The center will also provide resources in the form of access to the existing knowledge base, experts in the field, and state-of-the-art experimental and analytical capabilities and could provide a magnet for regional economic development. ■

### Acknowledgments

Many people have contributed to the success of this program over the last 25 years. Current team members include Peter Adcock, Guido Bender, Rod Borup, Eric Brosha, Amanda Casteel, Jerzy Chilistunoff, John Davey, Christian Eickes, Robert Fields, Fernando Garzon, Dennis Guidry, Mike Inbody, Karl Jonietz, David King, Dennis Lopez, Rangachary Mukundan, Susan Pacheco, John Petrovic, Piotr Piela, Bryan Pivovar, Geri Purdy, John Ramsey, Tommy Rockward, John Rowley, Andrew Saab, Troy Semelsberger, Wayne H. Smith, Tom Springer, Jose Tafoya, Francisco Uribe, Judith Valerio, Will Vigil, Mahlon S. Wilson, Jian Xie, Piotr

Zelenay, and Yimin Zhu. We also want to acknowledge the support of people in management, including Charryl Berger, Paul Follansbee, Ross Lemons, Tom J. Meyer, Richard Silver, and Donna Smith.

### Further Reading

- Gottesfeld, S. May 1990. "Preventing CO Poisoning in Fuel Cells," U.S. Patent No. 4,910,099.
- Raistrick, I. October 1989. "Electrode Assembly for Use in a Solid Polymer Electrolyte Fuel Cell," U.S. Patent No. 4,876,115.
- Wilson, M. S. May 1993 "Membrane Catalyst Layer for Fuel Cells," U.S. Patent No. 5,211,984.
- , August 1993. "Membrane Catalyst Layer for Fuel Cells," U.S. Patent No. 5,234,777.

**For more information**, including references, a fuel-cell primer, and an animation showing how a PEM fuel cell works, visit our website at <http://www.lanl.gov/mst/fuelcells/>.

*For further information, contact Ken Stroh (505) 667-6832 (stroh@lanl.gov).*

# Ideas That Change the World **60** years

1943 - 2003

## Highlights of the Laboratory's Anniversary Celebration



### Celebration Kickoff

Pete Nanos, then Interim Director, kicked off the anniversary celebrations with an address to the Laboratory. He reflected on national service as the sustaining role of the Laboratory since 1942. He termed the Laboratory's scientific achievements as the "gold standard for the country" and lauded the partnership and contributions of the University of California. In July 2003, the Board of Regents of the University of California confirmed Nanos as the seventh director of the Los Alamos National Laboratory.

*Historians mark the beginning* of Los Alamos National Laboratory with two dates—the initial meeting of J. Robert Oppenheimer's Scientific Committee in Los Alamos on March 6, 1943, and the signing of the first operating contract between the federal government and the University of California on April 20, 1943. The 60th anniversary of the Laboratory was commemorated with numerous activities, starting in April and concluding in September.

Planned by a task force of volunteers, the anniversary activities celebrated the Laboratory's historic contributions and accomplishments; appreciated people, communities, and institutions as enablers of the Laboratory; and anticipated future directions and challenges. The winning entry in a Laboratory-wide slogan contest provided the 60th anniversary theme, "Ideas That Change the World."

Celebrate, appreciate, and anticipate—these words sum up the mood of the Laboratory during the celebrations. We recapture that mood in these pages.

*"As our country continues to deal with security threats at home and abroad, the work that is being done at this national lab is more important today than at any other time."*

—Richard C. Atkinson,  
University of California President



(Left to right) Harold Agnew, John Hopkins, Pete Nanos, Sig Hecker, and John Browne.

### Former Directors Discussed the Lab's Scientific Accomplishments

Participating in a Director's forum were Harold Agnew (1970–79), Sig Hecker (1986–1997), and John Browne (1997–2003). John Hopkins represented Don Kerr (1979–1985). Director Pete Nanos moderated the forum, which was complemented by the

evening program "Three Decades of Directorship at Los Alamos," hosted by the Los Alamos Historical Society. During this program of public tribute, Harold Agnew received the University of California Presidential Medal for a lifetime of outstanding leadership and commitment.

Written and designed by Dennis J. Erickson and Andrea M. Gaskey

### 2002 Los Alamos Medals Awarded to Louis Rosen and George Cowan

Instituted by former Director John Browne, the medals recognize extraordinary scientific achievement. Rosen was cited for vision, leadership, and sustained contributions to nuclear science and application. Cowan was recognized for pioneering work in radiochemical techniques and for scientific leadership in the Laboratory and the community.



### Pit Manufacturing Milestone Announced

University of California President Richard Atkinson and NNSA Administrator Linton Brooks converse following the pit manufacture news conference. Director Pete Nanos joined U.S. Senator Pete Domenici, Ambassador Brooks, and President Atkinson in announcing the Los Alamos manufacture of the first nuclear weapons pit in 14 years that meets U.S. stockpile design and quality specifications. Some 700 Laboratory employees and contractors were praised for their efforts that began in 1996.



### Monte Carlo Conference Marked the 50th Anniversary of the Metropolis Algorithm

Marshall Rosenbluth gave the keynote address at a June conference held in Los Alamos to commemorate the 50th anniversary of the publication of this famous and widely applied algorithm. The article, authored by Nick Metropolis, Arianna

Rosenbluth, Marshall Rosenbluth, Mici Teller, and Edward Teller, provided the basis for the Monte Carlo method to become a powerful means to study the properties of physical systems. The *Journal of Computing and Information Science in Engineering* recently categorized the Metropolis algorithm as one of the top ten of the twentieth century.

### A Lifetime of Turbulence

Frank Harlow delivered the inaugural Heritage Lecture on the occasion of 50 years as a Los Alamos theorist. In a talk entitled "A Lifetime of Turbulence," Harlow reflected on five decades of work in dealing with the complexity of turbulence and its application through models. A renowned physicist and beloved mentor, Harlow is credited by many colleagues with giving birth to the science of computer fluid dynamics.



### Origins of Early H-Bomb Discussed at Special Classified Forum

Conrad Longmire, Richard Garwin, and Harris Mayer returned to Los Alamos to discuss their research and work, based on Edward Teller's theoretical design, that led directly to the first hydrogen bomb tested in the Los Alamos Mike event in late 1952.

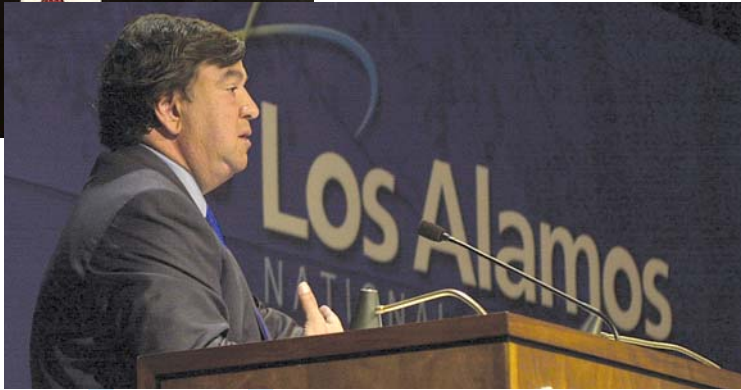




### Lab and U.C. Say, "Thank You," on Anniversary Recognition Day

On April 22, Director Pete Nanos, University of California President Richard Atkinson, and U.C. Regents Chair John Moores expressed appreciation for 60 years of sustained support and partnership to numerous distinguished guests representing federal, state, tribal, and local governments and agencies; sister labora-

tories; and academia. NNSA Administrator Linton Brooks, Sen. Jeff Bingaman (D-NM), Rep. Tom Udall (D-NM), and New Mexico Gov. Bill Richardson commented on the Laboratory's six decades of scientific achievement and continued importance. Special recognitions were extended to neighboring counties, cities, and pueblos. In addition, Nanos and Brooks reaffirmed the 1994 accord agreements with the governors of Santa Clara, San Ildefonso, Jemez, and Cochiti Pueblos.



*"The nation looks forward to your future leadership."*

—Linton Brooks, Administrator,  
National Nuclear Security Administration

# Appreciate

*"Our people make the difference. Individual excellence is absolutely critical..."*

—Pete Nanos, Director,  
Los Alamos National Laboratory

### Family Festival Celebrated the Contribution and Commitment of Extended Lab Workforce and Families

Thousands of Laboratory employees and their families were hosted at a July Saturday event at Sullivan Field in Los Alamos. The festival began with greetings from Director Nanos and a proclamation from U.C. President Richard Atkinson. Later, participants enjoyed the games, food, and entertainment activities for kids. Special U.C. funds enabled the event.



### Appreciation Extended to Neighboring Communities

Several community events reflected the Laboratory's appreciation for the support and partnership of neighbors. These events included Community Days in Santa Fe, Chamber Fest in Los Alamos, the Eight Northern Indian Pueblos (ENIP) Arts and Craft Show at San Juan Pueblo, and Spirit Day in Española. Lab presence included participation by Laboratory leaders and an information booth, which in turn featured a special display highlighting the diverse and skilled people of the Laboratory. The Lab director, for example, participated in the dedication of a new ENIP Visitors Center at San Juan Pueblo as the permanent home for the Arts and Crafts Show.





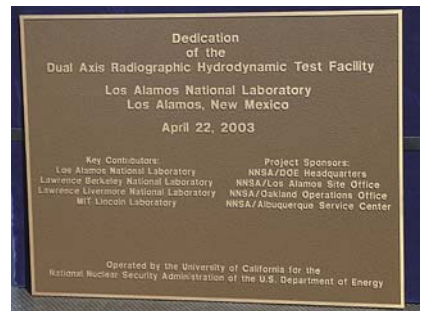
### Sen. Pete Domenici Cut Ribbon at NISC Dedication

During the Anniversary Recognition Day, Sen. Pete Domenici (R-NM), with the assistance of Associate Director Don Cobb, cut the ceremonial ribbon to dedicate the Nonproliferation and International Security Center. This impressive facility, immediately adjacent to the Nicholas C. Metropolis Center for

Modeling and Simulation, will house 400 employees. Sen. Domenici used the NISC dedication to deliver his “tough love” message, recommending DOE restructure the Lab’s operating contract and open it to bid when the current U.C. contract expires in September 2005.

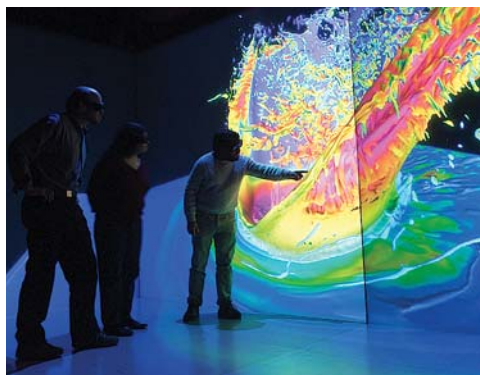
### Plaque Commemorates Completion of DARHT Facility

NNSA Administrator Linton Brooks was the senior DOE official participating at the dedication of the recently completed Dual-Axis Radiographic Hydrodynamic Test (DARHT) Facility. The state-of-the-art high-explosive firing site equipped with two intense flash-x-ray machines will be the stockpile stewardship program’s primary experimental facility for the coming decade.



### Plutonium Futures Conference Anticipated Challenges

Held in Albuquerque as the third of a series, this conference provided an international forum for research on physical and chemical properties, environmental interactions of plutonium and other actinide elements, and materials management issues. A medal symbolizing the conference, the Lab’s 60th anniversary, and the July 16, 1945, Trinity test was presented to each conference participant.



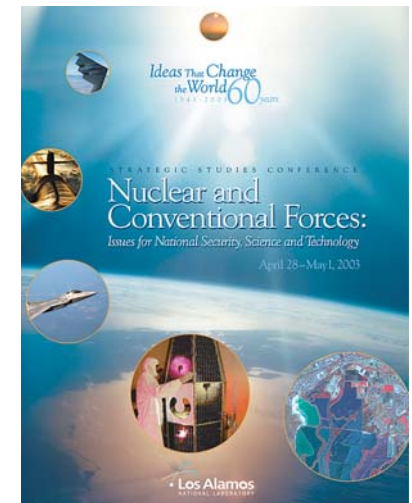
### Science Day Focused on Prediction

Featuring special talks, panel discussions, and topical sessions, the day-long event spanned institutional direction, societal contribution, and scientific accomplishment in research fields such as superconductivity, nanotechnology, sensors, biology, and actinide chemistry. Lab presenters included Deputy Director Bill Press, senior scientists, and recipients of the Distinguished Postdoctoral Research Award. Director Nanos also announced recipients of the newly created Fellows’ Prize for Outstanding Leadership in Science and Engineering.



### Issue Forums Stimulated Discussion and Surfaced Perspectives

Scientific and technical issues were discussed in different forums. An internal Laboratory forum with restricted attendance focused on “Nuclear Weapons Testing” from both policy and technical perspectives. Other forums were specifically designed for public participation. Among them were “Water, Drought, and New Mexico” and “Risk: What Does It Mean to You?,” held in Santa Fe, and “Nuclear Power in the 21st Century,” held in Los Alamos.



### Conference Focused on Nuclear and Conventional Forces

As a featured 60th anniversary event, the week-long conference “Nuclear and Conventional Forces: Issues for National Security Science and Technology” convened national and international experts who addressed emerging defense requirements given the new realities of the international security environment. The conference, part of the Los Alamos Strategic Studies program, was also used to educate the next generation of Lab leaders.

Photos: LeRoy Sanchez, Ed Vigil, Kristen Honig, Richard Robinson, and Mick Greenbank