SCIENTIFIC AMERICAN

OCTOBER 1990
$3.95

What killed the dinosaurs—a meteor or a volcanic eruption?

Light-bending crystals for optical computers.

TRENDS IN COSMOLOGY: new observations challenge theory.

***Chess computers** beat novices, then experts, now grandmasters.*
Is it only a matter of time until one defeats the world champion?

The World Championship-winning Mercedes-Benz team now uses only one make of tires: Goodyear Eagles.

A Goodyear Eagle "contact patch." Where an Eagle demonstrates its superiority.

Mercedes-Benz is known throughout the world for quality, technical innovation and engineering brilliance.

Nowhere are all of those qualities more apparent than in the highly exotic and very sophisticated Championship-winning Mercedes-Benz race cars.

These are the cars that, in 1989, won seven of the eight endurance races that constitute the FIA

SCIENTIFIC AMERICAN

October 1990 Volume 263 Number 4

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017. Copyright © 1990 by Scientific American, Inc. All rights reserved. Printed in the U.S.A. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher. Second-class postage paid at New York, N.Y., and at additional mailing offices. Authorized as second-class mail by the Post Office Department, Ottawa, Canada, and for payment of postage in cash. Subscription rates: one year $36, two years $64, three years $88 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. only 800-333-1199; other 515-247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537.

DEPARTMENTS

THE POWER OF BELIEF: #2 in a series

Who will nourish the seeds of tomorrow's technologies?

The handmaiden of every new technology is risk. There are no guidebooks to the New. ■ Success or failure often hinges on instinct, and the willingness to invest in a belief. ■ Decades ago, Motorola committed millions to an emerging technology that many dismissed as a gimmick. Today, cellular phones are used in more than 40 countries.

Motorola's EMX family of switches are now the most widely used cellular telephone switching systems in the world.

Winner 1988 Malcolm Baldrige National Quality Award

And, we produced the first hand-held cellular portable phone. In 1989, we introduced the Motorola Micro T·A·C™ Personal Telephone, shirt-pocket small, with a fraction of the parts of the original cellular phone. ■ These developments, along with others, are the product of an annual R&D investment more than twice the national average. Such is the measure of our belief.

The new, body-friendly Motorola Micro T•A•C™ Personal Telephone is the smallest cellular telephone ever made.

Building On Beliefs

MOTOROLA

THE COVER evokes the idea of chess-playing computers. In the 40 years since this magazine published its first article on computer chess, these machines have improved steadily and now routinely best grandmasters. The next step is beating the world champion. The authors of "A Grandmaster Chess Machine" (page 44) expect that day to come soon after they complete their next version of Deep Thought, now the most advanced chess system. Due to begin playing in 1992, it will be 1,000 times faster and will consider one billion positions a second.

THE ILLUSTRATIONS

Cover painting by Carol Donner

Page	Source
45	© 1989 Louie Psihoyos/Matrix
46–47	Paul Grange
48–49	Johnny Johnson
50	Courtesy of Carnegie-Mellon University
53	Carol Donner
54–60	Patricia J. Wynne
63	Roger S. Cudney, University of Southern California
64	Paul Grange
66–70	Gabor Kiss
72	R. M. Pierce, R. S. Cudney, G. D. Bacher and J. Feinberg, University of Southern California
74	Gabor Kiss
76–77	Patricia J. Wynne
79–80	George Retseck
81	Johnny Johnson (*top*), Alessandro Montanari, University of California, Berkeley (*bottom*)
82	Glen A. Izett, U.S. Geological Survey
83	Canada Center for Remote Sensing, Department of Energy, Mines and Resources
86	George Retseck (*top*), Patricia J. Wynne (*bottom*)
87	Keith Cox, University of Oxford
88	George Retseck
89	George Retseck and Patricia J. Wynne
90–92	George Retseck
95	Philip L. Richardson, Woods Hole Oceanographic Institution
96–99	Ian Worpole
101	Ian Worpole
102	Johnny Johnson
103	Hiroshi Hatanaka, Teikyo University (*top*), Johnny Johnson (*bottom*)
106	Ian Worpole
108–109	Hank Iken
110–111	P. James E. Peebles, Princeton University, and John Horgan
112	Michael Goodman
113	Edward Bell
114	Department of Astrophysics, University of Oxford
115	Lennox L. Cowie, University of Hawaii
116	Hank Iken
126	Forrest M. Mims III
127–128	Michael Goodman
129	Forrest M. Mims III

SCIENTIFIC AMERICAN®

Established 1845

SCIENTIFIC AMERICAN, INC.
415 Madison Avenue
New York, NY 10017
(212) 754-0550


MARY BRODIE WON'T FEEL THE LUMP IN HER BREAST FOR ANOTHER TWO YEARS.
Like a lot of women, Mary Brodie understands the importance of regular breast self-examination. And because she's never felt a lump, she thinks everything is fine. It's the same conclusion a lot of women reach.
Unfortunately, it's wrong.
The tiny tumor that's forming in her breast is too small to feel. But with mammography, it's not too small to see.
And tomorrow Mary is getting her first mammogram. Thanks, in part, to a new x-ray film created by Du Pont that makes it safer for women to start mammography early.
And for Mary, early detection means a two year head start on the rest of her life.
At Du Pont, our dedication to quality makes the things that make a difference.
DU PONT
BETTER THINGS FOR BETTER LIVING.

LETTERS

Big Bang over Hawaii

To the Editors:

It is unbelievable that anyone could be so insensitive to the beauty of the island of Kauai as to propose installing and operating an electromagnetic launcher at Barking Sands ["Science and the Citizen," SCIENTIFIC AMERICAN, April]. The western half of Kauai has some of the most beautiful undeveloped terrain in the world.

Operation of the electromagnetic launcher, producing sonic booms "as often as every 10 minutes," would destroy the entire western end of the island as a habitat for people and wildlife. It would be impossible to enjoy the beauty of Nuuanu Pali, the "Grand Canyon of the Pacific," with its many exciting hiking trails, or of the beach at Haena Point, or of the waterfalls and valley farms at Hanalei Bay. One might as well propose to dam the Yosemite Valley for power or to divert the waters of Niagara Falls for agriculture.

ROBERT H. MCDOUGLE
Marina Del Rey, Calif.

Light-Water Reactors

To the Editors:

I was disturbed by the offhand and casual way in which nuclear reactors were presented in "Advanced Light-Water Reactors," by Michael W. Golay and Neil E. Todreas [SCIENTIFIC AMERICAN, April], as desirable options for coping with current and future energy shortages. While no energy source is without setback or risk, I feel the authors failed to acknowledge the profound danger that constantly accompanies any nuclear power program.

An oil, coal or hydroelectric power accident can be devastating, as evidenced by the *Exxon Valdez* oil spill, but it does not approach the magnitude of a serious nuclear reactor accident. Redundant safety backups cannot guarantee protection against unforeseeable earthquakes, plane crashes and other accidents. This is to say nothing of the problem of handling radioactive wastes. Conservation and alternative sources would serve to bolster our national energy security.

THOMAS A. SIDLEY
Hollywood, Calif.

To the Editors:

Golay and Todreas make a strong case for restoring nuclear power to a central role in generating energy. We find it regrettable, however, that their survey of the various competing technologies only briefly mentions the natural uranium heavy-water reactor (HWR). That technology has been developed to maturity in Canada and now provides about 46 percent of the electric energy in Ontario. Stations are also located in New Brunswick, Argentina, Korea and elsewhere.

The neutron absorption cross section is much lower for deuterium (heavy water) than for ordinary water. Because the HWR has higher neutron efficiency, nonenriched uranium can be used, the fuel can be burned more efficiently and safer engineering features can be incorporated into the reactor's design. Generating stations using the CANDU (Canada Deuterium Uranium) design, developed by AECL Ltd., routinely lead the world in performance. As of December, 1989, five CANDU units ranked among the top 10 reactors in lifetime performance.

R. L. CLARKE
P. C. JOHNS
M. K. SUNDARESAN
Department of Physics
Carleton University, Ottawa

Golay and Todreas respond:

We argue that the U.S. nuclear technology development program should attempt to provide a set of attractive alternatives that could be employed as national needs dictate. This strategy does not exclude improving nonnuclear and efficiency technologies as well. If the worst concerns about global warming should turn out to be realistic, we shall need all the help we can get from these approaches.

Chernobyl was about as bad as any nuclear accident could be, yet it was not the most severe disaster in history. Through wise design and operations we should be able to make severe accidents infrequent enough to satisfy most reasonable people.

Our discussion of HWRs was brief, as was that of gas-cooled and liquid-metal-cooled reactors, because the topic of our article was light-water reactors. No slight to other concepts was intended. HWRs have generally compiled good operating records and demonstrated the value of exercising care in the design of power stations.

Pole Position

To the Editors:

In his article on Admiral Peary's exploration of the North Pole ["Science and the Citizen," SCIENTIFIC AMERICAN, March], Paul Wallich stated that "the actual position of the pole wanders, sometimes by as much as hundreds of meters, as the earth's mass shifts and its axis of rotation wobbles." The motion of the pole has been carefully tracked since 1899, and the total shift in nearly a century has been approximately nine meters. Although the error is not of much importance to the Peary controversy, I hate to see readers have an order of magnitude error in their feel for polar motion.

WILLIAM E. CARTER
National Oceanic and Atmospheric Administration
Rockville, Md.

Questions about Abortion

To the Editors:

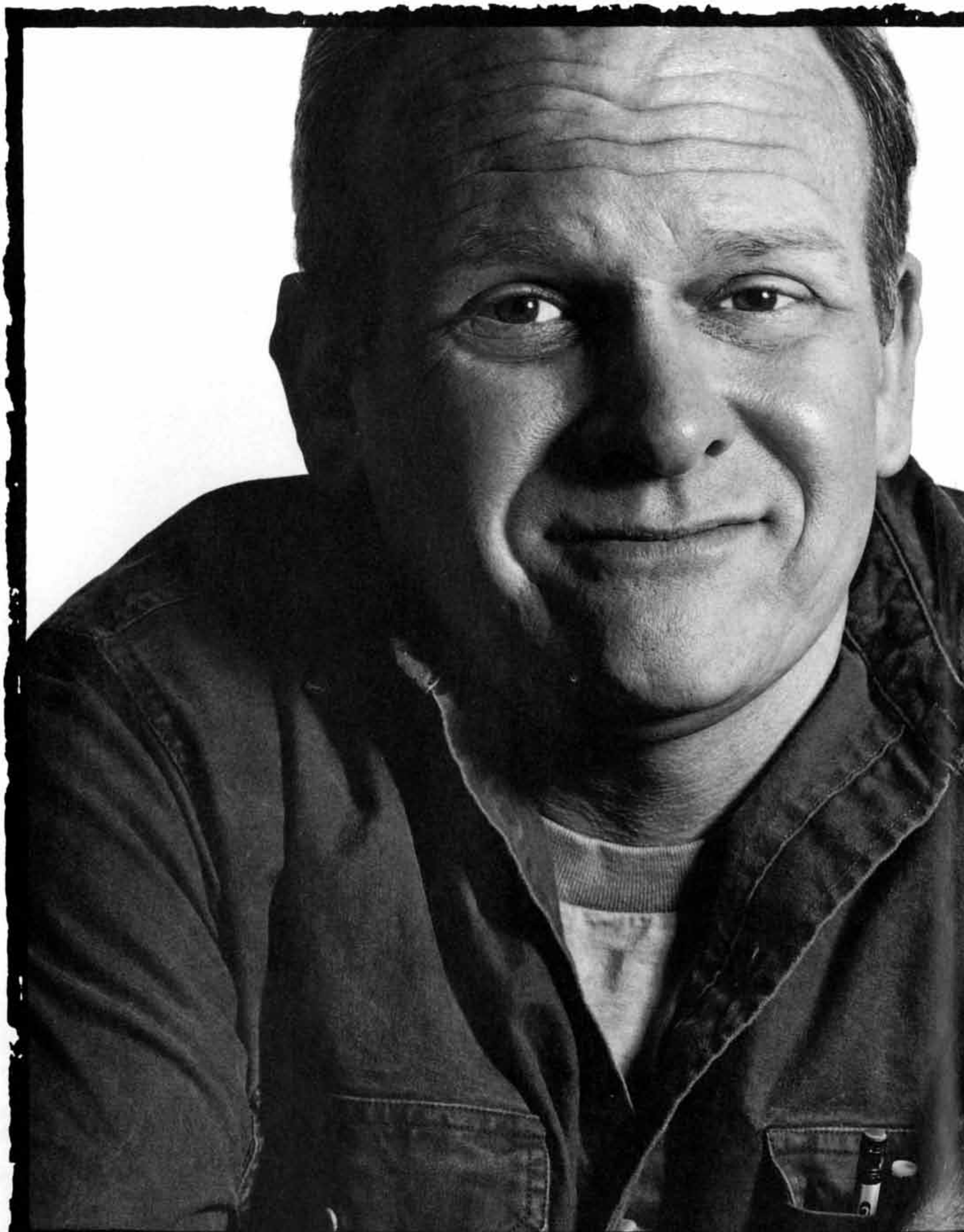
The studies mentioned in "Right to Lie?" by John Horgan ["Science and the Citizen," SCIENTIFIC AMERICAN, April], leave open the question of whether the psychological benefits of abortion are attributable to the abortion itself or to attitudes of those who elected for the procedure. It may be that those who choose to have an abortion are more motivated, which could account for their lower dropout rate from school. Are the single women who elect to carry their babies to term more likely to be under illusions about the actions of their boyfriends and families? Are they predisposed to psychological problems? Further research is needed to clarify such matters.

PAT MURTAGH
Winnipeg, Manitoba

ADDENDUM

Research for "Third World Ballistic Missiles," by Janne E. Nolan and Albert D. Wheelon [SCIENTIFIC AMERICAN, August], was conducted under the auspices of the Aspen Strategy Group.

We and our authors thank our readers for sharing their thoughts with us. Because of the great volume of mail that we receive, we are unable to answer more than a fraction of it.



"The idea of computers in the factory used to scare the daylights out of me. Now I run one."

"I figured I'd get burned either way—computers show up and I get fired, or computers don't show up and the plant closes down.

"But what happened is, they retooled the plant and while that was going on they sent me to school, to an IBM-sponsored course at the community college.

"Here are two things I learned. I learned a new job that's better than my old one. And I learned that our plant won't be boarded up any time soon."

Yes, you can teach old factories new tricks, and CIM (Computer Integrated Manufacturing) is one of them. CIM coordinates the manufacturing process, from design to distribution, as a single system. Needless to say, it can make our economy more competitive.

And yes, we'll have to teach people some new tricks, too. That's why IBM sponsors CIM education for students and workers at over 70 colleges and universities across America.

To learn more about CIM and IBM's commitment to CIM education, write to us at IBM, P.O. Box 3974, Dept. 972, Peoria, IL 61614.

IBM®

SCIENTIFIC AMERICAN
In Other Languages

LE SCIENZE
L. 5,000/copy L. 50,000/year L. 65,000/(abroad)
Editorial, subscription correspondence:
Le Scienze S.p.A., Piazza della Repubblica, 8
20121 Milano, Italy
Advertising correspondence:
Mondadori Pubblicità S.p.A., Via Rivoltana 8
20090 Segrate, Italy

サイエンス
Y 980/copy Y 10,753/year Y 13,393/(abroad)
Editorial, subscription, advertising correspondence:
Nikkei Science, Inc.
No. 9-5, 1-Chome, Otemachi
Chiyoda-ku, Tokyo, Japan

INVESTIGACION Y CIENCIA
500 Ptas/copy 5500 Ptas/year 6200 Ptas (abroad)
Editorial, subscription, advertising correspondence:
Prensa Cientifica, S.A.,
Viladomat, 291 6º 1ª
08029 Barcelona, Spain

POUR LA SCIENCE
29 FF/copy 296FF/year 350FF/year (abroad)
Editorial, subscription, advertising correspondence:
Pour la Science S.A.R.L.,
8, rue Férou,
75006 Paris, France

Spektrum DER WISSENSCHAFT
10.50DM/copy 105.60DM/year 115.20DM/(abroad)
Editorial, subscription correspondence:
Spektrum der Wissenschaft
Verlagsgesellschaft mbH
Moenchhofstrasse 15
D-6900 Heidelberg
Federal Republic of Germany
Advertising correspondence:
Gesellschaft für Wirtschaftspublizistik GWP mbH
Kasernenstrasse 67
D-4000 Duesseldorf,
Federal Republic of Germany

科学
3.80RMB/copy 45.60RMB/year $48/(abroad)
Editorial subscription correspondence:
ISTIC-Chongqing Branch, P.O. Box 2104,
Chongqing, People's Republic of China

В МИРЕ НАУКИ
2R/copy 24R/year $78.50/(abroad)
Editorial correspondence:
MIR Publishers
2, Pervy Rizhsky Pereulok
129820 Moscow U.S.S.R.
Subscription correspondence:
Victor Kamkin, Inc.
12224 Parklawn Drive
Rockville, MD 20852, USA

TUDOMÁNY
98Ft/copy 1,176Ft/year 2,100Ft/(abroad)
Editorial correspondence:
TUDOMÁNY
H-1536 Budapest, Pf 338
Hungary
Subscription correspondence:
"KULTURA"
H-3891 Budapest, Pf. 149
Hungary

مجلة العلوم
1KD/copy 10KD/year $40/(abroad)
Editorial, subscription, advertising correspondence:
MAJALLAT AL-OLOOM
P.O. BOX 20856 Safat,
13069 – Kuwait

Advertising correspondence all editions:
SCIENTIFIC AMERICAN, Inc.
415 Madison Avenue
New York, NY 10017
Telephone: (212) 754-0550 Telex: 236115

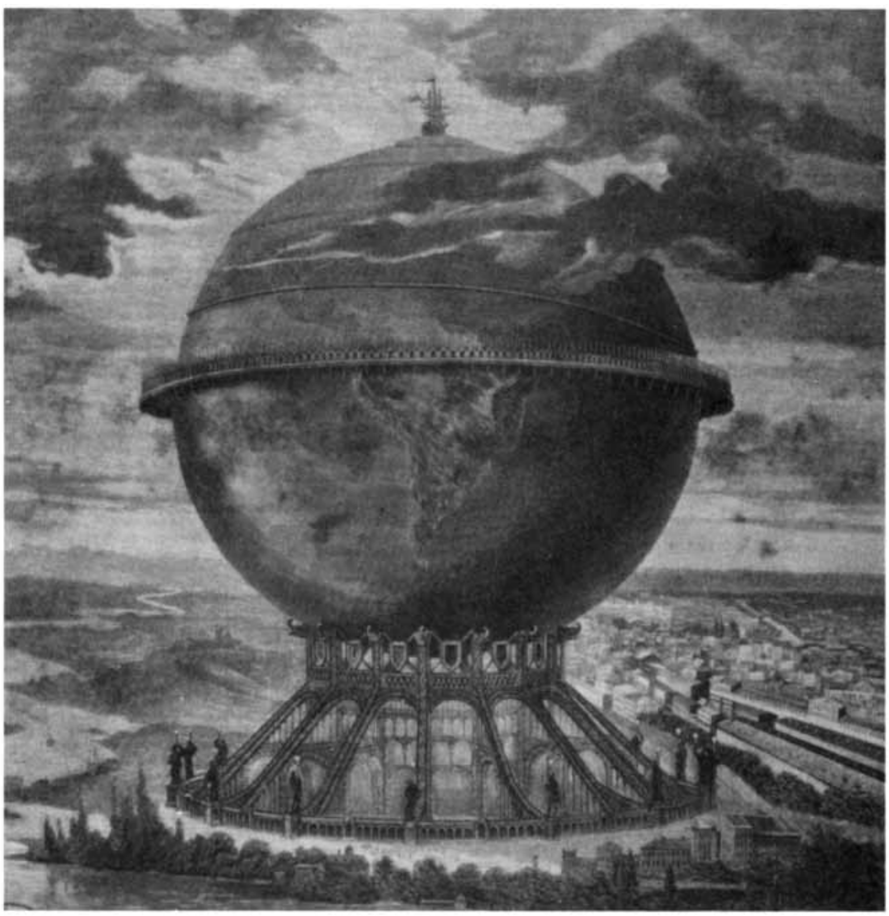
50 AND 100 YEARS AGO

OCTOBER, 1940: "When the editor of SCIENTIFIC AMERICAN requested this article, he suggested the title 'Bombers *versus* Battleships.' It would be about as easy and conclusive to present a discussion upon the subject of brothers *versus* sisters or, to select an analogy closer to the question, infantry *versus* tanks, whereas the self-evident fact is that we need both. We in the Navy consider the 'bomber *versus* battleship' question an abstract argument."

"After 1850, world temperatures rose slowly and irregularly for several decades, but since 1920 this rise has become more emphatic. And with this recent warming-up has come a sharp turn back toward despotic forms of government all over the earth. Man has felt weaker. With the cold of the last century came a marked increase in world population and such an abundance of energy and inventive genius as had never before been witnessed in human history. Since 1920, however, population increase has shown signs of tapering off, with statesmen of several countries worrying about the possibility of an actual decline in numbers."

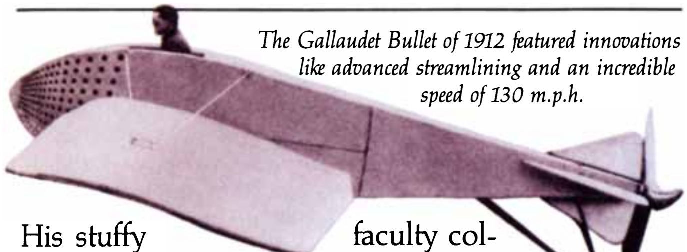
"A powerful blast of air racing six times faster than a tropical hurricane recently blew out an electric arc powerful enough to light for an instant all the lamps in Chicago. Only a slight puff of smoke trailed off as a reminder of the once powerful arc. The new circuit breaker quenches in a few inches of space an arc that theoretically would have to be pulled out some 40 feet to be extinguished in ordinary air."

"The newest of the synthetic rubbers, Chemigum, has just been announced by Goodyear as a result of several years' research. A new plant having an initial capacity of 10,000 pounds per day is being installed at Akron, Ohio. Chemigum is derived from petroleum through a cracking process, and tires made of it are said



A memorial for Christopher Columbus

HE WASN'T A WRIGHT BROTHER. BUT HE HAD THE RIGHT IDEA.



The Gallaudet Bullet of 1912 featured innovations like advanced streamlining and an incredible speed of 130 m.p.h.

His stuffy faculty colleagues were offended: his tinkering with "foolish flying gimcracks" was "undignified."

But young professor Edson Gallaudet was willing to give up his dignity, and his job, for a new idea about warping the wings of flying machines.

In 1898, he tested a scale model which proved his idea right. And five years later, Wilbur and Orville Wright flew using that same principle.

In 1908, Edson Gallaudet started what many credit as the first aircraft factory in America. Gallaudet Engineering Company became the earliest aircraft ancestor of General Dynamics.

Over the next 82 years, our history of building aircraft has included some of America's famous planes. And famous people.

Eddie Stinson, barnstormer and "birdman," who designed aircraft for many of the pioneer pilots, merged his company with ours. As did Jerry Vultee, whose planes set many distance and speed records, including Jimmy Doolittle's 12-hour cross-country flight.

The RB Racer, made by Dayton-Wright Airplane Company, was the first aircraft to have fully retractable landing gear. Dayton-Wright became an early part of General Dynamics.

F-16s of the USAF Thunderbirds.

During WWII, the Consolidated B-24 Liberator became the most-produced American bomber. After the war, our B-36 Peacemaker became the backbone of America's Strategic Air Command.

Our innovative, delta-wing design made the F-102 the world's first supersonic interceptor. And the B-58 Hustler the world's first supersonic bomber.

Today our F-16 Fighting Falcon is rated the finest fighter in the world. It well represents our long tradition of craftsmanship and creativity.

Once again, that tradition is about to be tested. In a technology competition against groups from West Germany and Japan, General Dynamics is teaming with four top American companies to develop the National Aerospace Plane.

To fly from runway to orbit, at speeds up to 17,000 m.p.h., we must invent new science. We must also invent new ways for American competitors to work together. But we are confident.

Since the days of Gallaudet, our company has been inventing not only better airplanes, but better ways to make them.

GENERAL DYNAMICS

A Strong Company For A Strong Country

to give superior performance to those made of German Buna. It also has possibilities for blending with natural rubber, so that it might help in an emergency to eke out slender supplies of the natural product."

SCIENTIFIC AMERICAN

OCTOBER, 1890: "Let us see what might be the consequences of a celestial meeting of the earth (traveling 18 miles per second) and a comet that had at least an equal velocity. If the comet had a consistent nucleus, the terrestrial crust would be staved in by the impact, and the torrents of lava that it conceals would produce a terrible commotion in contact with the waters of the ocean. In addition, the axis of the earth would be abruptly displaced. This is the sole plausible hypothesis to explain the inclinations of planets upon their orbit."

"The 40-inch telescope that is to be made by the Clarks, of Cambridgeport, for the University of Southern California, will be only one-ninth greater in diameter than the Lick glass telescope, but its light-grasping power will be about one-fourth greater. The existence of a large city on the moon would readily be detected by the telescope. But as to inhabitants of other planets, the 40-inch lens will leave us as much in the dark as we have ever been. For any such achievement as that we shall have to wait until a genius comes who can invent an instrument for seeing as much superior to the present telescope as an arc light is brighter than a tallow dip."

"Mr. M. Alberto de Palacio, of Bilboa, Spain, drew an original design for the 1892 Universal Exposition, of which we publish an engraving. It is an iron sphere nearly 1,000 feet in diameter which symbolizes the geographical completion of the earth which was realized by Christopher Columbus' discovery of the New World. The great pyramids of Egypt, the Sphinx, and the Colossus could lie in the hollow interior like jewels in their case. At night the sphere will be illuminated by the lines of light which will form the outlines of the continents and islands, thus casting over the city torrents of refulgent brilliancy. Under the majestic central rotunda will be placed a gigantic statue of the great discoverer surrounded by the navigators and missionaries who rendered his discovery fruitful."

The Problem With Buying Software From A Hardware Company

Once a company gets locked into hardware and software from one computer vendor, it operates at the mercy of that vendor.

Many companies buy and build software that runs on only one kind of computer. Locking themselves into one vendor's hardware.

These companies can then find themselves trapped into buying a ten million dollar computer when a one million dollar computer from another vendor would actually do a better job.

These expensive hardware traps can easily be avoided by using the right software.

Oracle software runs on virtually every computer – mainframes, minicomputers, workstations and PCs. From virtually every vendor.

Oracle software lets companies develop applications that run on all their computers. And share data among them as well.

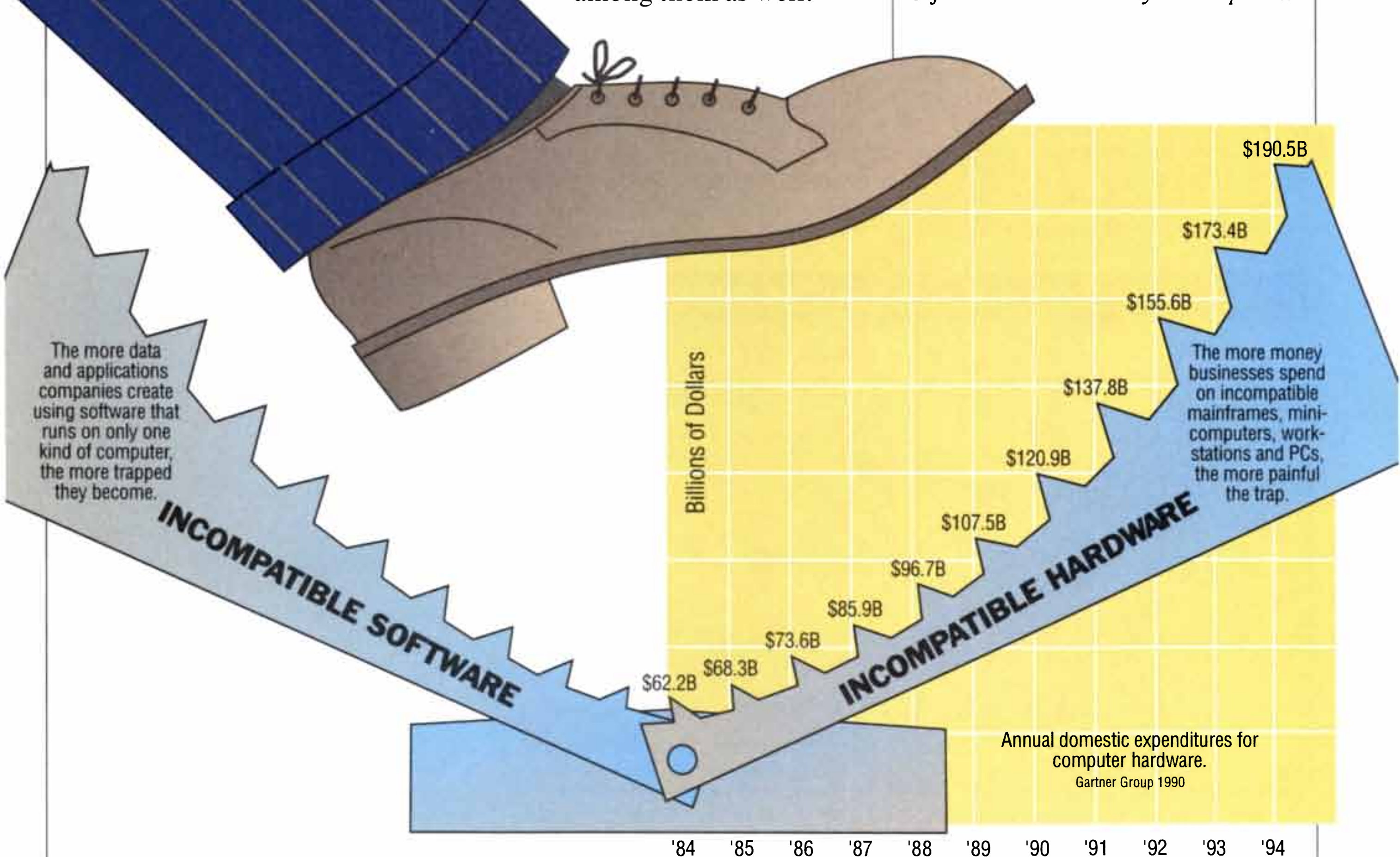
Most important of all, it lets companies exploit newer, faster and cheaper technologies as they emerge.

Call 1-800-633-0573 ext. 1812 for more information.

We'll tell you how you can move forward without worrying about what you're stepping into.

ORACLE®

Software that runs on all your computers.



SCIENCE AND THE CITIZEN

Straits of Magellan, transplanting retinas, marijuana receptor, sex and brains

New Order

Artificial evolution creates proteins nature missed

Identifying the sequence of building blocks that make up a particular protein or gene has become routine in molecular biology laboratories. But understanding what a given sequence means biologically is far more difficult. Consequently, the goal of constructing new drugs and biologically active substances from scratch has remained elusive. Molecular engineers have restricted their efforts to simple modifications of sequences that are products of natural evolution.

Not content with this timid approach, several research groups are pioneering an entirely different strategy. They avoid the difficulty of predicting how sequences will behave by creating a speeded-up microcosm of evolution by natural selection in the laboratory. Random sequences are "selected" to do a specific job. Directed evolution, as the technique has been called, can result in effective substances unlike anything known in nature—and many that nobody would have thought of making. The approach could enormously accelerate the production of new research tools and pharmaceuticals, many of which work by binding to particular targets in the body.

Molecular biologists have been talking about mimicking evolution to produce useful molecules for some years. But they have not been able to investigate the idea thoroughly until recently, because the experiments are technically difficult. Now, four reports have firmly established the technique.

The experiments vary in their particulars but are similar in principle. The process begins by creating a mixture of random sequences of amino acids or nucleic acid bases. That can be done easily with automatic nucleic acid synthesizers, which are becoming standard equipment in molecular biology laboratories.

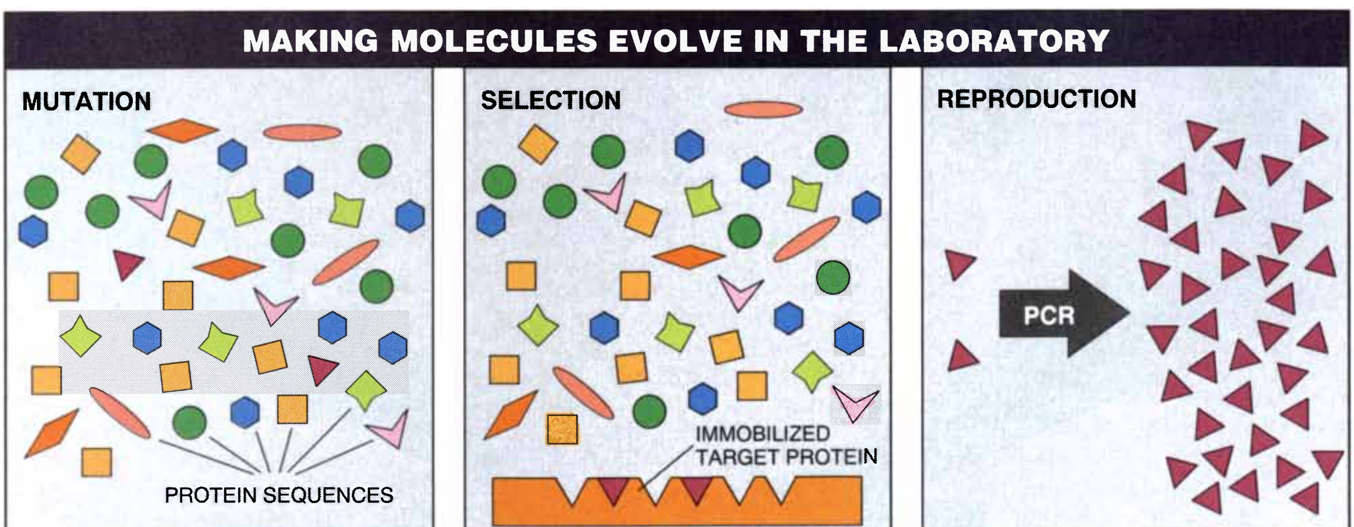
Normally such gene machines build specific sequences by attaching bases one by one to the end of a growing chain. But if the machines are fed a mixture of bases instead, they produce a random jumble of sequences. These nucleic acid sequences can be converted into protein sequences.

Production of the random sequences, whether of nucleic acid or protein, is analogous to mutation in natural evolution—albeit speeded-up mutation. The next step is selection. The random sequences are selected for their ability to bind to an immobilized molecule chosen to be the target. The target molecule can be a biological molecule such as an antibody or protein that has been attached to a substrate.

When the soup of random sequences is brought into contact with the immobilized target molecules, a few of the thousands or millions of sequences that happen to have the right shape will bind to the target. Those that do not are washed off. The ones that do bind strongly—and there may be just a few—are then forcibly removed by changing the chemical conditions.

The final step corresponds to reproduction. Millions of copies are made of the few molecules that bound strongly. The selection and reproduction steps are repeated several times: typically a few strongly binding sequences come to dominate the population after just a few cycles.

Writing in the journal *Science*, Craig Tuerk and Larry Gold of the University of Colorado at Boulder describe an experiment in which they selected among sequences of RNA for those that would bind strongly to an immobilized protein. To keep things to a manageable scale, they investigated only short sequences, eight bases long. Still, there were calculated to be more than 65,000 different sequences present. They used



Using nucleic acid synthesizers, researchers create thousands or even millions of random protein sequences. Many may not occur in nature.

The random sequences are exposed to an immobilized protein with a target receptor. A few of them bind strongly to the protein. The others are washed away.

The sequences are removed from the target protein. A process called the polymerase chain reaction (PCR) produces a large quantity of the sequence.

Like a silverwinged dragonfly the Cessna swoops down over a pan in Ngamiland. The four conservationists aboard hold their breath.

Below them are the ghosts of a bygone era, two black rhino in a strange and desolate world. Once this was called a Garden of Eden.

Once it was teeming with the black rhino of northern Botswana.

Perhaps, one day, we will regain our Paradise Lost.

The Trail Back to Paradise

Every time a new member of a vanishing species is sighted, there is another ray of hope.

At South African Airways, we understand it's of the utmost importance that some of the planet's former glory should be regained.

That's why whenever you fly to or from Southern Africa with us, South African Airways' African Wildlife Heritage Trust will benefit directly.

And when you fly with us you will discover a different kind of traveller's paradise.

That may be the reason why readers of the UK's prestigious Executive Traveller Magazine voted us the Best Airline to Africa for the second year running.

SAA
SOUTH AFRICAN AIRWAYS
We didn't invent flying, we just perfected it.

YOU COULD SPEND THOUSANDS OF DOLLARS TRYING TO TURN YOUR PCs INTO WORKSTATIONS.

a technique known as the polymerase chain reaction (PCR) to make millions of copies of the RNA sequences that bound strongly to the protein.

After four successive rounds of selection and "reproduction," Tuerk and Gold were left with just two RNA sequences that bound strongly. One was identical to a naturally occurring RNA known to bind to the target. The other sequence was unknown, and the researchers say it would never have been suspected of binding to the target. If such a sequence had been valuable as a pharmaceutical, the technique would have found it even though no chemist had thought of it.

The other three groups used a different technique to produce random sequences. They spliced random nucleic acids into the genes of bacteriophages—viruslike agents that infect bacteria. The genes give rise to proteins on the outside of the bacteriophage, so many millions of random protein fragments were expressed on the phage surfaces.

The workers then allowed the phages to come into contact with immobilized target molecules and washed away the phages that did not bind. The few phages that did were then recovered and allowed to reproduce in bacteria. All the groups found that after several rounds of selection and reproduction, only a few of the initial millions of protein sequences were present.

A group headed by Steven E. Cwirla and William J. Dower at Affymax Research Institute in Palo Alto, Calif., used sequences of six amino acids. Writing in the *Proceedings of the National Academy of Sciences*, the group reports that it quickly found several sequences that bound to the chosen target molecule, a monoclonal antibody. All the sequences bore a strong similarity to a portion of a known natural protein (beta endorphin) that is bound strongly by the antibody. (The exact natural sequence, though, was not found.)

Jamie K. Scott and George P. Smith of the University of Missouri at Columbia screened about 40 million random protein sequences six amino acids long for the ability to bind to two monoclonal antibodies. They also quickly "evolved" sequences similar to naturally occurring ones. In addition, they found an unknown sequence. A team headed by James J. Devlin at Cetus Corporation used sequences of 15 amino acids selected for the ability to bind to a substrate. Several binding sequences were quickly found, all containing a particular short stretch of amino acids that was presumably responsible for the binding ability.

Although it is too early to predict how valuable directed evolution will turn out to be, some researchers are already eyeing commercial applications. The Cetus group, which was racing Scott and Smith to publication (both reports were in the same issue of *Science*), has applied for a patent on its technique as a means of drug discovery. Gold hopes to adapt his technique, which he has named SELEX, to select nucleic acids directly for the type of proteins they encode. "We're looking at a quantum jump" in the ability to identify biologically active molecules, Dower says.

Aside from new drugs, the technique will teach investigators much about how proteins and nucleic acids behave. Directed-evolution experiments are also likely to fuel speculation about how life evolved, even though the experiments are carried out under unnatural conditions. The experiments, Gold says, provide hard data for "people who wonder what DNA and RNA are capable of." And that is just about all biologists. *—Tim Beardsley*

A Latin American caterpillar calls ants to a free meal

Ulterior motives often underlie friendships. That is certainly the case with ants and caterpillars. Caterpillars commonly feed ants a secretion rich in sugar and amino acids. In return, the ants refrain from devouring the caterpillars and even protect them from wasps and other predators.

Entomologists have observed ants and caterpillars throughout the world engaged in this odd alliance. Researchers have also suspected that many caterpillars somehow "call" the ants to them. But how, exactly? Philip J. DeVries, an entomologist at the University of Texas at Austin, has answered that question—at least in the case of a Latin American caterpillar known to biologists as *Thisbe irenea*.

The caterpillar, which metamorphoses into a small brown and white butterfly with iridescent blue trim on its lower wings, has a pair of ridged, rod-shaped organs located just behind its head. The caterpillar "plays" these tiny organs, which are called papillae, by scratching its bumpy head against them, according to DeVries.

The papillae generate a high-pitched, rhythmic chirping that is too subtle for humans to hear, but ants get the message in vibrations that travel through the ground or leaves. DeVries was able to record the sound by holding a microphone to the ground.

Devries says the papillae remind him of a Latin American percussion instrument called the *gueiro*, a gourd encircled with grooves that one plays by bumping a stick along its length. Next question: Can *Thisbe irenea* play "La Cucaracha"? *—Holger Wittekindt*

Venus Unveiled

Magellan's radar peeks under a cloudy coverlet

Unprecedentedly sharp pictures of Venus sent by the space probe *Magellan*—between perplexing breaks in radio contact—show a rugged terrain shaped by vulcanism. It is exactly what one would have expected of the Roman goddess of love, who was wife to Vulcan, god of fire and metalworking.

These first radar images, based on an orbit and a half of scanning, were taken mainly to test *Magellan*'s equipment. The picture on the left shows intersecting faults on a strip 20 kilometers wide and 80 kilometers long, which had appeared as a mere blur in earlier images made from the earth. A neighboring strip of comparable size, on the right, contains a particularly clear fracture line.

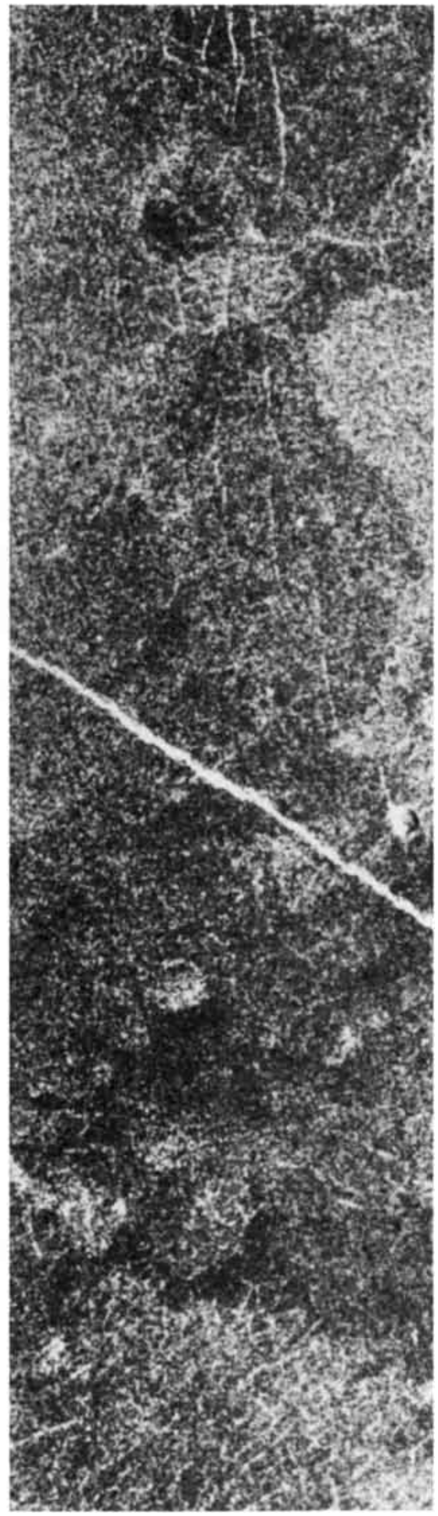
"We can see that it is an extensional scarp, probably due to the pressure of the uplift of the volcano and the region around it," says Ellen Stofan, a planetary geologist at the Jet Propulsion Laboratory in Pasadena, Calif. "What we didn't know before was the amount of fracturing in the lava plains around this fault, which shows that there's been a lot of deformation from tensional stress." Stofan says one of *Magellan*'s missions is to determine whether volcanoes still shape the Venusian crust or whether most interior heat is vented through the movement and subduction of rigid tectonic plates.

Full-scale mapping will begin in September, if communications problems can be solved, and will involve about 1,200 orbits over a period of 243 terrestrial days, one full rotation of Venus. *Magellan* uses a sideways-looking radar system to peer beneath the second planet's obscuring clouds. The radar can detect features measuring 120 meters on a side, a resolution about 10 times finer than that achieved by the Soviet Union's *Venera* probes, which used an earlier version of the technology seven years ago.

Such systems are often called synthetic-aperture radar because they exploit the spacecraft's motion to simulate the resolving power of a much larger antenna (or aperture, in the parlance of telescopy). *Magellan* shines a beam of microwaves downward and to the side, tracing a strip of land about 20 kilometers wide. Returning signals are recorded at various points in the orbital track, stored and then transmitted to the earth. There computers analyze echo delays to infer range and compare differences in wavelengths caused by the relative motion between the radar and its target. These calculations generate two-dimensional images.

Unlike such earth-orbiting observatories as the *Cosmic Background Explorer* and the *Hubble Space Telescope*, *Magellan* has only a single scientific instrument to accomplish three tasks. It images the terrain, measures the height of features and records thermal emissions from the planet's surface. *Magellan*'s main antenna is used both for mapping and to transmit data to handlers at J.P.L., a compromise in design that necessitates periodic interruptions in the mapping.

Take a good look at these pictures. The National Aeronautics and Space Administration meant them to be hors d'oeuvres, but they may turn out to be the whole meal if *Magellan*'s handlers lose track of their $750-million charge for good. —*Philip E. Ross*



CONTRABAND confiscated by the U.S. Fish and Wildlife Service and stored at its laboratory in Ashland, Ore., includes a number of different protected species.

Fowl Play

A U.S. agency uses science to fight wildlife bandits

The U.S. Fish and Wildlife Service has enlisted the help of advanced forensic techniques in its never-ending battle against poachers. The service has built a $4.5-million laboratory in Ashland, Ore., devoted to crimes with animal victims.

The sleuths have a formidable task. The world trade in illegal wildlife has reached an estimated $2 billion a year and involves thousands of species, many of them rare. Agents of the wildlife service must determine whether a handbag is made of alligator skin (which can be legally traded) or crocodile skin (which cannot) or whether the meat in a suspected poacher's freezer is a legally slain deer or a protected bighorn sheep.

Sometimes evidence is deliberately altered. Poachers have been known to disguise the sex of their kills and to stick arrows in bullet holes. (Only one sex of some species is protected, and other species can be hunted only with a bow and arrow.) Moreover, wildlife agents have no reliable way to detect such illegal exotica as bear bile, turtle oil and Oriental potions containing rhinoceros horn.

Kenneth W. Goddard, director of the laboratory, hopes to develop techniques that will solve such problems. (He says he is open to suggestions, particularly on chemical tests for turtle oil.) Goddard, who headed a police forensics unit before coming to the wildlife service laboratory three years ago, sees it as the future Scotland Yard of wildlife crime.

The young laboratory has already had some successes. Edgard O. Espinoza has found a way to distinguish elephant ivory (which is banned in international trade because of rampant elephant poaching) from mammoth and mastodon ivories (which are rare but can be traded) and from such pseudoivories as hippopotamus teeth. Elephant ivory, he found, displays tiny fissures called Schreger lines in a unique pattern, which can be detected with an optical microscope.

Another wildlife service scientist, Mary-Jacque Mann, is using a forensic scanning electron microscope to find residues left by gunsmoke on the hands and faces of poachers. Beth Ann Gilroy is using the same instrument to examine the downy breast feathers of birds: microscopic features on the feathers can reveal whether they are from a protected species.

Agents can identify the species of animal by-products by analyzing the proteins found in flesh or blood. Protein analysis, says Peter A. Dracht, who performs the technique at the laboratory, has helped to exonerate a hunter who was suspected of passing off bighorn meat as deer.

DNA fingerprinting—which exploits the fact that even members of the same species have different DNA—is also being explored. Steven R. Fain of the wildlife service says the technique could reveal that a bear's gallbladder (which cannot be sold legally) came from the same animal whose head is hanging in a suspect's living room.

DNA fingerprinting could also help trace the geographic origin of an animal and perhaps show whether it comes from a region where hunting is prohibited. Fain is building up a collection of tissue and blood samples from elk and black and grizzly bears from different regions. He can already distinguish California bears from Alaska ones. Recently he extracted DNA from tanned leather, a feat that could extend the range of the technique.

Not all the methods used by the wildlife laboratory are technologically sophisticated. The staff solved a case of wild elk rustling simply by putting food containing a fluorescent dye in the elks' habitat. The dye showed up days later in the urine of animals being passed off as captive-bred.

The wildlife service seems unlikely to run out of cases to investigate. A warehouse at the service's Ashland laboratory contains a waist-high pile of turtle-skin boots, snakeskin watchbands by the gross, tortoiseshell lampshades and racks of handbags fashioned from the skins of endangered reptiles and mammals. —*T.M.B.*

Seeing the Light

A glimmer of hope for retinal transplants

Investigators at Washington University in St. Louis report that an experimental retinal-transplant technique restores at least some communication between the eye and the brain in animals blinded by the loss of their rods and cones—the eye's light-detecting cells. The discovery, described at recent scientific meetings, fans the hope that the technique might eventually repair blindness in some of the millions of people with damaged photoreceptors. "If they're right, it's an extraordinary observation," says Joe G. Hollyfield of the Baylor College of Medicine, who studies degenerative retinal disorders. "I'm just not sure yet."

Surgeons are not sharpening their scalpels just yet, however. The data are still preliminary, and no one knows whether the transplants can actually restore the ability to see light or images.

The investigators, Martin S. Silverman and Stephen E. Hughes of Washington University's Central Institute for

the Deaf, replace missing photoreceptors with a layer of healthy ones extracted intact from a donor retina. The graft is obtained by affixing the retina to a block of gelatin and then shaving off thin slices until only the paper-thin photoreceptor layer remains. The gel dissolves after the transplant is inserted into its normal position in the eye.

The workers keep the original layer of cells intact because their normal orderly arrangement is thought to be as important to the perception of images as the presence of the photoreceptors themselves. These specialized neurons normally convert light into electric impulses that are transmitted to other neurons in the retina, which in turn relay the signals to the brain.

By the end of 1989 Silverman and Hughes had shown in rodents that the transplanted cells survive in the eye. They also demonstrated that the cells convey information about light to other parts of the retina.

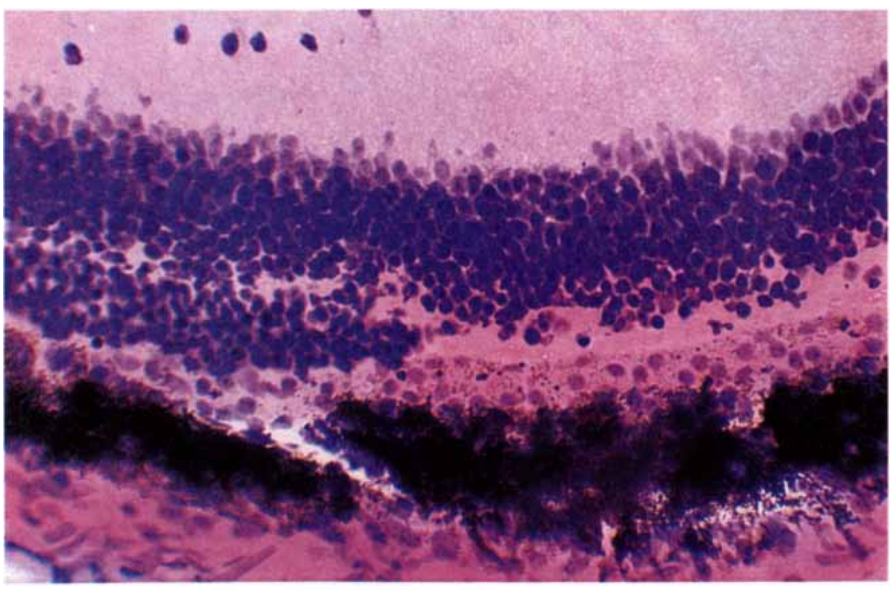
Although the workers have no information about pattern discrimination in the recipients, they report that a few important prerequisites for such discrimination are being met in animal experiments. For instance, they say, the grafted cells form synapses, or connections, with the appropriate neurons in the retina. In the absence of synapses, any signals emitted by the cells would be vague at best.

The researchers also report that after transplantation of the photoreceptors, the pupils of the eye constrict on exposure to light and dilate in the dark. That finding means, they say, that signals from the graft are reaching the brain. Of course, the pupillary response is reflexive and so does not indicate whether the transduced signals are getting to the cerebral cortex, where images are resolved. On the other hand, other experiments by the team reportedly show that neurons in the cortex fire in response to flashes of light.

To determine whether some amount of visual acuity can be restored, Silverman and Hughes plan to study pattern discrimination in primates. As a first step, they and Henry J. Kaplan, head of ophthalmology at Washington University, have adapted their technique to macaque monkeys, whose visual system resembles that of humans.

The workers naturally hope their technique will eventually restore some ability to discern form in blind individuals, but they realize that for many such people, simply being able to see light would be a major boon. "Given our results, we think at least light sensitivity might be possible," Silverman remarks.

If the technique does pan out, it could help people suffering from retinitis pigmentosa, an inherited disorder, and those who lose vision because of overexposure to light. In addition, as many as 10 percent of all adults over 60 become at least partially blind from macular degeneration, damage to the central part of the retina, where vision is most acute. In that condition, vision loss is caused to a great extent by the death of photoreceptors.—*Ricki Rusting*



***PHOTORECEPTORS** survived transplantation into the retina of a mouse whose own light-detecting cells had degenerated. The graft* **(half layer of purple dots)** *was put in the cell's normal position, between the rest of the retina* **(full purple layer)** *and the retinal pigmented epithelium* **(black)**. *Photo by Martin Silverman.*

Adjusted Agenda

AIDS activists accelerate alternative approaches

Mounting public pressure from AIDS advocacy groups is changing the direction of research into the fatal disease. Organizations such as the AIDS Action Council in Washington, D.C., as well as more vocal groups like New York's ACTUP (AIDS Coalition to Unleash Power), have pressed for more resources to be channeled toward fighting the opportunistic infections that usually end the lives of AIDS victims. With no vaccine or magic bullet in sight to stop the virus, they argue, efforts to combat the complications of the disease should assume greater prominence than they have so far. Now this argument seems to be winning widespread support.

In late August a strongly worded report from the National Commission on AIDS, an unofficial advisory group, charged that "a crisp, well-articulated clinical research strategy is simply not in evidence." The report echoed a call earlier in the month—made at hearings of the House subcommittee on human resources—for more research on opportunistic infections. "Rapid trials of a variety of antiparasitic drugs in vitro and in animal models would be simple and relatively inexpensive, and yet this has not been done," Renslow Sherer, director of the AIDS prevention service at Cook County Hospital in Chicago, said at the hearing.

In addition to the public pressure, some experts have come to believe that new treatment regimens will soon convert AIDS into what Anthony S. Fauci, director of the National Institute of Allergy and Infectious Diseases (NIAID), calls a "chronic manageable disease." As a result, the focus of funding is already shifting. Over the past year NIAID has begun to devote more effort to opportunistic infections, says John R. Killen, Jr., deputy director of the institute's AIDS program.

At the hearing, Fauci stated that the proportion of AIDS clinical-trial participants who are enrolled in studies on opportunistic infections doubled between July, 1989, and June, 1990, to 21 percent. He also explained that the reason NIAID's new drug-discovery program had until recently awarded no grants for AIDS-related opportunistic infections was that it had received no applications. To correct the balance, NIAID awarded $2.8 million to six research groups in August.

Officials say they plan ultimately to

devote 40 percent of their AIDS effort to opportunistic infections. Killen, however, cautions that awards will have to be driven by scientific opportunities. That is not enough for some AIDS activists, who see caveats as evidence of uninspired leadership. "Isn't it the job of NIH to make those 'opportunities'?" asks Jeffrey Levi of the AIDS Action Council. "They haven't said how we are going to target opportunistic infections and set up timetables and goals and measures of success."

Even so, NIAID is clearly making efforts to be responsive. A persistent criticism of the AIDS Clinical Trials Group, which administers most of the National Institutes of Health's AIDS tests, is that it is available only to patients who can attend one of the 46 participating academic research centers. So NIAID recently inaugurated a community-based clinical-trials research program, although its budget is only one tenth that of the AIDS Clinical Trials Group. The first study, which will cover therapies for the opportunistic infection toxoplasmosis, began this past summer.

Yet even some AIDS activists are concerned at the shift in emphasis away from combating the AIDS virus. "The AIDS community may overstate the significance of opportunistic infections," says Martin Delaney of Project Inform, a San Francisco group organizing community-based trials with the sanction of the Food and Drug Administration. "A person with no immune system is not going to be helped by any number of opportunistic infection treatments."

Delaney also says more community-based trials will be worthwhile only if they succeed in making experimental drugs available to a broader spectrum of patients and accelerate approval of new therapies. NIAID's Killen agrees there is a trade-off. AIDS patients do not want to abandon proved therapies like AZT in order to participate in trials of new drugs, yet conventional trials have strict exclusion criteria that would prevent participants from taking other drugs. While community trials may accommodate a broader range of patients, they compromise on the amount of data taken from each patient.

Whatever mix of strategies is adopted, experienced research managers are needed to continue the fight against AIDS. That may be one of NIAID's most intractable problems. The institute is struggling, with 26 out of about 126 positions in its AIDS division unfilled. The reason is obvious to all who work there. "Most people here could go into industry and get two, three or four times the salary," Killen says, "and a lot less grief." —*T.M.B.*

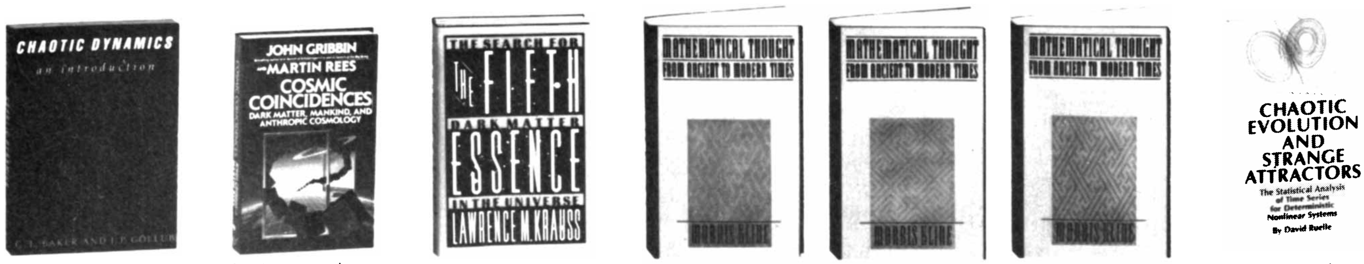
3 BOOKS FOR ONLY $1 EACH

as your introduction to the LIBRARY OF SCIENCE

You simply agree to buy 3 more books—at significant savings—within the next 12 months.

(Publishers' Prices shown.)

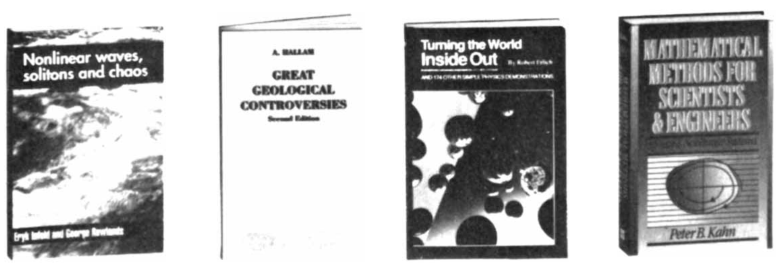
Values to $96.90



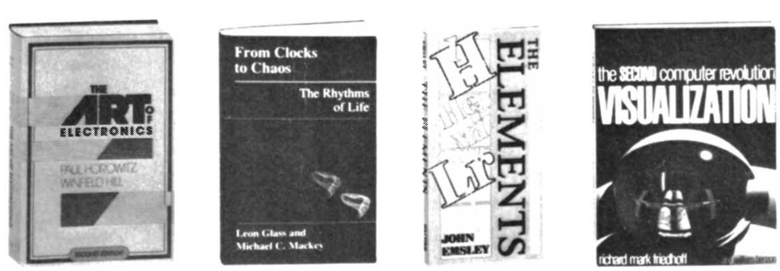
37515 $17.95 | 40647 $19.95 | 48390 $19.95 | 60765-2 $44.85 (3 books count as 2 choices) | 37508 $12.95

81830 $31.95 | 48849-2 $39.95 (Counts as 2 choices) | 79600 $24.95 | 36848 $18.95 | 64425-2 $49.95 (Counts as 2 choices) | 58103 $24.95 | 37923-3 $76.50 (Counts as 3 choices)

68753-2 $59.95 (Counts as 2 choices) | 55583 $35.00 | 52499 $29.95 | 81530 $21.95 | 67700 $25.00 | 87255 $19.95 | 83110 $24.95



64806 $29.95 | 50605 $24.95 | 84868 $14.95 | 60749-2 $54.95 (Counts as 2 choices)



34420-2 $49.95 (Counts as 2 choices) | 49104 $13.95 | 44176 $19.95 | 85886-2 $49.50 (Counts as 2 choices)

MEMBERSHIP BENEFITS• In addition to getting 3 books for only $1.00 each when you join, you'll also receive discounts on books you choose to purchase. Discounts generally range from 15% to 30% off the publishers' prices, and occasionally even more. • Also, you will immediately become eligible to participate in our Bonus Book Plan, with savings of 60% off the publishers' prices. • At 3-4 week intervals (15 times per year), you will receive theLibrary of Science News, describing the coming Main Selection and Alternate Selections, together with a dated reply card. • In addition, up to two times a year, you may receive offers of Special Selections which will be made available to a group of select members. • If you want the Main Selection, do nothing, and it will be sent to you automatically. • If you prefer another selection, or no book at all, simply indicate your choice on the card and return it by the date specified. • You will have at least 10 days to decide. If, because of late mail delivery of the News, you should receive a book you do not want, we guarantee return postage.

If reply card is missing, please write to Library of Science, Dept. 2-CT6, 3000 Cindel Drive, Delran, NJ 08075, for membership information and an application.

Scientific American 10/90

Teaching Teachers

A Nobel laureate's plan for improving science education

Everyone talks about the scientific illiteracy of American youth, but nobody does much about it. One exception is Leon Lederman, who has helped to create a program to improve the quality of science and mathematics instruction in the Chicago public schools. The program, called the Academy for Mathematics and Science Teachers, grew out of a phone call in the spring of 1989 from James D. Watkins, secretary of the Department of Energy, to Lederman, a Nobel laureate in physics and then director of the Enrico Fermi National Laboratory.

Watkins wanted to know what the Energy Department, which funds both Fermilab and Argonne National Laboratory (also in the Chicago area), could do for Chicago's public schools. Lederman, no stranger to innovative programs in education—in 1986 he helped found a high school for students gifted in the sciences—told Watkins he would give the matter serious thought.

By early this year Lederman, with the help of a committee, had drafted a formal proposal. The committee noted that the severe problems of the Chicago public schools, reflected by a 41 percent dropout rate, would require an unconventional approach. It advocated creating an academy for training Chicago's mathematics and science teachers, who serve approximately 400,000 students, in course content and teaching methodologies. "It is our belief," the committee wrote, "that mathematics and science education is valued because of urgent national needs and that inner city students, so far vastly underrepresented, are a major potential pool of science literates."

The current plan is for all 15,000 mathematics and science teachers in the city's 600 public schools to enroll in the academy over the next seven to eight years. At any given time, 500 science teachers will attend the academy, and 2,000 will graduate each year. They will be taught by a staff of 40. After a 10-week period of intensive training, the science teachers will return to their classrooms, where the academy will provide follow-up support.

The academy began operating this fall in an 80,000-square-foot building on the campus of the Illinois Institute of Technology. Funding from federal research agencies, including the Energy Department, will reach $25 million annually, starting with $5.7 million received this past spring. Contributions from Illinois (city, state and private sector) are expected to reach $10 million. Lederman and other organizers hope eventually to export the program to 25 or so major cities that share Chicago's problems. —*David Cooke*

LEON LEDERMAN helped to create a school for Chicago's science teachers.

Stardust Memories

Carbon dust grains reveal clues to stellar chemistry

Sneezing is inadvisable in Edward Anders's University of Chicago laboratory—at least when he's examining dust grains from beyond the solar system. In 1987 Anders and his co-workers found their first extrasolar grains—tiny diamonds embedded in primitive, carbonaceous meteorites. Later, with Ernst Zinner and his colleagues at Washington University in St. Louis, the researchers uncovered grains of silicon carbide. And in May the team identified grains of graphite.

The three types of carbon grains, thought to be a few million years older than the solar system, provide the only known compounds that retain a memory of its source—in this case, novas or carbon-rich, red giants. In addition, the carbon grains carry isotopes of noble gases created inside these stars. The grains and isotopes provide a "record of processes deep inside the stars" and regions near them, Anders says. In contrast, the matter in the solar system—melted, vaporized

SHARP PRESENTS THE CASE FOR A VGA 286 NOTEBOOK COMPUTER.

VGA DISPLAY.
The first in a notebook computer. It's bright, paper-white, with a crisp 640x480 resolution and 16 shades of gray.

TRUE NOTEBOOK SIZE.
8.5" x 11" x 1.4", 4.4 lbs.

12Mhz 286.
Runs the latest software with exhilarating efficiency. And it's fully expandable.

20MB HARD DISK.
With a speedy 23ms access time.

THE PC-6220.

About the only thing you *don't* get with the Sharp PC-6220 is another piece of luggage to carry. It's a mere 8½" x 11" and weighs only 4.4 lbs., yet has the kind of power you need. To run Windows.™ To do desktop publishing. To have instant access to your programs and files.

Find out how you can put a serious computer in your briefcase. Without taking everything out of it. Call 1-800-BE-SHARP.

SHARP®
FROM SHARP MINDS COME SHARP PRODUCTS™

and otherwise homogenized—contains no such record.

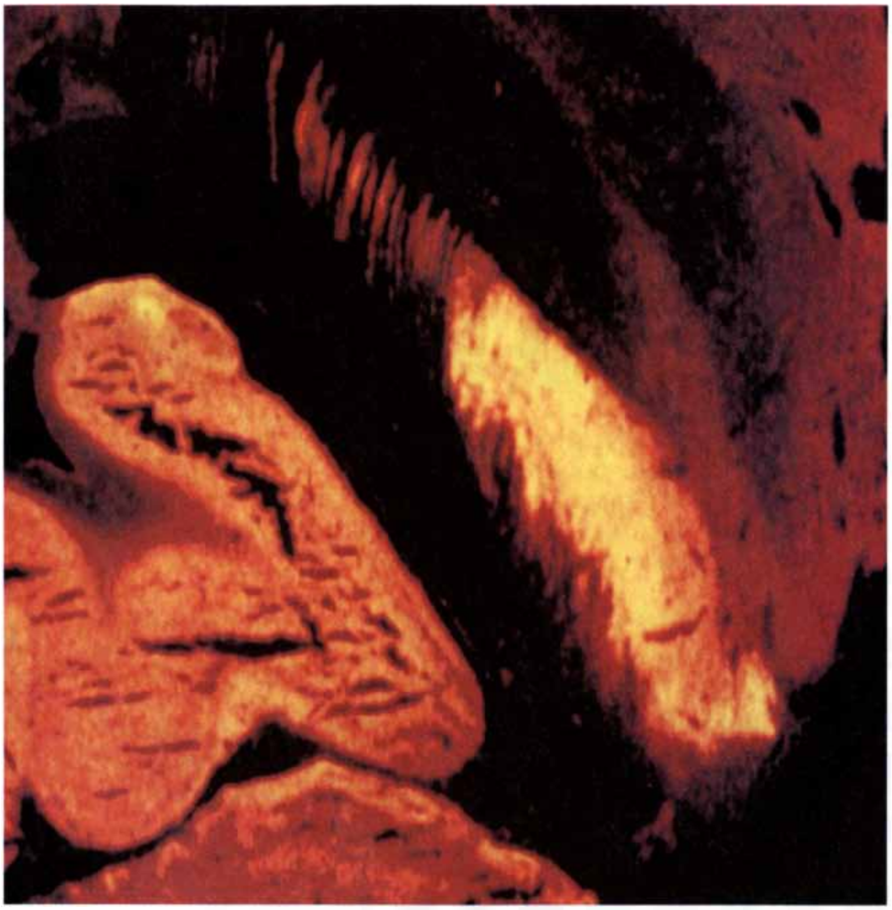
Knowing the relative proportions of diamond, silicon carbide and graphite grains in meteorites should help researchers better understand the chemical reactions in stellar atmospheres. For example, even though diamond is thermodynamically less stable than graphite, it is 100 to 200 times more abundant in the meteorites. The relative profusion of diamonds indicates that they were probably "made by the same chemical vapor deposition process" that creates diamond in laboratories, Anders says. Unfortunately, the diamonds cannot be studied individually; composed of 1,000 atoms or so, they are about .0015 micron across. Anders quips that the grains would make good engagement rings for bacteria.

The silicon carbide grains, however, are larger (reaching 12 microns) and can be analyzed to determine whether they came from different red giants or novas. So far Zinner and his colleagues have found that some of the silicon carbide grains came from at least four different red giants. Since the grains would have quickly decomposed in open space, they probably originated within a light-year of the solar system.

The discovery of the graphite grains (which are one to six microns in size) shows that graphite does indeed form when the matter ejected by red giants or novas condenses, in spite of calculations by other researchers indicating numerous obstacles to such formation. Still, the low amounts of graphite—in the sample, only one part in 10,000 is graphite—may indicate that most of the carbon in the universe is not incorporated as graphite, as many researchers believe.

As Anders sees it, red giants or novas ejected material that condensed into grains, which then sailed through interstellar space for several million years before reaching the coalescing solar system. Some were eventually incorporated into the parent bodies of meteorites that found their way to the earth and into laboratories. Robert Walker and his colleagues at Washington University are using in situ techniques to learn more about how the grains became embedded, as well as what other kinds of grains may exist.

Although many theories suggest how the solar system began coalescing, further study of the grains could provide evidence that the outflows of a nearby source, perhaps a red giant or a nova, triggered the formation of the solar system. Such questions, raised by the dust, will not be easily swept away. —*Philip Yam*



CANNABINOID RECEPTORS abound in a cross section of a human brain. Yellow regions have the most receptors; red regions have fewer. The brilliant yellow feature in this photograph by Miles Herkenham of the National Institute of Mental Health is the substantia nigra. The cerebral cortex is at the lower left.

Cannabis Comprehended

The "assassin of youth" points to a new pharmacology

Marijuana is celebrated—or castigated—for producing euphoria and altering perception. But the cannabis plant has for centuries been used for its medicinal properties as well as for intoxication. If it were possible to develop drugs that achieve marijuana's medicinal benefits but dispense with its mood-altering effects, they might make valuable additions to the physician's black bag.

Cannabinoids, the active chemicals in marijuana, have a range of physiological effects. They combat convulsions and nausea, dilate the air passages in the lungs and decrease pressure in the eyes. They can also kill pain and lower blood pressure. But because of their side effects, the Food and Drug Administration permits synthetic cannabinoids to be used only to suppress nausea caused by cancer chemotherapy and then only if other remedies fail.

Researchers at the National Institute of Mental Health in Bethesda, Md., have made a discovery that could lead to new pharmaceuticals based on marijuana. Lisa A. Matsuda, Tom I. Bonner and their colleagues have isolated and cloned a gene that gives rise to the receptor molecules that cannabinoids bind to in the brain. The workers recently described their findings in *Nature*.

Luck played a part in the discovery. The investigators had been searching for receptor genes using molecular screening techniques, but they were not looking specifically for a cannabinoid receptor. When they turned up what looked like a novel receptor gene in rat brain cells, they implanted it in cultured cells and then tested the cells' responsiveness to various substances. The last substances the group tested were cannabinoids.

The cannabinoid receptor is distinct from other types of drug receptors in the brain, such as those that bind to opiates. But like some other receptors, the cannabinoid receptor is bound up with proteins in the cell membrane called G proteins. These proteins relay chemical signals to the cell when their associated receptor binds to a target molecule.

Cannabinoid receptors presumably did not evolve for the pleasure of marijuana users. It seems more likely that the brain produces a natural cannabinoidlike substance. That compound has not yet been identified, but Allyn C. Howlett of St. Louis University, who first found evidence for cannabinoid receptors two years ago, says it is not one of the known neurotransmitters. Miles Herkenham of the National Institute for Mental Health, who has mapped cannabinoid receptors in the brain, finds they are particularly common in the hippocampus, a region involved in memory.

Now that Matsuda has cloned the gene, it will be much simpler to study and screen drugs that bind to the cannabinoid receptor. "Once you can clone the receptor, you can screen 1,000 compounds each day," observes Solomon H. Snyder of Johns Hopkins University, one of the investigators who established in the 1970s that the brain contains opiate receptors and produces natural opiates known as endorphins. Synthetic analgesics that arose from that discovery are now being tested in clinical trials.

Some drug companies studied synthetic cannabinoids a decade or more ago but gave up because of their side effects. Pfizer, for example, experimented with levonantradol, a compound that proved to be a potent antiemetic and analgesic in clinical trials. But the company abandoned commercial development of the drug because it sedated patients too much, according to Lawrence S. Melvin, Jr., a Pfizer pharmacologist.

Susan J. Ward, a pharmacologist at Sterling Drug, says her company has developed a new range of compounds, known as aminoalkylindoles, that bind strongly to cannabinoid receptors. Sterling has also developed compounds that prevent binding to the receptor. But it is Snyder's impression that "none of the drug candidates have been free of psychoactivity. They are good analgesics, but they make you high, so they won't be developed."

With the cloning of the cannabinoid receptor, that obstacle may vanish, Snyder says. As a molecular tool, the clone could lead the way to subtypes of the cannabinoid receptor. Drugs designed to affect receptor subtypes might well have advantageous effects. According to Snyder, Matsuda's data already provide a hint that a receptor subtype might exist. *—T. M. B.*

PROFILE: VIVE LA DIFFÉRENCE

Doreen Kimura plumbs male and female brains

Even as she looks for her car in an underground garage, Doreen Kimura is pondering the differences between female and male brains. Kimura, professor of psychology at the University of Western Ontario in London, Canada, believes women may rely more often on landmarks to orient themselves, whereas men may use spatial cues such as the distance of the car from the stairwell or the direction they turned on exiting. Ultimately Kimura finds her car using signs—stereotypically female, she says.

Kimura, a petite woman in her fifties whose stature belies her toughness, is perhaps the world's authority on neurological differences between the sexes. More than a decade ago she found that speech is processed in different regions in women's and men's brains and, more recently, that hormonal levels can affect people's performance on certain verbal and spatial tests. Her work has contributed to an understanding of the functions of the left and the right hemispheres of the brain and has offered clues to the evolution of language. That research has won her the respect of neurologists, psychologists and anthropologists alike—and disagreements with some linguists and feminists.

Driven by a desire to understand the neural mechanisms of human behavior, Kimura usually can be found in her office or laboratory, a cup of coffee within an arm's length, poring over data. Although she claims she does little other than work, her office gives her away: a photograph of her daughter adorns her desk, and on the side of her filing cabinet—hidden from direct view—are two pictures of Keith Richards, whom she saw last year on the Rolling Stones' revival tour.

Her interest in education came early. By the age of 17, Kimura was teaching at a rural Canadian school. Later she entered McGill University, the second person from her small town in Saskatchewan to go to college. There, in a course given by Donald O. Hebb, she found her lifelong interest in the brain. In 1949, a few years before Kimura came to McGill, Hebb's book *Organization of Behavior* was published, one of the first attempts to trace the neural basis for psychological processes.

Hebb's encouragement of independent research shaped Kimura's approach to science. She, too, demands autonomy from her students, making her a rigorous mentor—only five people have received Ph.D.s under her in the past 20-odd years. "You have to think you are right when everyone tells you you're wrong," Kimura explains. "You can't be aquiescent and do science."

Kimura learned to follow her instincts when she conducted experiments for her Ph.D. in the late 1950s at McGill. Using a dichotic listening test—in which the left ear and the right ear are simultaneously played a different series of numbers—she discovered that neurological patients reported that they heard better with the right ear.

DOREEN KIMURA likes raw data and rock and roll.

Because it had been established that the right ear communicates more directly with the left hemisphere of the brain, Kimura decided the left hemisphere was processing the incoming information more completely than the right. Her colleagues challenged her conclusion, calling it a fluke. But Kimura persisted. "There was no way I had done something wrong," she laughs. "So I just went ahead and tested quite a few normal people and found a very strong right ear effect." Then, using tapes of melodies, Kimura found that her subjects recognized tunes better with the left ear, which connects mainly with the right side of the brain.

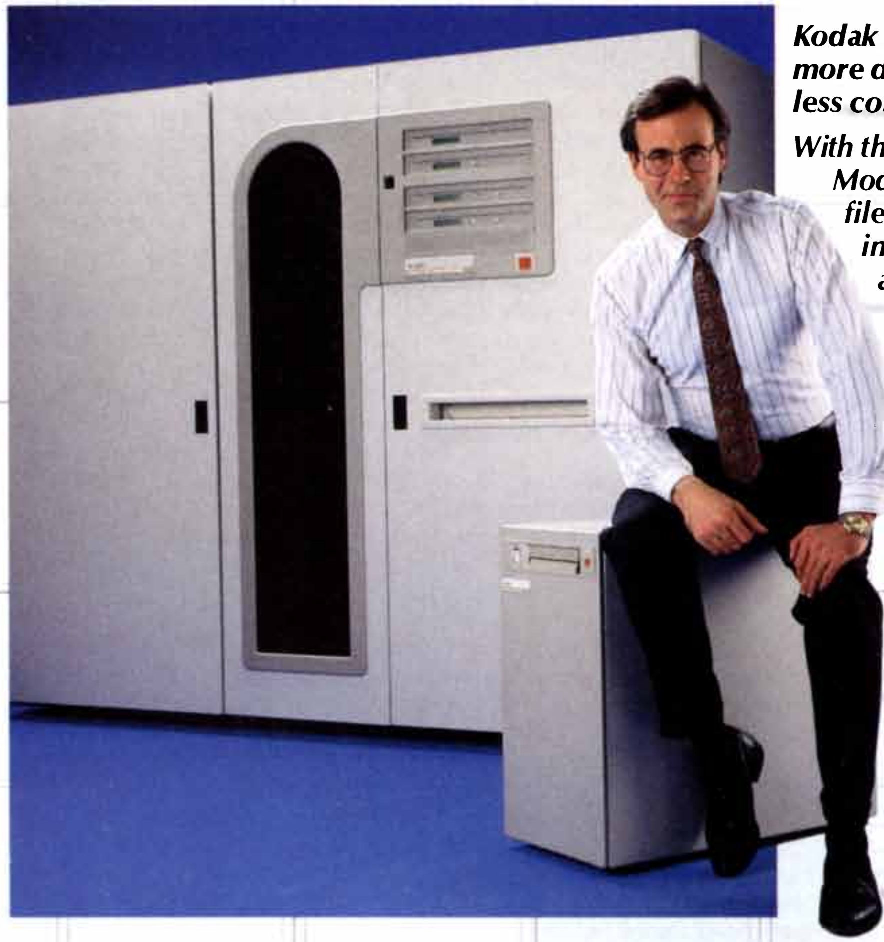
Those studies provided evidence that the two brain hemispheres process different kinds of aural information—supporting long-standing evidence that the left hemisphere processes language and the right hemisphere processes nonverbal information, including spatial cues. In addition, Kimura had shown that the hemispheres could be studied without surgery or neurological damage.

In 1964 she applied the same technique to studying vision, presenting various kinds of information to the right or left visual field and observing which hemisphere registered which information. She found that the right hemisphere was better at processing spatial material; the left recognized letters more readily.

Although she discovered ways to study normal subjects, Kimura gets behavioral clues from neurological patients, often those who have suffered lesions to different regions of the brain. In 1974 she set up a neuropsychology unit at the University Hospital in London. By correlating the site of the lesions with behavioral changes in patients, Kimura arrived at unique understandings about speech and language.

What initially struck her when she observed patients who had suffered lesions in the left hemisphere was not just that they had lost the ability to speak but that "these people were having great difficulty organizing their hands." Additionally, they could not make certain oral movements, Kimura notes, puffing up her cheeks and rolling her tongue to demonstrate. She concluded that instead of being language specific, the left hemisphere regulates certain aspects of motor control that are also important for speech.

By emphasizing the relationship between motor control and speech, Kimura argues that language is not uniquely human. The left hemisphere most probably controlled complex motor skills, including tool making, before speech.



Evidence of primates' ability to learn gestural languages supports this idea, Kimura says.

Some linguists, who believe that language is fundamentally human, were quick to challenge her hypothesis. "They already know how the brain works," Kimura deadpans and then laughs. "The fact that damages in the left hemisphere produce difficulties in making oral movements is something they don't talk about." But she also sees such intellectual resistance as problematic for science at large. "If you come with very strong preconceptions about how things are organized, you don't find anything out," she insists.

Kimura herself approaches data with what she hopes is an open mind. In her laboratory, Kimura can barely tear herself away while her teaching assistant prints out the results of a new experiment. Soon after, her desk is covered with raw data. "I let it tell me things. I really like looking at data—it is one of my favorite activities," she says.

Kimura's knack for ferreting out findings led her to an important discovery about sex differences in the human brain. Studies of stroke patients done in the 1970s had suggested that women's brains were more symmetric than men's—that women's verbal abilities were more equally distributed between the two hemispheres. Kimura, however, found the sex differences lay not between the hemispheres but within them.

After compiling 10 years of data from the university's neurological clinic, Kimura found that women suffered speech problems if they had damage in the front part of the left hemisphere. Men tended to have similar impairments if the damage was in the back part of that hemisphere. This intrahemispheric variation appears to be "the critical difference between the sexes."

Wondering how these brain differences arise, Kimura turned to the study of hormones. Since hormones play a part in organizing male and female characteristics in the brain, Kimura thought hormonal fluctuations would affect function. Sure enough, Kimura and a student, Elizabeth Hampson, found remarkable variations—earning Kimura some disfavor from feminists.

Kimura and Hampson discovered that women performed certain verbal and motor tasks better (women usually excel in these areas) when their estrogen levels were high. Their performance also showed a reciprocal correlation to spatial tests (specifically those that men usually perform well). When estrogen levels were low, women did better on those spatial tests. "We are interested in sex differences as one way of studying how individuals differ from one another," says Kimura, who is already examining how men's testosterone levels affect their spatial ability.

Research on sexual differences in brain organization and performance is politically charged—even when those differences are slight. Kimura attributes some of the resistance to her findings to the way in which the estrogen study was reported: one Ontario headline urged that men prevent women from parking the car at certain times of the month. But also "the idea that there may be some very early determinants of brain organization, and consequently function, is just unappealing to [some people's] view of life," Kimura states. "Obviously, I don't share that. My interest is in finding out how the brain really works. I don't have any ax to grind, political or otherwise."

The hormonal findings have recently overlapped with studies on body asymmetry, which Kimura describes as the "weirdest stuff we are doing." Other research has shown that bodies are generally asymmetric. Kimura is finding in preliminary studies that men tend to be larger on the right side, women larger on the left. These somatic sizes, which may be reflected in differences in the brain hemispheres, correspond to performance on certain tests. "Things like math are better in right larger individuals, and the trends are the same in males and females," Kimura says. To her "amazement," she says, physical asymmetry seems to predict some degree of cognitive function.

Kimura believes sexual dimorphism evolved as a result of task specialization in hunter-gatherer societies. Women stayed closer to camp and to their children, focusing on immediate surroundings and developing fine-motor skills—the precursors of language. Consequently, perhaps, women remain better at certain fine-motor skills. Meanwhile men wandered afield, oriented more to external space, giving them an edge in specific spatial tests.

"The fact that hormonal systems are still so dynamic is puzzling," Kimura says. She suggests the neural mechanism is flexible enough to allow the degree of sexual dimorphism to change. In other words, behavior and capability are not totally fixed in the genetic code.

The flexibility that Kimura perceives in human behavior is reflected in her creative approach to science. She tries to remain open to new, sometimes "crude," ideas. "That's what you collect data for. You don't collect it to confirm your hypothesis," Kimura asserts. "If you do, you are doing boring research." —*Marguerite Holloway*

Outdoor Performance. Engineered by Rockport.™
Each of these handsome new Rockport® shoes are engineered from the bottom up. Rockport's unique Rocker Profile helps you walk easier, more naturally, more efficiently. Our unique blend of Vibram® in the midsole provides shock absorption and resilience, while the outsole provides incredible traction. Inside, the Rockport® Walk Support System® takes over, offering a dozen specific technologies to support your foot, cushion it with extraordinary comfort, and propel you forward naturally.
The biggest surprise is when you hold one in your hands. Rockports actually weigh less than most athletic shoes.
Rockports are men's and women's dress, casual, and fitness walking shoes.
For the store nearest you, call 1-800-343-WALK.
Rockport®
Innovations in Comfort.
ROCKPORT and WALK SUPPORT SYSTEM ARE REGISTERED TRADEMARKS and ROCKPORTS MAKE YOU FEEL LIKE WALKING and ENGINEERED BY ROCKPORT are trademarks of THE ROCKPORT COMPANY, INC. VIBRAM is a registered trademark of VIBRAM S.P.A. © 1990 THE ROCKPORT COMPANY, INC. ALL RIGHTS RESERVED.
Hanover • REI
and other fine stores
Rockports make you feel like walking™

SCIENTIFIC AMERICAN October 1990 Volume 263 Number 4

A Grandmaster Chess Machine

In the 40 years since this magazine published the original prospectus for a chess computer, machines have vanquished first novices, then masters and now grandmasters. Will Gary Kasparov be next?

by Feng-hsiung Hsu, Thomas Anantharaman, Murray Campbell and Andreas Nowatzyk

In January of 1988, at a press conference in Paris, world chess champion Gary K. Kasparov was asked whether a computer would be able to defeat a grandmaster before the year 2000. "No way," he replied, "and if any grandmaster has difficulties playing computers, I would be happy to provide my advice."

Ten months after Kasparov's statement, in a major tournament held in Long Beach, Calif., Grandmaster Bent Larsen, a former contender for the world title, was defeated by a chess-playing machine we had designed in a graduate project at Carnegie-Mellon University. The machine, a combination of software and customized hardware called Deep Thought, won five other games, drew one and lost one, tying Grandmaster Anthony Miles for first place. Because machines are disqualified from winning money in tournaments, Miles pocketed the first prize of $10,000. (Deep Thought nonetheless defeated Miles a year later in an exhibition play-off match.)

By the summer of 1990—by which time three of the original Deep Thought team had joined IBM—Deep Thought had achieved a 50 percent score in 10 games played under tournament conditions against grandmasters and an 86 percent score in 14 games against international masters. Some of these games and dozens of others against less distinguished opponents had been played under the auspices of the U.S. Chess Federation, which used the results to derive a chess rating of 2552. That rating indicates a playing strength in the bottom half of the grandmaster range. An average tournament player, by contrast, is rated around 1500 [*see illustration on page 47*]. In the games played after August of 1988, when the computer reached its current analytical speed of 750,000 positions per second, its performance rating exceeded 2600.

The next generation of the machine, expected to play its first game some time in 1992, will run on far more powerful hardware. The equipment will increase the speed of analysis by more than 1,000-fold, to about a billion positions per second. This change alone might well make Deep Thought's descendant a stronger chess player than Kasparov—or any other human being in history.

Why would anyone want to teach a machine how to corner a wooden king on a checkered board? First, chess has long been regarded in the West as the preeminent game of wits and therefore—in Goethe's words—"the touchstone of the intellect." Many people argue that a successful chess machine would prove that thinking can be modeled or, conversely, that chess does not involve thinking. Either conclusion would surely change the conception of what is commonly called intelligence.

Furthermore, computer chess presents an appealing engineering problem. The case was stated in the pages of this magazine 40 years ago by Claude E. Shannon, the founder of information theory [see "A Chess-Playing Machine," by Claude E. Shannon; SCIENTIFIC AMERICAN, February, 1950]:

> The investigation of the chess-playing problem is intended to develop techniques that can be used for more practical applications. The chess machine is an ideal one to start with for several reasons. The problem is sharply defined, both in the allowed operations (the moves of chess) and in the ultimate goal (checkmate). It is neither so simple as to be trivial nor too difficult for satisfactory solu-

FENG-HSIUNG HSU, THOMAS ANANTHARAMAN, MURRAY CAMPBELL and ANDREAS NOWATZYK constructed Deep Thought, the world's leading chess machine, while completing doctorates in various fields of computer science at Carnegie-Mellon University. Hsu, Anantharaman and Campbell have since joined the IBM Thomas J. Watson Research Center; Nowatzyk is with Sun Microsystems. Hsu began the project, worked on it full-time and acted as the system architect. He received a B.S. in electrical engineering from National Taiwan University. Anantharaman wrote most of the host software and implemented various algorithms. He received a B.S. in electrical engineering from Banaras Hindu University in India. Campbell maintained the chess opening book and wrote the precomputation software for the evaluation function. He received B.S. and M.S. degrees in computer science from the University of Alberta. Nowatzyk designed and implemented the automatic tuning of the evaluation function. He received diplomas in physics and computer science from the University of Hamburg in West Germany.

WORLD CHAMPION Gary Kasparov poses with an IBM PS/2, used to communicate with Deep Thought, before beginning a match against the machine late in 1989. Kasparov won despite Deep Thought's grandmaster-strength rating.

tion. And such a machine could be pitted against a human opponent, giving a clear measure of the machine's ability in this type of reasoning.

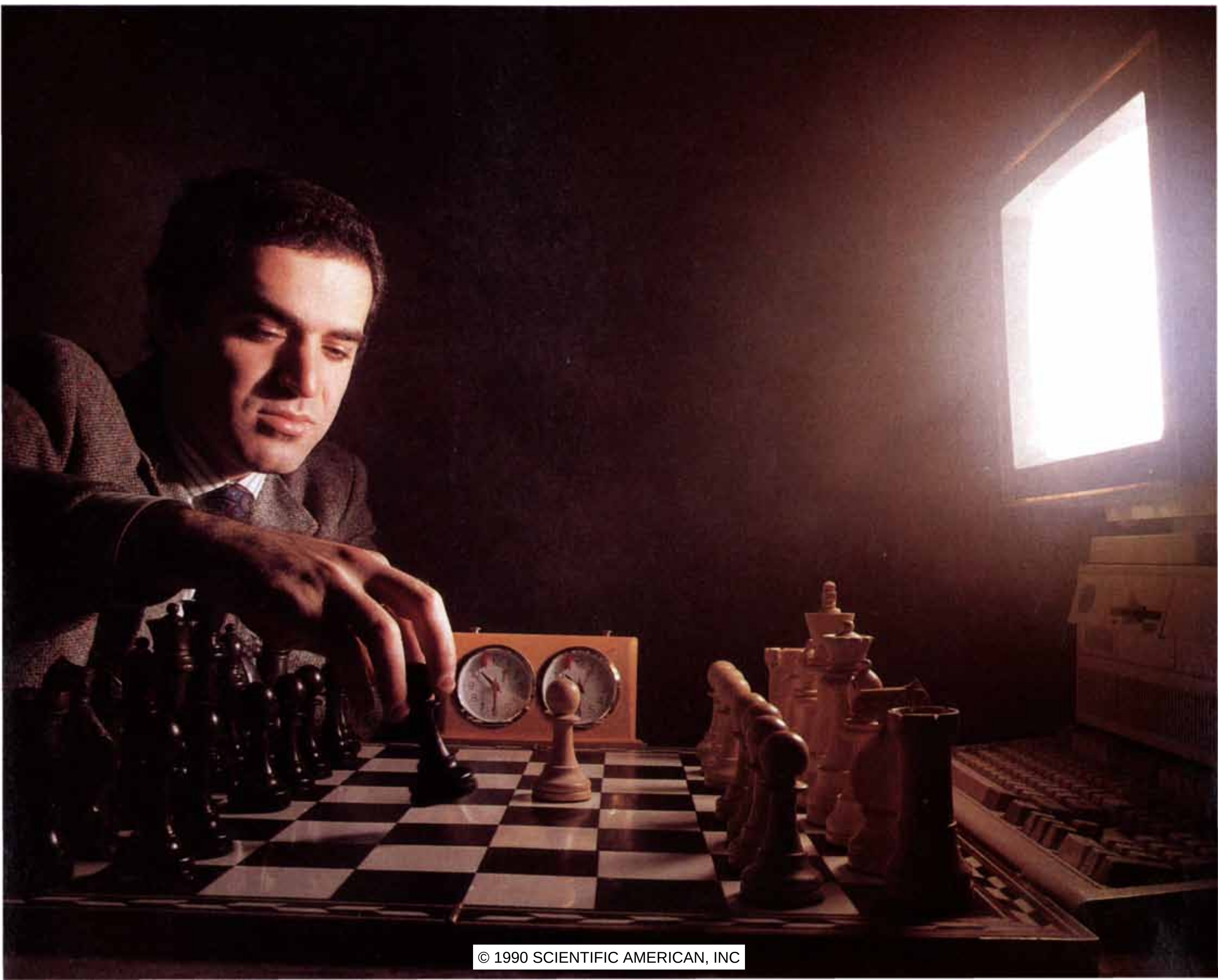
Perhaps the greatest practical consequence of chess programming comes from its demonstration of the efficacy of computer analysis. The perfection of related techniques promises to advance network design, chemical modeling and even linguistic analysis.

The idea of a chess-playing machine dates back to the 1760s, when Baron Wolfgang von Kempelen exhibited the Maezal Chess Automaton in Europe. The machine, nicknamed the Turk because it played its moves by means of a turbaned and mustachioed marionette, was apparently actuated by a complicated mechanism in a cabinet underneath. It generally played well and once sent Napoleon Bonaparte into a fury by beating him in 19 moves. Edgar Allan Poe, among others, later guessed the automaton's secret—a diminuitive chess master made its moves from a secret compartment—but Poe cited the wrong reason: he argued that the Turk's occasional losses were inconsistent with the perfection of a true machine.

Alan M. Turing, the British mathematician, computer scientist and cryptographer, was among the first to consider the problem of a chess-playing computer. He found it easier, however, to work his simple move-generating and position-evaluating program by hand than by machine. Konrad Zuse in Germany and other workers made similar efforts, but the seminal work was done by Shannon. He built on the discoveries of John von Neumann and Oskar Morgenstern, who, in their general theory of games, had devised a so-called minimax algorithm by which the best move can be calculated.

The process basically represents an arbitrarily large number of positions that might result from every possible series of moves, assigns them a numerical score and works backward from this information to derive the best first move. It begins when a move generator calculates all the moves the computer might play from the position at hand, then all the opponent's possible replies and so forth. Each step along the chain of events is called a half move in chess parlance or a ply in the terminology of computer science.

Each new ply in the branching tree of analysis encompasses roughly 38 times as many positions (the number of moves in a typical chess position) as did the previous one, or six times as many positions if "alpha-beta pruning" is used [*see box on page 48*]. Most positions therefore reside in the tree's outermost buds, from which the tree grows until either the game or the computer's allotted time is exhausted. An evaluation function then scores each end position, assigning perhaps "1" to a checkmate of the opponent, "-1" to a checkmate by the opponent and "0" to



a draw. More nuanced advantages can also be registered and balanced against one another. The computer can, for example, count material values—those of pieces and pawns—and calculate positional values according to parameters that represent piece placement, pawn structure, occupation of an unobstructed vertical line of squares (called a file), control of the center and so on.

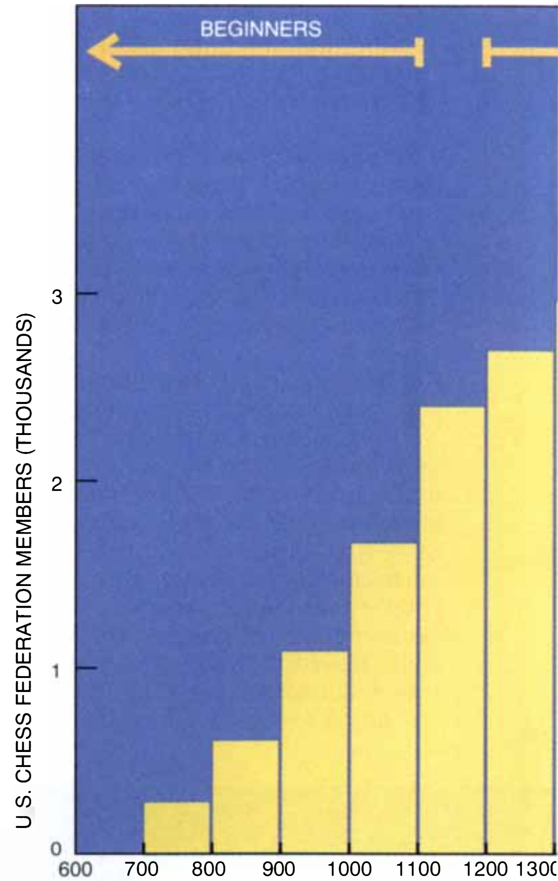
One can strengthen a computer's play by improving its search ability or refining its positional judgment. Flawless play would result if a computer could generate all possible games and classify the end positions as mating or being mated, or neither. Such a computer might surprise its opponent on the first move by announcing, "White to play and mate in 137 moves," or, alternatively, by resigning the position as hopeless. Such exhaustive analysis is easy in games as simple as tic-tac-toe but impractical in chess, in which 10^{120} discrete games are possible. Equally perfect play could be obtained from an examination of only one ply, provided the positional evaluation were as good as that jocularly claimed by Richard Réti, a master who flourished in the 1920s, when he said that he saw only one move ahead—the best.

Such pretensions were far from the minds of the earliest chess programmers, who were not even able to program machines to observe the rules of chess until 1958. Another eight years passed before MacHack-6, a program written by Richard D. Greenblatt of the Massachusetts Institute of Technology, became the first computer to reach the standard of average tournament players.

As the number of people developing chess programs increased, they divided into two philosophical camps. Let us call them the emulation camp and the engineering camp. The former asserted that chess computers should play as humans do, perhaps by means of explicit reasoning about move decisions. The latter took a less restrictive view, arguing that what works well for humans may not be applicable to computers. The emulation camp had most of the say in the early days, when computer chess was more a matter of theory than practice.

In the 1970s the engineering camp moved to center stage when the depth of search was found to correlate almost linearly with the program rating. Each additional ply added about 200 rating points to the computer's playing strength [*see illustration on opposite page*]. Programmers therefore scrambled for access to ever faster computers and found engineering tricks by which to squeeze deeper searches from the available processing power.

MOUNTING CHALLENGE of computers is seen in this superposition of the U.S. Chess Federation's 35,000 rated members and the ratings that are achieved by machines at varying depths of analysis (*green*). Computers gain about 200 points for each additional half move, or ply, they examine. Deep Thought now searches 10 plies for a strength near 2600. Its successor will analyze 14 or 15 plies for a vastly higher rating.



Managing the search is half the battle. At the very beginning of chess programming, search routines generated positions with practically no discernment. They treated transposing variations—which lead to the same position—as if they were distinct. Such needless double counting is now avoided by keeping track of positions in memory arrays known as hash tables. Hash tables provided an even greater benefit by helping the alpha-beta algorithm to weed out many irrelevant lines of play.

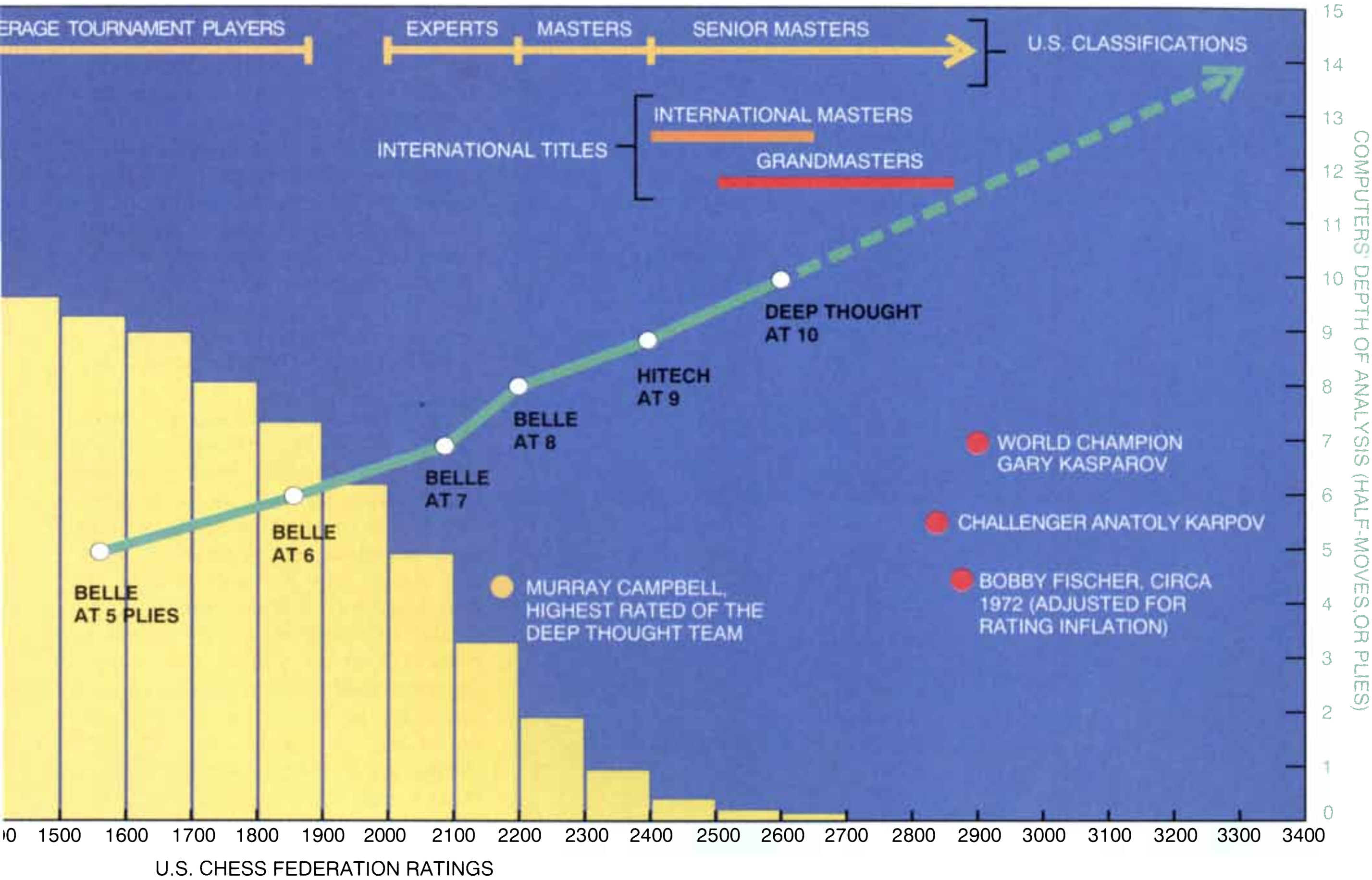
The greatest problem in a search is knowing where to end its multitudinous branches. One cannot examine all lines indefinitely, but one would like at least to avoid cutting off analysis in unstable positions. Such positions result when the analysis stops in the middle of an exchange of pieces. Suppose, for example, the computer searches exactly eight plies ahead in all lines and discovers an eighth-ply position in which it seems to have won a knight in exchange for a pawn. Even if the very next move will enable the opponent to recover the knight and remain a pawn to the good, the computer will steer tenaciously toward the illusory material advantage.

This so-called horizon effect can cause computers to commit a form of chess suicide not seen in the games of even the weakest human players. Out of the blue and for no reason apparent to the naive observer, the machine will begin to throw away its pawns and pieces, leaving its position in tatters. To reduce the chance of such errors, virtually all programs now add a stage of quiescence search to the basic search. Such searches typically examine only sequences of captures of pawns or pieces until reaching a stable, or quiescent, position that is suitable to static evaluation.

The 1970s and early 1980s saw the prevalence of what were commonly called brute-force machines, because they owed their strength to sophisticated implementation of the basic and quiescence search strategies. The period was dominated almost continuously by the Northwestern University program Chess 4.0 and its 4.X descendants. The Northwestern program hopped from one generation of computing hardware to another, steadily increasing its rating until it surpassed expert level (2000) in 1979.

Several attempts to build special-purpose chess machines also began in the 1970s. The most famous one, AT&T Bell Laboratories' Belle, crossed the national master barrier of 2200 in 1983. The era of pure brute-force machines reached its peak around 1986 with the ascendancy of Cray Blitz, which ran on the Cray supercomputers, and Hitech, a special-purpose machine that generated moves on 64 chips—one for each square. Hitech won the 1985 North American Computer Chess Championship, and Cray Blitz won the 1986 World Computer Chess Championship on a tiebreaker by beating Hitech in the last round. Cray Blitz and Hitech searched 100,000 and 120,000 positions per second, respectively.

Deep Thought had a rather unusual history. First, it was developed by a team of graduate students who had no official sponsorship or direct faculty supervision. (Faculty members



conducting work on computer chess at Carnegie-Mellon had no connection with the Deep Thought team.) Second, the team's members had diverse backgrounds that led them to adopt unorthodox approaches.

In June of 1985 one of us (Hsu) concluded that a single-chip move generator could be built with the Very Large Scale Integration (VLSI) technology that was provided to the academic community by the Defense Advanced Research Programs Agency (DARPA). Hsu based his chip on Belle's move generator but found several refinements that made the design amenable to VLSI implementation. He also designed the chip so that its electronic features (including 35,925 transistors) could be packed efficiently, despite the rather coarse (three-micron) minimum-feature size offered by MOSIS, the silicon broker that provided DARPA-funded fabrication services. Hsu spent six months working on design, simulation and layout and then waited four months to receive the first working copies. He tested the chip by linking it to a scientific workstation, and he found that the chip could process up to two million moves per second, 10 times faster than Hitech's 64-chip array.

At this point, Hsu joined forces with Thomas Anantharaman, then a computer science graduate student in the university's speech-recognition group. Anantharaman had written a toy chess program that generated its moves via a software package. By substituting Hsu's chip tester for the package, Anantharaman speeded the program's analysis by 500 percent, to a total of 50,000 positions per second.

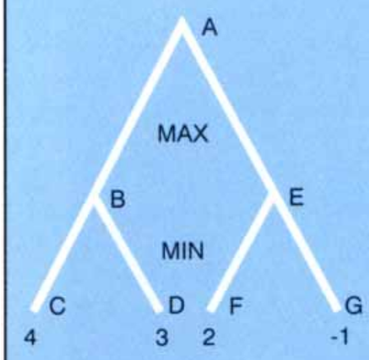
Hsu and Anantharaman became ambitious and, as a lark, decided to prepare their machine for the 1986 North American Computer Chess Championship, then only seven weeks away. Murray Campbell and Andreas Nowatzyk, two graduates in computer science, were then recruited to the project. A more sophisticated evaluation function was desirable, and Campbell, who had once played competitive chess, agreed to work on it. The second and even harder task, given the time constraint, was to augment the chip tester so that it could act as a simple searching engine. Such an engine would exploit the potential speed of the move-generator chip more fully than the workstation linkup had.

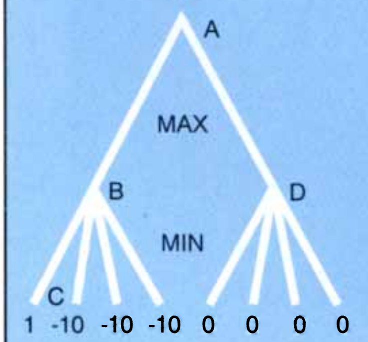
Hsu took drastic action to get this component ready in time: he decided the engine would ignore two basic aspects of chess: castling of the king and rook and the repetition of positions. (Any player can claim a draw by demonstrating that a position has been repeated three times, with the same player to move in each case.) To compensate for the omissions, a hybrid search strategy was adopted wherein the early plies were searched on a host computer that considered castling and repetition of positions. The later plies (which of course accounted for most of the positions) were analyzed on the engine.

We had no budget and so created our first machine, ChipTest, by scrounging parts from other projects. The value of all the components did not exceed $500 or $1,000—if the estimated unit cost of the DARPA-funded chips is taken into account. But neither the engine nor the host software was fully debugged in time for the championship, and the defective machine merely managed to obtain an even score. Still, that was not a bad result for seven weeks' work.

We learned much from this debut. Hsu observed, for example, that two other programs had played into lines in which their every move could be forced and where neither side knew the outcome in advance. In other words, the program that emerged with the better position owed its advantage to blind luck. Hsu proposed to remedy the defect with what he called the sin-

TREE-SEARCHING TECHNIQUES





To derive the best first move, a chess program generates a tree of variations, evaluates the final positions and works backward from them. Positional scores are made from the program's point of view; a high score, therefore, is advantageous.

A simple technique, called the minimax algorithm, winnows out the best variations by seeking maximum scores for the computer's moves and minimum scores for the opponent's replies. In the tree on the top, nodes B and E are minimized to 3 and -1, respectively, and node A is maximized to 3.

This simple approach can search an additional ply only by increasing processing power by a factor of 38 (the number of moves in a typical chess position). Alpha-beta pruning refines the algorithm by allowing the program to ignore irrelevant lines of play, so that a sixfold speedup suffices to analyze an extra ply. If, for example, the scoring begins at node C and proceeds to the right, the computer will assign B a value of 3 and then notice that E is less than or equal to 2. It therefore need not examine F.

Another method, called singular extension, gets more out of a computer's search by focusing on critical positions. In the tree on the bottom, the value of B depends heavily on that of C, whereas the value of D does not depend on any one of its successors. To increase the reliability of A's score, the singular extension algorithm would therefore search C one ply deeper than normal. This method allows Deep Thought to see extraordinarily far ahead in many tactically complicated positions.

gular extension algorithm. The algorithm looks deeper, that is, it extends the search to greater depths, in lines where the computer sees only a single good reply. The goal is to ensure that critical positions are given particularly careful attention [*see illustration above*].

When one side is about to win, say, a trapped bishop, the defender usually has fewer and fewer good replies the further the search depth proceeds. Toward the end there is only one good reply, after which the bishop is finally lost. Singular extension homes in on such cases. In one game, it enabled the computer to shock a master by announcing mate within 19 moves.

Anantharaman, the only person who understood the code he had written for the ChipTest host computer, programmed the singular extension algorithm on his own. Meanwhile Hsu completed the microcode, the instructions that control the hardware on its most elementary level. Searching at between 400,000 and 500,000 positions per second, ChipTest won the 1987 North American Computer Chess Championship in a clean sweep, defeating, among others, the world champion machine, Cray Blitz. Thus ended the reign of pure brute force. These days, almost all top programs incorporate at least some elements of selective search.

Our work had made it clear that ChipTest's hardware could be speeded up and that the search could be managed more intelligently. The roughly $5,000 in seed money for this new endeavor—called the Deep Thought project—was provided by Hsu's adviser, H. T. Kung.

The basic version of Deep Thought's chess engine contains 250 chips, including two processors, that plug into a single circuit board measuring about half again the length of this magazine [*see illustration on page 50*]. The engine is managed by a program—the so-called host software—running on a workstation. The machine's processors have hardly more raw speed than do ChipTest's, but improved control of the search algorithm enables them to search 30 percent more efficiently.

The evaluation hardware has four components. A piece placement evaluation (the only inheritance from ChipTest) scores pieces according to their central placement, their mobility and other considerations. A pawn structure evaluation scores pawns according to such parameters as their mutual support, their control of the center of the board and their protection of the king. A passed-pawn evaluation considers pawns that are unopposed by enemy pawns and can therefore be advanced to the eighth rank and promoted to queens. A file structure evaluation assigns values to more complicated configurations of pawns and rooks on a particular file.

We also began to consider ways of tuning the evaluation function's 120 or so parameters, specified in software. Traditionally, programmers had hand-tuned the weights that programs assigned to material—pawns and pieces—and to positional considerations. We believe ours is the only major program to tune its own weights automatically.

We acquired 900 sample master games and arbitrarily defined the optimum weights as those that produce the best match between the moves the machine judges to be best and those that the masters actually played. The software part of the evaluation function was completely rewritten by Campbell and Nowatzyk to reflect this strategy. Instead of just assigning a final numerical value to each position, the evaluation function—in its tuning mode—returns an equation containing a string of linear terms. In other words, it produces a vector.

Two tuning mechanisms were used. The first, which is called hill climbing, simply sets a given evaluation parameter at an arbitrary value and then performs, say, a five- or six-ply search on every position in the game data base to find the moves that the machine would play. It then adjusts the parameter and recalculates. If the number of matches between the computer's choices and the grandmaster's choice should increase, then the parameter is adjusted again in the same direction. The process continues until all the parameters have reached their highest level of performance. It would take years to optimize all the parameters by this method, however, and so we used it only in a few difficult cases.

The second tuning mechanism, proposed and implemented by Nowatzyk, was much quicker. It evolved from the simple notion of finding the best fit between the function of the machine's evaluation of positions and their presumed true values. The best fit provides the lowest average squared value of the error between the model and the true value. True values can be approximated, for these purposes, by the results returned from deep searches (if a known concept is being fine-tuned) or by comparing machine decisions with those of first-rate human players.

The sample games give strong hints about the relative values of positions: any position reached after a grandmaster's move is, after all, likely to be better than all of the others that would have been reached via alternative moves. Instead of computing a position's value from its parameters, Nowatzyk calculated the parameters on the basis of an assumed difference between the position a grandmaster chose and the alternative positions the grandmaster had rejected. His algorithm takes only a few days to compute, and unlike hill climbing, it does not improve parameters one at a time but improves the entire set of parameters simultaneously.

Our automatically tuned evaluation function appears to be no worse, if no better, than the hand-tuned functions of such well-known academic chess programs as Hitech and Cray Blitz. There still appears to be a gap, however, between Deep Thought's evalua-

ANATOLY KARPOV VERSUS DEEP THOUGHT

HARVARD UNIVERSITY, FEBRUARY 1990

Comments by International Master Michael Valvo

WHITE KARPOV	BLACK DEEP THOUGHT
1 e2-e4	c7-c6
2 d2-d4	d7-d5
3 Nb1-d2	g7-g6
4 c2-c3	Bf8-g7
5 e4-e5	

This move marked the end of Deep Thought's opening book.

... f7-f6!

Fantastic play by a machine! Black immediately attacks the head of the pawn chain and induces White to weaken his control of the light squares.

6 f2-f4 Ng8-h6
7 Ng1-f3 castles
8 Bf1-e2 f6xe5!

Deep Thought finds active counterplay.

9 f4xe5 c6-c5!

So that 10 d4xc5 would allow Black to obtain some advantage with Nf3-g4!

10 Nd2-b3 c5xd4
11 c3xd4 Nb8-c6
12 castles Qd8-b6
13 Kg1-h1 a7-a5

White's central pawn phalanx still guarantees him an edge, but his pawn on d4 is a major weakness and his pieces neither cooperate well nor occupy good squares.

14 a2-a4 Bc8-f5
15 Bc1-g5 Bf5-e4
16 Nb3-c5!

This move, which threatens a simultaneous attack on Black's Queen and Rook on the next move, is probably based on a little insight into the way computers operate. Deep Thought cannot resist capturing the pawn:

... Qb6xb2?

A mistake; better was 16 ... Nh6-f5!, preparing to sacrifice a Rook for a Knight after 17 Nc5-d7, Qb6xb2; 18 Nd7xf8, Nf5xd4! when Black's attack would have been very dangerous.

17 Nc5xe4 d5xe4

Deep Thought now thinks that it is down by the equivalent of one third of a pawn. Karpov later was impressed with the computer's evaluation, which approximated his own during the game.

18 Ra1-b1 Qb2-a3

Forced, as 18 ... Qb2-c3 loses to 19 Rb1-b3, whereas 18 ... Qb2-a2 runs into 19 Nf3-d2 followed by 20 Be2-c4.

19 Bg5-c1 Qa3-c3
20 Bc1-d2 Qc3-a3
21 Bd2-c1 Qa3-c3
22 Rb1-b3 Qc3-a1

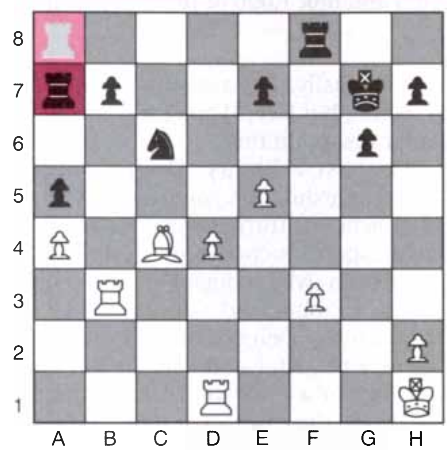
Karpov repeats moves to gain thinking time.

23 Be2-c4(check) Kg8-h8
24 Bc1xh6! Qa1xd1
25 Bh6xg7(ch) Kh8xg7
26 Rf1xd1 e4xf3
27 g2xf3

27 Rb3xb7 was better. But how could Karpov have guessed Deep Thought's next move?

... Ra8-a7!!

DEEP THOUGHT



KARPOV

The audience laughed at this "ugly" move. According to Karpov, however, it is Deep Thought's only chance.

28 Bc4-d5 Rf8-d8
29 Rb3-b5 Ra7-a6!

The computer defends very resourcefully. Now it threatens 30 ... Nc6-a7; 31 Bd5xb7, Na7xb5; 32 Bb7xa6, Rd8xd4; with equality.

30 Bd5-c4 Ra6-a7
31 Bc4-d5 Ra7-a6
32 Rb5-c5 Rd8-d7
33 Kh1-g2 Ra6-b6!
34 Bd5xc6 b7xc6
35 Kg2-f2!

Risky, given Black's slightly superior game, but Karpov still wants to win.

... Rd7-d5
36 Rc5xd5 c6xd5
37 Rd1-c1 Rb6-b4
38 Kf2-e3 Rb4xa4

Another, and perhaps simpler, way to draw is 38 ... Rb4-b3(ch); 39 Ke3-e2, Rb3-b4 with repetition of positions, as White can hardly afford to give up his d-pawn.

39 Rc1-c5 e7-e6
40 Rc5-c7(ch) Kg7-g8
41 Rc7-e7 Ra4-a3(ch)
42 Ke3-f4 Ra3-d3
43 Re7xe6 Rd3xd4(ch)
44 Kf4-g5 Kg8-f7!

A fine defensive interpolation.

45 Re6-a6 a5-a4

Deep Thought thinks it stands better and therefore refuses to force a draw by 45 ... h7-h6(ch); 46 Kg5xh6, Rd4-h4(ch); 47 Kh6-g5, Rh4-h5 (ch); 48 Kg5-f4, Rh5-f5(ch); followed by 49 ... Rf5xe5.

46 f3-f4 h7-h6(ch)
47 Kg5-g4 Rd4-c4?

47 ... g6-g5! draws, but the machine, still thinking it is ahead, refuses to give up a pawn for safety.

48 h2-h4 Rc4-d4
49 Ra6-f6(ch) Kf7-g7
50 Rf6-a6 Kg7-f7
51 h4-h5 g6xh5?

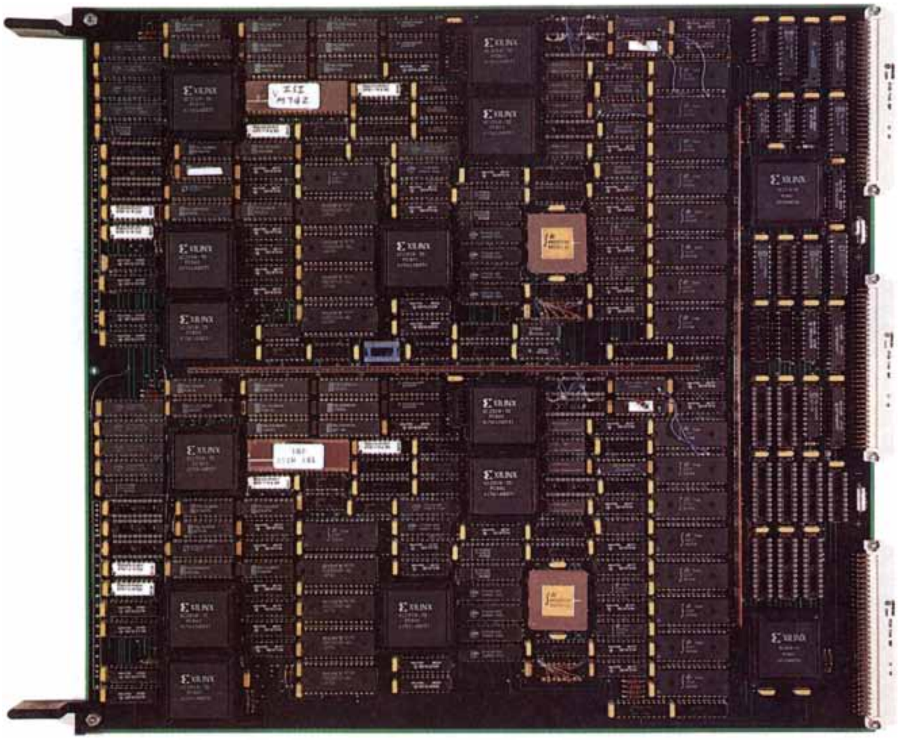
This loses. 51 ... g6-g5 would have held White to an edge too slight to win the game, Karpov later concluded.

52 Kg4-f5 kf7-g7
53 Ra6-a7(ch) Kg7-f8
54 e5-e6 Rd4-e4
55 Ra7-d7 Re4-c4
56 Rd7xd5 h5-h4
57 Rd5-d3 Kf8-e7
58 Rd3-d7(ch) Ke7-f8
59 Rd7-h7 h6-h5
60 Kf5-e5 h4-h3
61 f4-f5 Kf8-g8
62 Rh7xh5 a4-a3
63 Rh5xh3 a3-a2

A human player would play 63 ... Rc4-a4, to prolong the struggle, but then 64 Ke5-f6 would soon prevail.

64 Rh3-a3 Rc4-c5(ch)
65 Ke5-f6 Black resigns

Deep Thought considered itself to be down by the equivalent of at least six pawns, so at this point its designers decided to resign for it. The machine had about 20 minutes in which to make its moves, while Karpov had less than a minute left on his clock; however, that was more than enough time for a player of his caliber to force the win.



HARDWARE HEART of Deep Thought fits on a circuit board the size of a large pizza. Each of its two processors can search 500,000 positions per second. The successor machine will shrink Deep Thought to a chip and link 1,000 of them in parallel.

tion function and those of the top commercial chess machines, which typically are the fruit of many man-years of work. With better feedback from the automatic tuning procedures, we hope to close the breach soon.

It may seem strange that our machine can incorporate relatively little knowledge of chess and yet outplay excellent human players. Yet one must remember that the computer does not mimic human thought—it reaches the same ends by different means. Deep Thought sees far but notices little, remembers everything but learns nothing, neither erring egregiously nor rising above its normal strength. Even so, it sometimes produces insights that are overlooked by even top grandmasters.

The machine's inhuman insights were perhaps the reason behind Grandmaster Kevin Spraggett's decision to engage the machine as an assistant in preparing for a World Championship Candidate Quarter Final match with Grandmaster Artur Yusupov. The machine's participation had no discernible effect on the match, but it did establish an interesting precedent.

In October of 1989 an experimental six-processor version of Deep Thought played a two-game exhibition match against Kasparov in New York City. Although the new version was capable of searching more than two million positions per second, Kasparov disposed of it quite easily. The result was not unexpected, but Deep Thought's play was rather disappointing.

This past February, Deep Thought played an exhibition game against Anatoly Karpov, a former world champion and Kasparov's challenger in the 1990 title match (which begins in New York City in October and concludes in Lyons, France). Defects that surfaced in the experimental software in the six- and four-processor versions led us to revert to the two-processor version. Deep Thought, benefiting from a number of improvements in its evaluation function, played one of its best games in the first 50 moves, then blundered away a clearly drawn position. A stable six-processor version would have had enough speed to avoid the blunder [*see illustration on preceding page*].

Speed is the key to work now under way at the IBM Thomas J. Watson Research Center, where the next-generation machine is now being designed. It should outcalculate its predecessor by a factor of at least 1,000. The machine we have in mind will therefore examine more than a billion positions per second, enough to search 14 or 15 plies deep in most cases and from 30 to 60 plies in forcing lines. If the observed relation between processing speed and playing strength holds, the next-generation machine will play at a 3400 level, about 800 points above today's Deep Thought and 500 points above Kasparov's rating record.

To achieve this speed, Hsu is designing a chess-specific processor chip that is projected to search at least three million moves per second—more than three times faster than the current Deep Thought. He is also designing a highly parallel computing system that will combine the power of 1,000 such chips, for a further gain of at least 300-fold. Anantharaman and Campbell are improving various aspects of the current version of Deep Thought, so that these improvements can be incorporated into the next machine as well. Nowatzyk is pursuing other interests.

We believe the system will be strong enough, by virtue of its speed alone, to mount a serious challenge to the world champion. We further believe that the addition of a long list of other planned improvements will enable the machine to prevail, perhaps as soon as 1992.

Kasparov begs to differ, and we respect his opinion. In a private communication, he acknowledged that a machine searching a billion positions per second might defeat the general run of grandmaster, then added, "But not Karpov and me!" Kasparov contended that the very best players should be able to prepare themselves to exploit the special weaknesses presented by machines. He maintained that human creativity and imagination, in particular *his* creativity and imagination, must surely triumph over mere silicon and wire.

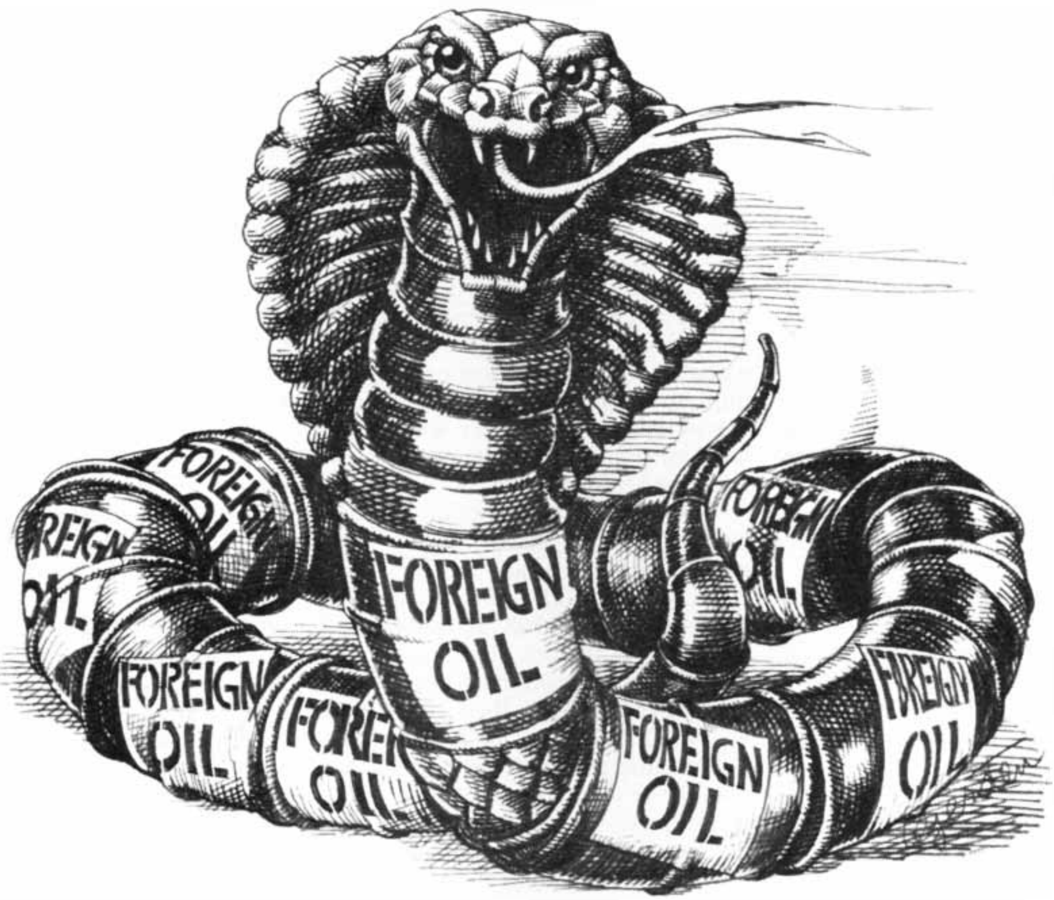
When the two opinions collide over the board, the ingenuity of one supremely talented individual will be pitted against the work of generations of mathematicians, computer scientists and engineers. We believe the result will not reveal whether machines can think but rather whether collective human effort can outshine the best achievements of the ablest human beings.

FURTHER READING

BELLE. J. H. Condon and Ken Thompson in *Chess Skill in Man and Machine*. Second Edition. Edited by P. W. Frey. Springer-Verlag, 1984.

CHESS 4.5—THE NORTHWESTERN UNIVERSITY CHESS PROGRAM. David J. Slate and Lawrence R. Atkin in *Chess Skill in Man and Machine*. Second Edition. Edited by P. W. Frey. Springer-Verlag, 1984.

LARGE SCALE PARALLELIZATION OF ALPHA-BETA SEARCH: AN ALGORITHMIC AND ARCHITECTURAL STUDY WITH COMPUTER CHESS. Feng-hsiung Hsu. Ph.D. Thesis. Carnegie-Mellon University Computer Science Department, CMU-CS-90-108, February, 1990.



DANGEROUSLY UNPREDICTABLE.

We now import 50 percent of all the oil we use. And with the unrest in the Middle East, this dependence has put America's economy in danger.

But the more we use nuclear energy, instead of imported oil, to generate our electricity, the less we have to depend on uncertain foreign oil supplies.

America's 112 nuclear electric plants already have cut foreign oil dependence by 4.3 billion barrels since the oil embargo of 1973, saving us $125 billion in foreign oil payments.

But 112 nuclear plants will not be enough to meet our rapidly growing demand for electricity. We need more plants.

Importing so much oil is a danger America must avoid. We need to rely more on energy sources we can count on, like nuclear energy.

For more information on nuclear energy and energy independence, write to the U.S. Council for Energy Awareness, P.O. Box 66080, Dept. RS10, Washington, D.C. 20035.

U.S. COUNCIL FOR ENERGY AWARENESS

Nuclear energy means more energy independence.

Parental Imprinting of Genes

Even when fathers and mothers contribute identical genes to their offspring, the genes may have different effects. Sex-specific gene imprints may influence normal development and trigger diseases

by Carmen Sapienza

In his early experiments with garden peas during the mid-19th century, Gregor Mendel made an observation that later became a virtual axiom for geneticists. He saw that when he crossed true-breeding round peas with true-breeding wrinkled peas, all the hybrid offspring were equivalent plants with round peas. The result was the same regardless of whether the round-pea plant served as the male or, in the reciprocal case, as the female in the cross. Mendel had discovered the principle of equivalence in reciprocal crosses: no matter which parent contributes a gene to its offspring, the gene will behave in the same way.

The importance of Mendel's observation in the history and practice of genetics cannot be underestimated. It has been shown to be true of many genetic traits, not only of garden peas, but also of fruit flies, mice, human beings and a host of other organisms.

Recently, however, geneticists and embryologists have described some traits that do not obey these rules. Instead the traits are governed by a phenomenon called genome imprinting: a process that temporarily and erasably imprints, or marks, the genes passed on by females and males in different ways. Offspring that receive marked genes from their mothers are consequently different from those that receive the genes from their fathers. In short, sometimes it does matter from which parent a gene is inherited.

Investigators in several laboratories around the world have been striving to identify the molecular nature of the genome imprint, the mechanism by which it is established, and the number and kinds of genes that are imprinted. Although we have not yet been notably successful in this endeavor, we have made some fascinating discoveries along the way, some of which are broadening our understanding of certain cancers, genetic diseases and other maladies. Research on genome imprinting may even turn up previously unsuspected influences on the inheritance of traits covered by classic Mendelian genetics.

Exceptions to the rule of identical reciprocal hybrids have actually been known for many years, but they have generally fallen into one of two categories. Within the first are traits linked to genes on the sex chromosomes, X and Y. Female mammals have two X chromosomes in all their cell nuclei, whereas males have one X and one Y. Color blindness and hemophilia are two of the many traits determined by genes on the X chromosome. The inheritance of these sex-linked traits follows a well-defined pattern that need not be equivalent in reciprocal hybrids.

If a color-blind father crosses with a mother who does not carry that trait, for example, then none of their sons will be color-blind. If the mother is color-blind and the father is not, then all the sons will be color-blind. In both cases, the daughters will carry the gene for color blindness but will not be color-blind. The inheritance and manifestation of the sex-linked trait depends on the sex of the offspring but not directly on the sex of the parent who possesses the trait.

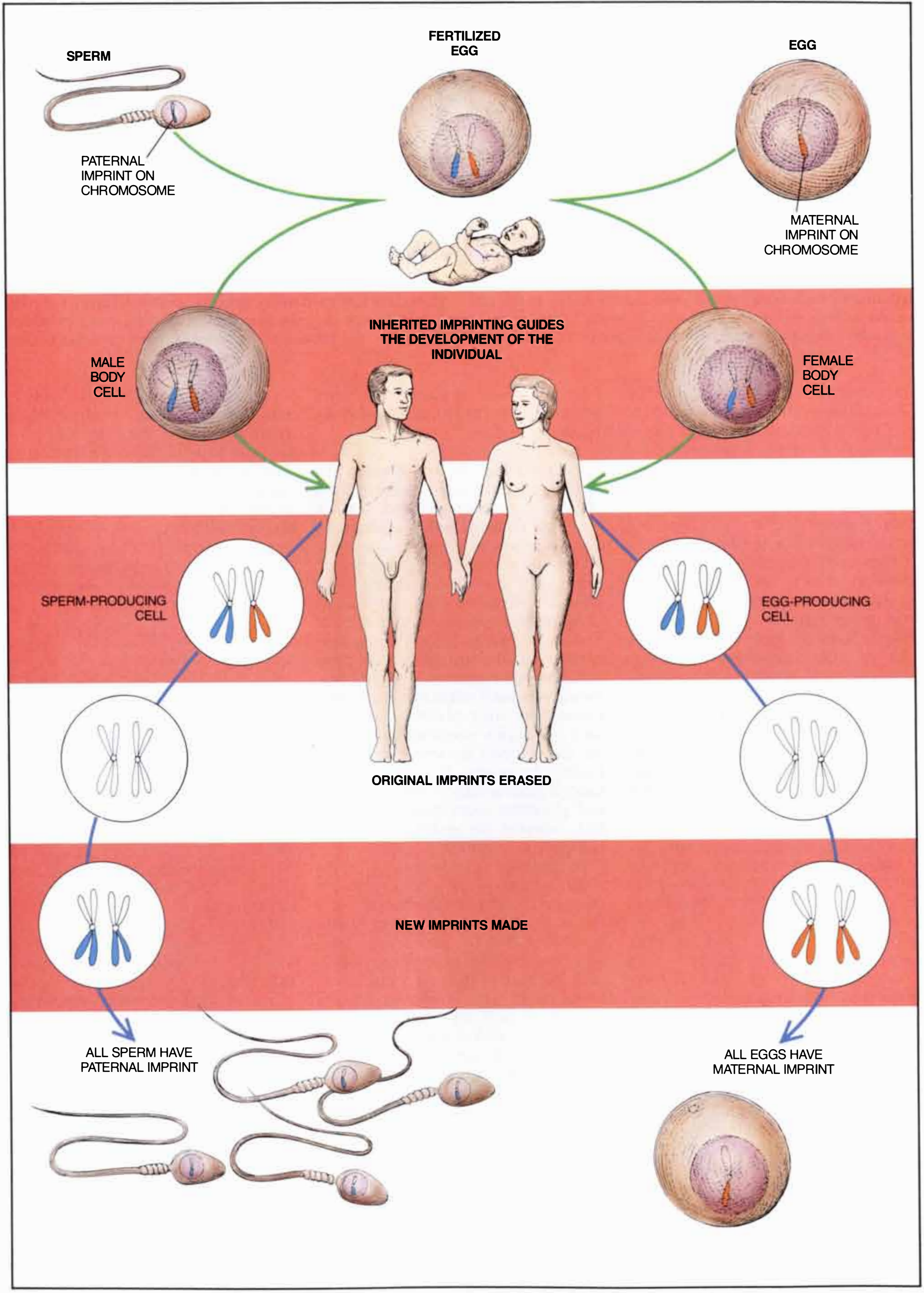
The second class of nonequivalent reciprocal crosses involves traits controlled by genes outside the cell nucleus. Certain subcellular organelles—mitochondria in animal cells and mitochondria and chloroplasts in plant cells—carry their own genetic information within them. These organelles pass between generations in the cytoplasm of the egg cell and are therefore inherited exclusively from the mother. Leaf color in some plants shows this type of inheritance, as does a type of neuromuscular disease in human beings called mitochondrial encephalomyopathy.

Genome imprinting, the third and most recently discovered kind of exception to the rule of equivalent reciprocal crosses, differs markedly from the previously discussed instances. The traits it affects are not necessarily carried by the sex chromosomes (although they can be), nor are they associated with maternally inherited organelles. Imprinting can, at least in theory, affect any gene in a cell. The sex-linked and maternal organelle-linked exceptions hinge on unequal genetic contributions by each parent. In contrast, because of genome imprinting, the parents may contribute exactly identical genes to their offspring, but if the genes have been differently imprinted, they will not have identical effects. These novel features have attracted the attention of biologists.

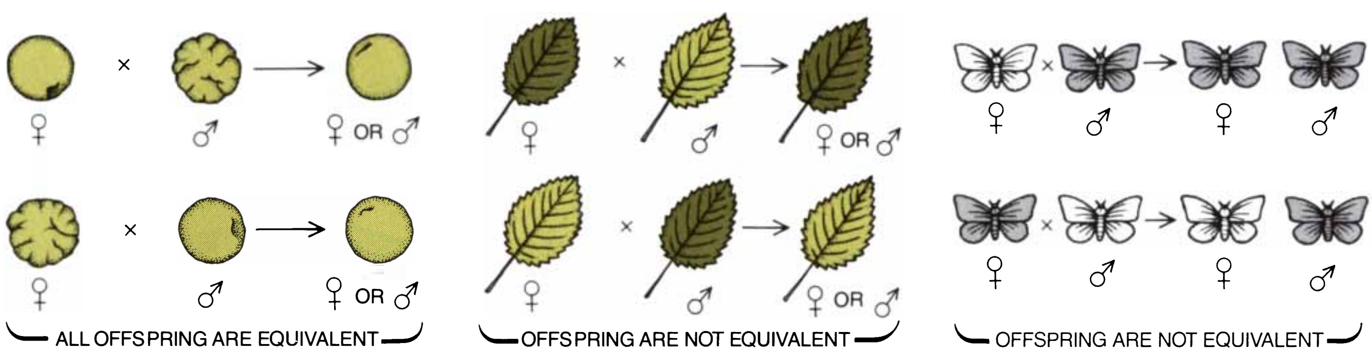
Much of the current interest in genomically imprinted traits stems from experiments performed by James McGrath and Davor Solter of the Wistar Institute of Anatomy and Biology in Philadelphia and M. Azim H. Surani and his colleagues of the Institute of Animal Physiology in Cambridge, England. McGrath and Solter developed an elegant microsurgical procedure called nuclear transfer, which enables them to exchange physically the genetic information of one

CARMEN SAPIENZA is a pioneer in the study of how genetic imprinting may influence the development of certain genetic diseases and cancers. He is head of the Laboratory of Developmental Genetics at the Ludwig Institute for Cancer Research in Montreal. Away from the laboratory, Sapienza is an avid baseball fan and a reader of fiction.

GENE IMPRINTS are carried on the chromosomes of sperm and egg cells. Both sexes therefore inherit some genes with a paternal imprint and some with a maternal imprint. When individuals produce sperm or eggs, however, the old imprints are erased and new ones specific to the sex of the individual are imposed on all the gamete chromosomes.


SPERM
FERTILIZED EGG
EGG
PATERNAL IMPRINT ON CHROMOSOME
MATERNAL IMPRINT ON CHROMOSOME
INHERITED IMPRINTING GUIDES THE DEVELOPMENT OF THE INDIVIDUAL
MALE BODY CELL
FEMALE BODY CELL
SPERM-PRODUCING CELL
EGG-PRODUCING CELL
ORIGINAL IMPRINTS ERASED
NEW IMPRINTS MADE
ALL SPERM HAVE PATERNAL IMPRINT
ALL EGGS HAVE MATERNAL IMPRINT

RECIPROCAL CROSSES



BREEDING experiments in which the gene for a trait is contributed by the mother in one case and the father in the other are called reciprocal crosses. For most traits, such as roundness in peas, it does not matter which parent carries the gene (*left*). Certain traits, however, such as leaf color in some plants, are inherited solely from the mother (*center*). For traits linked to sex chromosomes, such as wing color in *Abraxas* moths, the inheritance pattern is more complex (*right*).

mouse embryo for that of another. Their technique takes advantage of the fact that after a sperm fertilizes an egg but before the cells of the embryo begin to divide, the egg nucleus and the sperm nucleus remain briefly separate in the egg's cytoplasm. Each nucleus is visible under a light microscope and has a characteristic position and size.

With a very thin, hollow glass needle, the workers were able to remove selectively either the egg-derived nucleus or the sperm-derived nucleus (or both) from a fertilized egg. They were then able to replace the removed nucleus with one from another fertilized egg. McGrath and Solter found that if they substituted one sperm-derived nucleus for another, or if they replaced one egg-derived nucleus with another, the resulting embryos developed fully and were indistinguishable from normal mice of the same strain.

Nuclear transfer also enabled the researchers to create embryos with unusual genetic constitutions: embryos with both sets of chromosomes from one parent. McGrath, Solter and Surani became curious about how such embryos would develop. They therefore created gynogenotes (embryos with two complete sets of maternal chromosomes) and androgenotes (embryos with two sets of paternal chromosomes) and compared their development with that of normal embryos.

It is important to note that because the mice used in some of these experiments had been inbred for many generations, the males and females carried identical sets of chromosomes (except, of course, that the females did not have a Y chromosome and the males had only one X chromosome). If the development of a mouse embryo depends solely on its collection of genes, then theoretically it should not matter whether a mouse receives all its genes from one parent: gynogenotes, androgenotes and normal mice should all develop identically.

In fact, they do not develop the same way. Neither embryos with two maternally derived sets of genes nor embryos with two paternally derived sets of genes develop to term: usually they stop developing after dividing into only a few dozen cells. When development does continue, as occasionally happens, the gynogenotes and androgenotes display intriguingly different abnormalities. In the most advanced gynogenotes, the embryos themselves have relatively minor aberrations, but their placentas and yolk sacs (structures essential for nourishing the embryo) are severely stunted. Exactly the reverse is seen in the advanced androgenotes: the yolk sacs and placentas seem more nearly normal, whereas the embryos are puny and poorly developed.

Because the DNA sequences in the chromosomes of the gynogenetic, androgenetic and normal embryos were the same, McGrath, Solter and Surani concluded that the genes had somehow been modified or imprinted differently because of their maternal or paternal origin. This imprinting process seemed to turn off selectively some genes that would otherwise act during early embryonic development. Among the paternally contributed genes, some relevant to embryonic development appeared to be inactive; genes critical to the formation of the yolk sac and placenta seemed to be inactive in the maternal contribution.

Another graphic demonstration of imprinting was provided by Bruce M. Cattanach, a geneticist at the Medical Research Council in Oxford. Cattanach has been interested in the phenomenon of nonequivalent reciprocal crosses for many years. One of the tools he created to assist his study of these traits are lines of mice that carry one or more pairs of fused chromosomes. That is, in these mice, some chromosomes are physically connected and so are unable to separate during meiosis, the cell-division process that produces egg and sperm cells. Consequently, one dividing cell will occasionally receive two copies of a chromosome, and another cell will receive none.

If an egg with two copies of a chromosome is fertilized by a sperm without any, the resulting embryo will contain the normal number of chromosomes, but both copies of one chromosome will have come from the mother. Similarly, both copies will come from the father if the sperm carries both chromosomes and the egg lacks them. According to the rule of equivalence in reciprocal hybrids, all animals resulting from such crosses should be identical. Yet Cattanach's experiments have unequivocally shown that the rule does not hold true for several mouse chromosomes.

Cattanach found, for example, that mice inheriting both copies of chromosome 11 from one parent varied radically from the norm in size and weight. If raised under identical conditions, mouse pups with two chromosomes 11 from their mother were abnormally small, whereas those with two chromosomes 11 from their father were gigantic. Similar experiments done with other fused chromosome pairs produced other characteristic aberrations, including embryonic death and behavioral abnormalities.

Cattanach's experiments also demonstrated another important point: the effects of the genome imprint do not persist into the next generation, which means that the imprint is not a permanent modification of a chromo-

Zenith Data Systems Presents Everything You Love About The PC ...Plus The Benefits Of Graphical Computing.

ZENITH DATA SYSTEMS INNOVATES AGAIN™

Zenith Data Systems was the first to offer Microsoft® Windows™ version 3.0 and Asymetrix® ToolBook® *pre-installed* on every hard drive 386-based desktop PC. And that's just part of *The Seamless Solution* we've created for today's graphical computing environment.

With Microsoft Windows version 3.0 and the fingertip simplicity of the Microsoft Mouse, your Zenith Data Systems PC lets you glide graphically, *seamlessly* between applications. While Asymetrix ToolBook lets you design your own applications under Windows version 3.0. So your productivity will soar. With the greatest of ease.

A $643 value,* Microsoft Windows version 3.0, Asymetrix ToolBook and the Microsoft Mouse are all yours with every Intel386,™ 386 EISA or 386SX desktop PC.

Add our award-winning 14" VGA Flat Technology Monitor, and you've got *The Seamless Solution.* So your Windows version 3.0 and ToolBook applications will come to life with breakthrough clarity.

Zenith Data Systems. Transforming the PC into a powerful graphic environment. For more information and the name of your nearest Zenith Data Systems Medallion Reseller, call:

1-800-523-9393

The Zenith Data Systems Sphere

The universal symbol of simplicity, the sphere perfectly represents The Seamless Solution℠ from Zenith Data Systems.

ZENITH data systems

Groupe Bull

Microsoft and *Windows version 3.0* are trademarks of Microsoft Corporation. Asymetrix and ToolBook are registered trademarks of Asymetrix Corporation. Intel386 is a trademark of Intel Corporation.
*Based on suggested retail price if purchased separately. © 1990 Zenith Data Sys

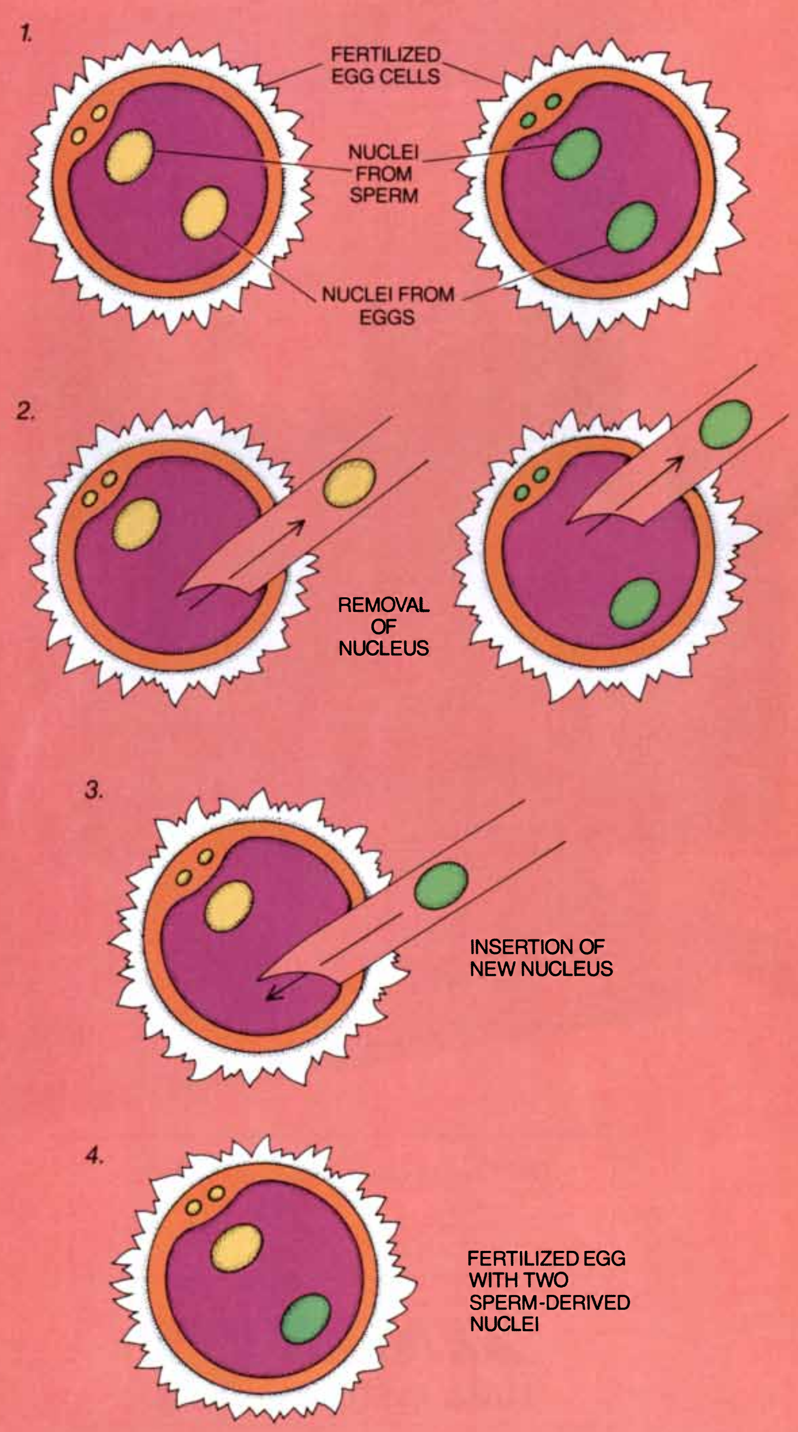
some. A small male mouse that received both copies of chromosome 11 from its mother would usually produce offspring of normal size. In some way, then, the genes that he received from his mother must be stripped of their female imprint. They must then be re-marked as the genes of a male.

Inspired by the discoveries of McGrath, Solter, Surani and Cattanach, investigators in several laboratories (including my own) have sought to determine whether the genome imprint could be the result of a direct modification of DNA. In particular, we have studied the role of a process called DNA methylation, in which small molecular groups are chemically linked to the large DNA molecule [see "A Different Kind of Inheritance," by Robin Holliday; SCIENTIFIC AMERICAN, June, 1989]. Much to our excitement, experiments have often shown that particular genes are methylated differently, depending on whether they are maternally or paternally inherited, and that these modifications are erasable. No one has yet been able to prove, however, whether such modifications are the primary mechanism for imprinting the genome or whether they merely reflect some hidden, more fundamental biochemical phenomenon.

Not all my colleagues agree with me, but I would wager that DNA methylation is only a consequence of the primary mechanism. Fortunately for all of us studying the subject, we can still do many important experiments without knowing the ultimate causes of genome imprinting. However genome imprinting is accomplished, it reveals itself in ways that are often strange and sometimes heartbreaking.

NUCLEAR TRANSFER METHOD

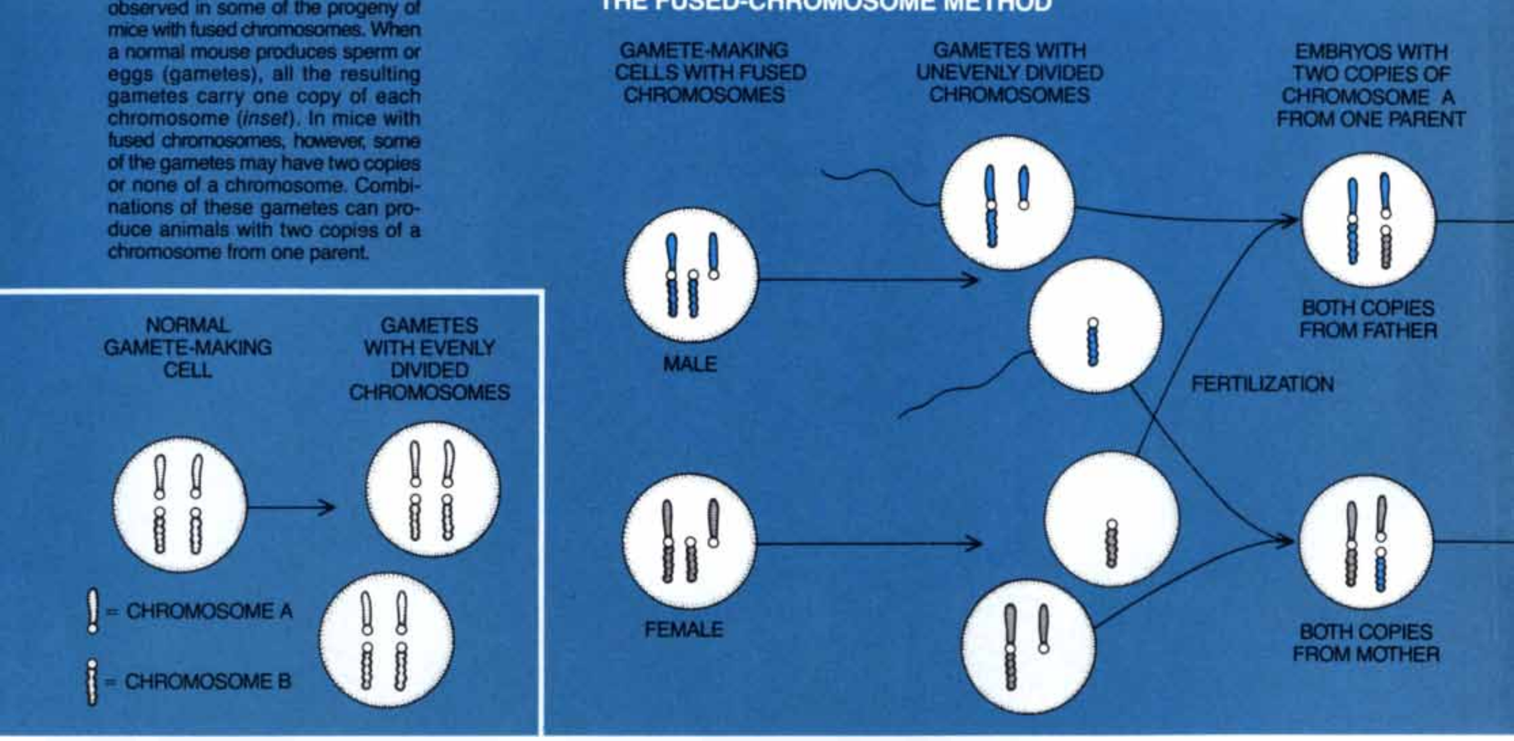
One technique for studying genome imprinting involves making embryos with chromosomes that all come from one sex.



The heartbreaking manifestations appear in several human diseases. Nature occasionally provides human parallels to conditions like those studied by Cattanach in mice. Robert D. Nicholls and his colleagues at Harvard Medical School recently made such a discovery about patients with Prader-Willi syndrome, a disease with symptoms of mental retardation, extreme obesity, short stature and disproportionately small hands and feet. After examining the genomes of Prader-Willi patients and their parents, Nicholls concluded that many patients had inherited both copies of chromosome 15 from their mothers.

Nicholls, Joan H. M. Knoll, also of Harvard, and their colleague Charles A. Williams of the University of Florida then made a related but contrasting discovery about patients with Angelman syndrome. (In the past, such persons were often described as "happy puppets" who exhibited excessive laughter, jerky movements and other symptoms of motor and intellectual retardation.) These patients frequently had partial deletions of their maternally inherited chromosome 15, with the result that only their paternal chromosome 15 was present in its entirety.

It is striking that two diseases with such different clinical features may be linked to differential imprinting of the same genes on the same chromosome. But unlike the abnormally large and



observed in some of the progeny of mice with fused chromosomes. When a normal mouse produces sperm or eggs (gametes), all the resulting gametes carry one copy of each chromosome (*inset*). In mice with fused chromosomes, however, some of the gametes may have two copies or none of a chromosome. Combinations of these gametes can produce animals with two copies of a chromosome from one parent.

small mice that Cattanach produced with his experiments, Prader-Willi syndrome and Angelman syndrome cannot be easily explained as opposite sides of the same coin, caused by an excess or shortage of the same gene products. The studies by Nicholls and his colleagues highlight the difficulty of predicting how particular traits may react to the imprinting process. It may therefore be profitable to reexamine many inherited human disorders for evidence of imprinting effects.

Even traits for which there is no obvious need to invoke the influence of genome imprinting may be affected. This contention may be illustrated by several childhood cancers, including embryonal rhabdomyosarcoma (a muscle tumor), Wilms' tumor (a form of kidney cancer) and osteosarcoma (a form of bone cancer).

To explain how genome imprinting may be involved in these diseases, a description of how cancers are thought to occur may be in order. Many cancers are thought to result from the accumulation of successive mutations in specific genes within a single cell. Chromosome 11, for example, carries a gene, called *Rd*, that is part of a family of what are variously known as recessive oncogenes, tumor suppressor genes or anti-oncogenes [see "Finding the Anti-Oncogene," by Robert A. Weinberg; SCIENTIFIC AMERICAN, September, 1988]. Without the product of *Rd*, a muscle cell transforms into a cancer cell and eventually gives rise to a rhabdomyosarcoma tumor.

Because each cell nucleus contains two copies of chromosome 11, both copies of *Rd* must be inactivated before a cell will transform. The first copy of *Rd* may be turned off in a variety of ways, the simplest being a mutation that alters the DNA sequence of the gene itself. If such a mutation occurs, then any additional genetic event that inactivates the remaining copy of *Rd* will probably cause the cell to become tumorigenic. Frequently the genetic event that destroys the second *Rd* either deletes the chromosome completely or deletes the part of it containing the gene.

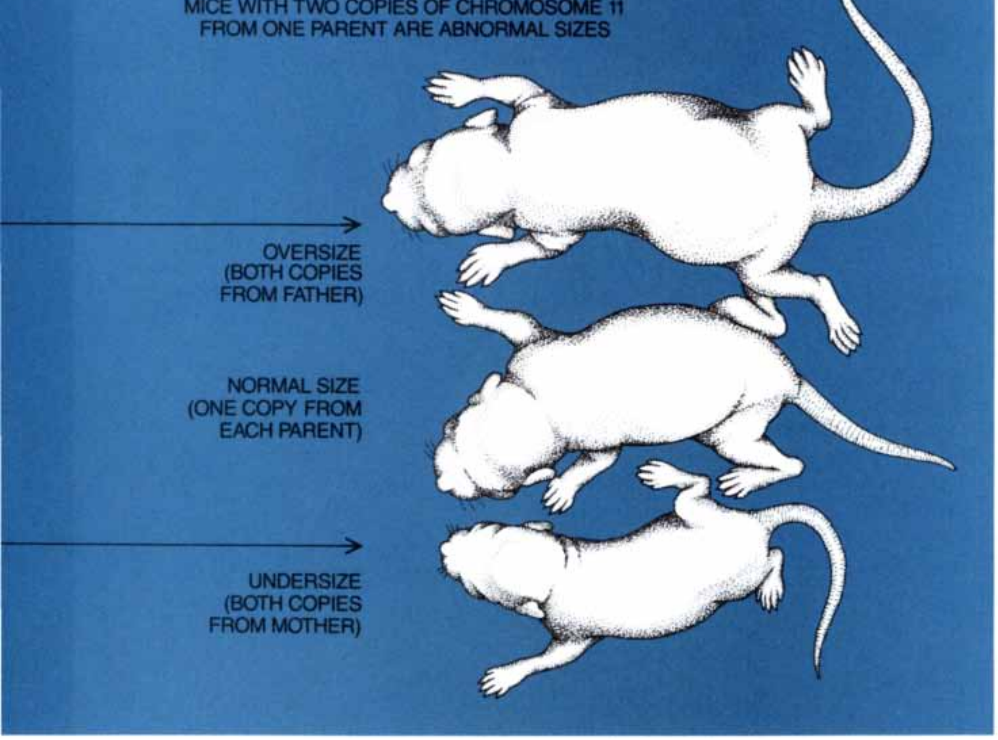
I and my colleagues Heidi J. Scrable and Webster K. Cavenee of the Ludwig Institute for Cancer Research in Montreal have proposed an alternative mechanism for some pediatric cancers. We suggest that the event inactivating the first copy of a recessive oncogene is not a true mutation. Instead it may be genome imprinting, which turns off genes differently in males and females. (The second event, which would trigger the final transformation of the cell, would probably still be a chromosome deletion or other accident.) According to our hypothesis, the rhabdomyosarcoma tumor cells should have an imprinted, inactive copy of *Rd* on their surviving chromosome 11, and that chromosome should come from the same parent in most cases of embryonal rhabdomyosarcoma.

Our laboratory and those of Grady F. Saunders of the M.D. Anderson Cancer Center in Houston, Manfred Mannens of the University of Amsterdam, Masao Sasaki of the University of Kyoto and several others have indeed found evidence for our theory. The first inactivation event in embryonal rhabdomyosarcoma, Wilms' tumor and osteosarcoma almost always occurs in a recessive oncogene carried on a paternal chromosome. The data support the involvement of genome imprinting in the genesis of these tumors—yet they also pose a major theoretical difficulty.

The difficulty arises because most of us who studied genome imprinting thought of it as a consequence of the vastly different physiology and biochemistry of gamete (sperm and egg) formation in males and females. Throughout most of their adult lives, males produce billions of sperm from a continuously dividing population of stem cells. Females, conversely, are born with all the eggs they will produce during their reproductive years. As these eggs mature, one or a few at a time, they are ovulated regularly.

If differential imprinting only reflected the dichotomy between male and female production of gametes, then it would seem that all males should imprint their genomes in one way and all females should imprint their genomes in another way. Lacking evidence to the contrary, we had assumed that all members of a sex would imprint their genomes identically.

But the data linking pediatric cancers to inactive, paternally derived recessive oncogenes belied our assumption. If



all males imprinted and inactivated the *Rd* gene, for example, then the incidence of embryonal rhabdomyosarcoma would be very high: just one genetic change in the maternally donated *Rd* would cause some cells in any individual to become tumorigenic. Because these tumors are rare (affecting approximately one in 20,000 children), it is unlikely that everyone carries an inactive recessive oncogene inherited from his or her father. Conversely, in people with these diseases, that is precisely what seems to have happened.

A tempting explanation for this discrepancy is that not all males (or females, presumably) imprint and inactivate the same genes. What might be the basis for such differences between individuals?

If one is a geneticist at heart, then genetics is always the first place to which one turns for answers to such questions. A genetic explanation requires the existence of one or more genes responsible for the imprinting. (These should not be confused with those genes that are themselves imprinted.) In other words, imprinting must be a process arising from the activity of varied genes that operate differently in males and females. Persons of the same sex with different imprint-controlling genes will also differ in their constellations of imprinted genes. Most males, for example, will not turn off the *Rd* gene, but rare individuals who have an aberrant copy of an imprint-controlling gene will.

The idea that the expression or lack of expression of a gene may be controlled by other genes is not new: it describes an old, well-known phenomenon in genetics called dominance modification. Many traits respond to the activity of other genes that are said to modify their expression. Genome imprinting may be viewed as a special case of dominance modification. The only unusual feature of the modifier genes postulated to control genome imprinting is that they act differently in males and females.

Huntington's disease (HD) is an interesting example of a trait that appears to be modified by sex-specific genome imprinting. A fatal neurological disorder, HD is inherited as a dominant trait, which means that anyone who inherits the HD gene from one parent will get the disease. HD generally strikes individuals in mid-life; the average age at diagnosis is 38 years. Approximately 10 percent of the cases, however, are characterized as juvenile onset and affect children as young as two and a half years.

As medical researchers noticed early in their studies, approximately 90 percent of affected children come from families in which the father is the affected parent. Clearly, in these cases, both the fathers and their children have the same disease trait, but modification of the trait through paternal inheritance has caused it to become obvious at a much earlier age.

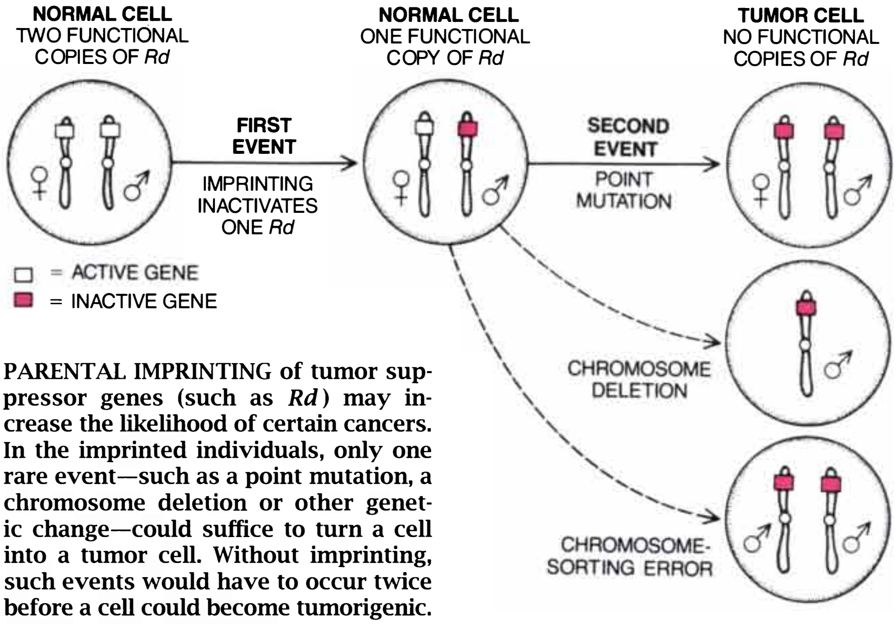
Many hypotheses have been put forward over the years to explain the genetic behavior of HD, but the model proposed by Charles D. Laird of the University of Washington comes closest to explaining the data. Laird's innovation was to model the genetics of HD after a specific class of variably expressed traits, called variegating position effects, in fruit flies. These traits are unusual because, unlike most other mutant traits, they are not expressed by every cell in the affected tissues. Instead the tissues become mosaics of some cells that appear mutated and other cells that appear perfectly normal. The proportion of cells that express the mutation is controlled by modifier genes that can swing the balance from almost wholly normal tissue to almost completely mutant tissue.

Laird reasoned that if the varying age of HD onset reflected a variable mosaicism in the expression of the HD gene, then the mosaicism was probably influenced by modifier genes. A modifier gene that pushed the balance of mosaicism to a more nearly mutant tissue would result in an earlier age of HD onset, whereas a modifier gene that pushed the balance toward normal tissue would result in a later age of onset. (There are reports of HD patients who were not affected until their 70s.)

Laird further explained that the preponderance of paternally inherited juvenile-onset cases resulted from the location of the modifier gene on the X chromosome. Because males have only one X chromosome, any aberration of its modifier gene would not be compensated for, as it would in females with two X chromosomes. Males are therefore more likely to have offspring in which the balance of mosaicism is pushed toward more nearly mutant tissue and an earlier age of HD onset. Ten percent of juvenile-onset cases of HD do come from families in which the mother, and not the father, is affected, but such cases are to be expected: females with aberrant modifier genes on both X chromosomes should exist, although they would be rare.

This model seems to explain the genetics of this complicated disease well. In my own work, I have slightly refined Laird's model to reflect more accurately the way in which the modifier genes must behave, and this small addition provides an even closer approximation to the real genetic behavior of HD.

To my mind, Laird's variegating position-effect model of the genetics of HD brings the study of genome imprinting almost full circle. A variegating position effect in fruit flies was the subject of the first in-depth study of genome imprinting of which I know, that of Janice Spofford of the University of Chica-



PARENTAL IMPRINTING of tumor suppressor genes (such as *Rd*) may increase the likelihood of certain cancers. In the imprinted individuals, only one rare event—such as a point mutation, a chromosome deletion or other genetic change—could suffice to turn a cell into a tumor cell. Without imprinting, such events would have to occur twice before a cell could become tumorigenic.

go. Spofford's work, first published in 1959, apparently received only limited attention. Other reports of genetic traits affected by which parent had transmitted a particular gene have also appeared sporadically during the past 30 years. Traits in organisms as diverse as fruit flies, yeast, maize, mice and humans seem to be affected by genome imprinting.

Nevertheless, genome imprinting is generally regarded, even by many biologists, as a curiosity that affects only a very few traits. On several occasions, I have been asked why I waste my time (and, by inference, the questioner's) on a phenomenon of such little importance. I have always replied that the number of traits affected by the genome imprinting is unknown but likely to be large. My response usually elicits an incredulous, open-mouth stare, followed by a one- or two-minute lecture on Mendel's principles, which culminates in a table-thumping pronouncement that most traits do not work "that way."

Up to a point, such critics are correct. If most traits showed absolute dependence on the parent from which the relevant genes came, geneticists would certainly have noticed. The key to the disagreement, however, may well be the level at which researchers examine a particular trait. When Spofford studied the mosaic expression of mutant traits in fruit flies, she was able to observe the same traits in all the offspring of reciprocal crosses, just as Mendel had. Yet there were still sufficient differences in the expression of the traits for her to discriminate which parent had contributed them.

Similarly, work in my own laboratory and that of my colleague Alan C. Peterson of the Ludwig Institute has shown that several traits in mice are mosaic. If one asks whether these traits are merely present, the answer is yes, but if one asks how strongly the traits are expressed, then the effects of parental imprinting can be seen. For some traits, the degree of expression may not be very important, but for others, it may be critical: for example, it may spell the difference between having Wilms' tumor or not, or manifesting HD at the age of five rather than 70.

Indeed, one of the most potentially fruitful areas for the investigation of genome imprinting may be in the inheritance of complex human diseases. How much this approach will contribute to the understanding of these diseases remains to be seen, but the theoretical, if not practical, successes we have achieved in the understanding of Huntington's disease and some pediatric cancers makes me hopeful that progress can be made.

FURTHER READING

DIFFERENTIAL ACTIVITY OF MATERNALLY AND PATERNALLY DERIVED CHROMOSOME REGIONS IN MICE. B. M. Cattanach and M. Kirk in *Nature*, Vol. 315, No. 6019, pages 496–498; June 6, 1985.

DIFFERENTIAL IMPRINTING AND EXPRESSION OF MATERNAL AND PATERNAL GENOMES. D. Solter in *Annual Reviews of Genetics*, Vol. 22, pages 127–146; 1988.

A MODEL FOR EMBRYONAL RHABDOMYOSARCOMA TUMORIGENESIS THAT INVOLVES GENOME IMPRINTING. H. Scrable et al. in *Proceedings of the National Academy of Sciences*, Vol. 86, No. 19, pages 7480–7484; October, 1989.

"Cambridge SoundWorks May Have The Best Value In The World. A Winner."

David Clark, *Audio* Magazine, Sept. '89

Henry Kloss, creator of the dominant speaker models of the '50s (Acoustic Research), '60s (KLH), and '70s (Advent), brings you Ensemble, a genuinely new kind of speaker system for the '90s, available only factory direct from Cambridge SoundWorks.

Cambridge SoundWorks has created Ensemble,™ a speaker system that provides the sound once reserved for large, costly speakers. It virtually disappears in your room. And because we market it directly, Ensemble costs hundreds less than it would in stores.

Your living room works *with* Ensemble, not against it.

No matter how well a speaker performs, at home your room takes over. Putting other speakers where the room helps the bass may hinder the upper ranges, or vice-versa. Ensemble consists of four units: *two* bass units and two high frequency satellites. The compact woofer units produce the bass that normally requires large speakers. Place them behind furniture, on bookshelves or under a couch. The small satellites blend into *any* decor. The result: a full range, musically accurate speaker system without big boxes.

Try Ensemble in your home risk-free for 30 days.

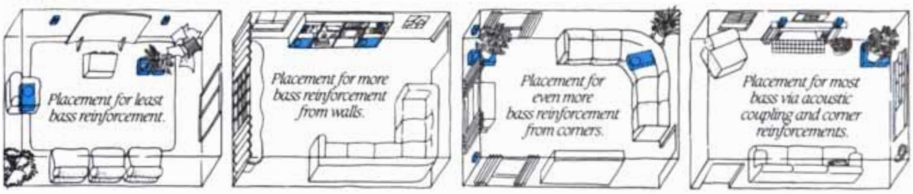
Ensemble is sold factory direct...by helpful, experienced audio experts. So you can audition Ensemble the *right* way—in your home for 30 days, with no risk, no salesman hovering nearby. At $499* it is *the* value on today's speaker market. For literature, expert advice, or to order, call 1-800-252-4434 9 AM to Midnight, (ET), seven days a week.

Canada 1-800-525-4434 Fax: 617-332-9229.

Ensemble is a Trademark of Cambridge SoundWorks, Inc. Reprinted with the permission of Audio Magazine, © 1989 Diamandis Communications.

Unlike seemingly similar satellite systems which use a single large sub-woofer, Ensemble uses two separate, compact bass units. They fit more gracefully into your living environment, and help minimize the effects of the listening room's standing waves.

Made In U.S.A.



CAMBRIDGE SOUNDWORKS

Suite 105OCT 154 California St., Newton, MA 02158

☐ Send more information on all products.
☐ Send Ensemble (black-laminate woofers) $599.*
☐ Send Ensemble (vinyl-clad woofers) $499.*
I'm paying by ☐ Check ☐ MC ☐ Visa ☐ AmEx

Acct. Number__________ Exp._____
Signature__________
Name__________
Address__________
City_____ State_____ Zip_____
Phone (Area Code)_____ Number_____

FOR IMMEDIATE SERVICE: 1-800-AKA-HIFI.
We ship worldwide, including APO and FPO.
MA residents add 5% sales tax.
*With vinyl-clad utility bass units. $599 with furniture-grade black-laminate bass units. Plus freight ($7–$25). Delivery time usually 2–7 days.

Here's your chance to achieve a small moral victory.

Gabriel Cortez
Colombia
Age 4

What would you do if you saw a lost, frightened child?

You'd probably stop, pick him up, brush away his tears and help him. And there's a reason.

You know what's right.

And right now, you can do just that. Through the Childreach Sponsorship program of PLAN International USA (formerly Foster Parents Plan) – one of the most respected sponsorship organizations in the world for over 53 years. Act on instinct and help a desperately poor child faraway. A child who may not have enough to eat. The chance to go to school. Or hope.

For only $22 a month, you can change a child's life through better nutrition, health programs, schooling. And you can help the sponsored child's family and community work together for a better future.

You'll receive the child's picture and personal story and in letters and progress reports, you'll see for yourself the difference your *personal* caring is making. Imagine. Only 72 cents a day could change a child's life!

Please don't wait.

If you saw a helpless child on the street, you'd reach out that instant. Reach out now to a child overseas. Achieve a small moral victory today!

Become a Childreach Sponsor with PLAN International USA.

Call

1-800-222-6356

childreach
SPONSORSHIP
Help so personal, you touch a child for life.

YES, I want to reach out and make a difference.

Enroll me as a Childreach Sponsor to . . .
___ **the child who needs my help most**
___ Girl ___ Boy ___ Either
___ South America
___ Central America/Caribbean
___ Asia
___ Africa
☐ My $22 check for the first month is enclosed. Please send my Sponsorship Kit with my child's picture and personal story.
☐ I haven't made a decision yet, but I'd like to know more about the child I could help.

☐Mr. ☐Mrs. ☐Miss ☐Ms.
Name __________
Address __________
City __________
State _____ Zip _____

Mail to: PLAN International USA
P.O. Box 877 A32
Providence, RI 02901

PLAN INTERNATIONAL USA

Childreach Sponsorship is a program of PLAN International USA, formerly Foster Parents Plan. Non-profit. Non-political. Non-sectarian. Tax deductible. A copy of our financial report is available upon request from N.Y. Dept. of State, Office of Charities Registration, Albany, N.Y., or PLAN International USA.

L453

The Photorefractive Effect

A laser beam passing through a crystal can suddenly burst into a spray of light. This photorefractive effect may be the key to developing computers that exploit light instead of electricity

by David M. Pepper, Jack Feinberg and Nicolai V. Kukhtarev

When Arthur Ashkin and his colleagues at Bell Laboratories first noticed the photorefractive effect some 25 years ago, they considered the phenomenon a curiosity at best and a complete nuisance at worst. Today photorefractive materials are being shaped into components for a new generation of computers that exploit light instead of electricity.

Ashkin was experimenting with a crystal of lithium niobate ($LiNbO_3$) that he hoped would convert one color of intense laser light to another (a process technically called second-harmonic generation). As part of his tests, he directed a laser beam through the crystal. At first, the crystal performed quite admirably, allowing light to pass through undisturbed. But after a few minutes, the crystal began to distort the beam, scattering light around the laboratory. Somehow the laser light had altered the optical properties of the crystal itself. This photorefractive effect would persist in the crystal for days. If the workers bathed the crystal in a uniform beam of light, however, the crystal would once again transmit an undistorted beam.

DAVID M. PEPPER, JACK FEINBERG and NICOLAI V. KUKHTAREV wrote part of this article on the beaches of Hawaii and simultaneously participated in an experiment on the photochromic effect: they developed suntans. Pepper is a senior staff physicist in the optical physics department at Hughes Research Laboratories and adjunct professor at Pepperdine University. In 1980 he received his Ph.D. in applied physics from the California Institute of Technology. This is his second article for SCIENTIFIC AMERICAN. Feinberg is associate professor of physics and electrical engineering at the University of Southern California. In 1977 he earned his Ph.D. in physics from the University of California, Berkeley. Kukhtarev is a senior scientist at the Institute of Physics in Kiev, Ukraine, U.S.S.R. In 1983 he received a doctorate degree for his studies of the theory of dynamic holography.

During the past 25 years investigators have discovered a wide variety of photorefractive materials, including insulators, semiconductors and organic compounds. Photorefractive materials, like film emulsions, change rapidly when exposed to bright light, respond slowly when subjected to dim light and capture sharp detail when struck by some intricate pattern of light. Unlike film, photorefractive materials are erasable: images can be stored or obliterated at whim or by design.

By virtue of their sensitivity, robustness, and unique optical properties, photorefractive materials have the potential to be fashioned into data-processing elements for optical computers. In theory, these devices would allow optical computers to process information at much faster rates than their electronic counterparts. Employing photorefractive materials, workers have already developed the optical analogue to the transistor: if two laser beams interact within a photorefractive material, one beam can control, switch or amplify the second beam. Photorefractive materials also lie at the heart of devices that trace the edges of images, that connect networks of lasers and that store three-dimensional images.

Because the optical properties of photorefractive materials can be modified by the very light that passes through them, they are categorized as nonlinear optical media. In linear optical media—such as lenses, prisms and polarizing filters—light beams merely pass through one another, without changing the properties of the material.

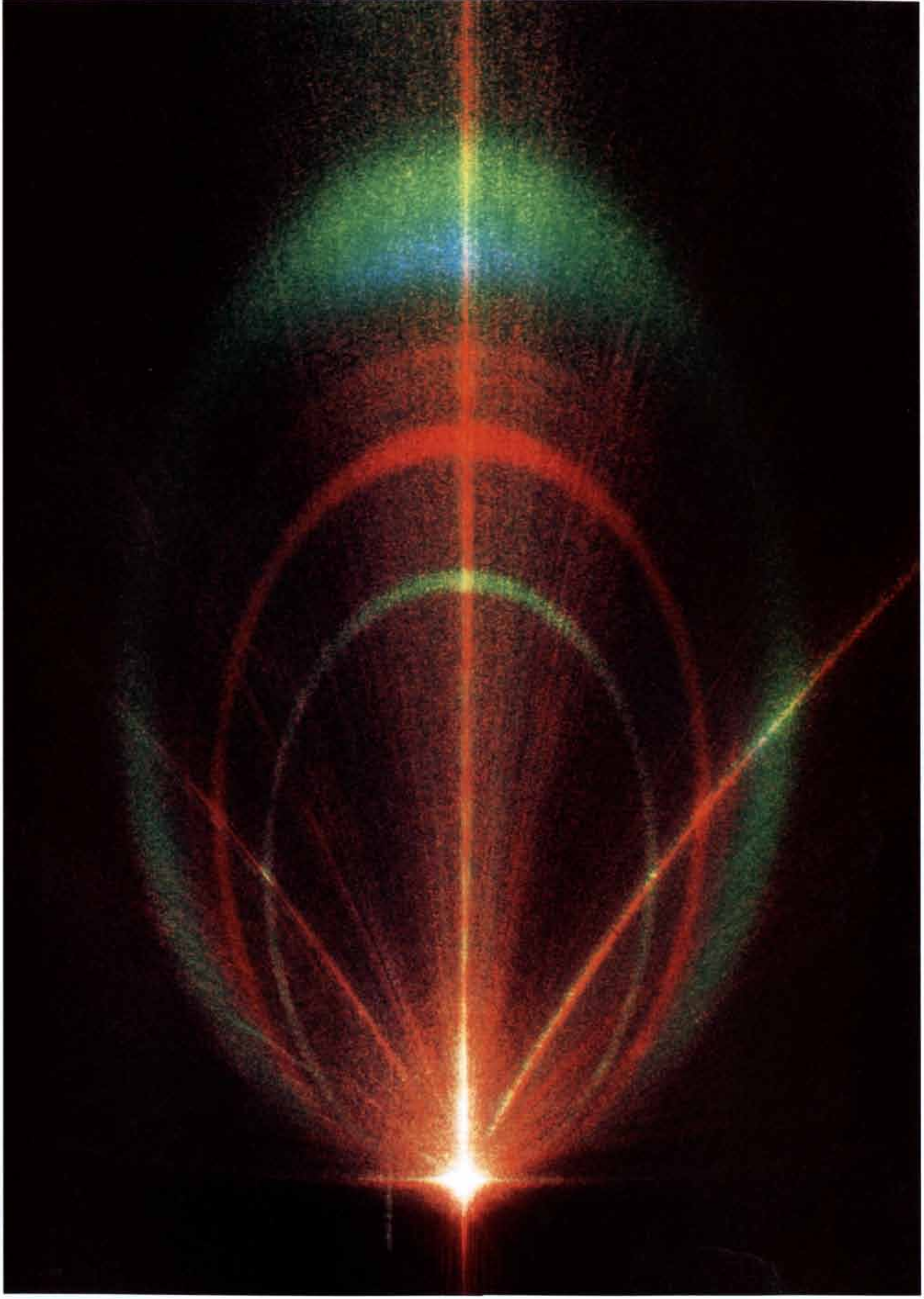
The photorefractive effect is closely related to another nonlinear phenomenon known as the photochromic effect. The light that strikes a photochromic material can change the amount of light that the medium absorbs. Photochromic materials, which are incorporated in certain brands of sunglasses, darken in bright sunlight and lighten in dark rooms.

In photorefractive materials, the light that bombards the material affects how fast light travels through it. More specifically, the photorefractive effect is a process in which light alters the refractive index of a material. (The refractive index is the ratio of the speed of light in a vacuum to that in the material.)

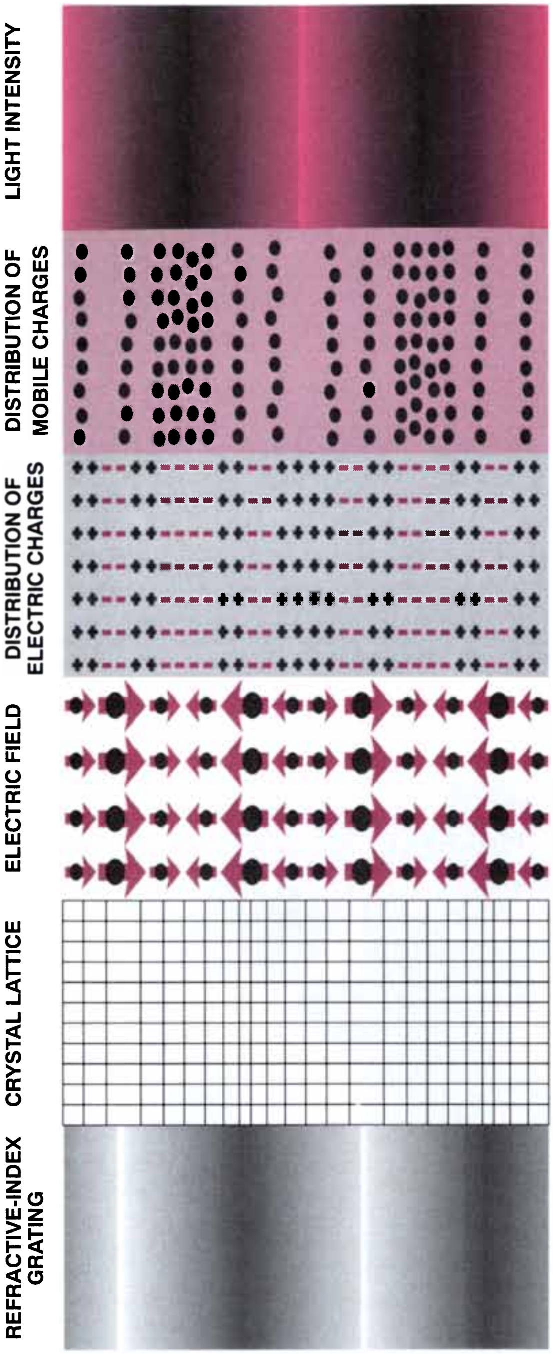
Most transparent materials will change their refractive index if bombarded by light of sufficient intensity. Light is a traveling electromagnetic wave whose electric field strength is proportional to the square root of the intensity of light. For instance, an optical beam whose intensity is 100 million watts per square centimeter is equivalent to an electric field strength of about 100,000 volts per centimeter. When such intense light is directed at a transparent material, it disrupts the positions of the atoms, changing the refractive index by a few parts in one million. As a result of the change, the material can act like a prism or a lens to deflect light.

The term "photorefractive" is usually reserved, however, for materials whose refractive index changes in response to light of low intensity. In photorefractive materials, light beams as weak as one thousandth of a watt per square centimeter can alter the arrangement of atoms in a crystal, changing the refractive index as much as a few parts in 10,000. And unlike most transparent materials, the change in photorefractive crystals is semipermanent: if an altered crystal is isolated from all sources of light, the change in the refractive index can last from milliseconds to years, depending on the material. In this manner, one can store information in the form of images in a crystal.

LASER BEAM striking a 5-mm crystal of barium titanate scatters into a fan of light owing to the photorefractive effect.



HOW LIGHT ALTERS THE OPTICAL PROPERTIES OF A PHOTOREFRACTIVE CRYSTAL



1. Two laser beams interfere in a photorefractive crystal to form a pattern of bright and dark regions.

2. Mobile electrons migrate away from bright regions of the crystal.

3. Mobile electrons accumulate in the dark regions, leaving regions of positive charge.

4. An electric field forms between regions of positive and negative charge.

5. The electric field distorts the crystal lattice.

6. The distortion causes light to travel slower through some regions and faster through others. More specifically, the refractive index is altered periodically. The refractive-index grating is shifted one quarter of a period in space from the light pattern.

How can a weak beam of light cause such a strong change in the refractive index of a crystal? In the late 1960s F. S. Chen of Bell Laboratories advanced the basic model of the photorefractive effect. Just as a single ant can move a large mound of sand one grain at a time, a weak beam of light can gradually build up a strong electric field by moving electric charges one by one. In photorefractive crystals, charges diffuse away from bright regions and pile up in dark regions. As more and more charges are displaced, the electric field inside the crystal increases, attaining a strength as high as 10,000 volts per centimeter. The electric field will distort the crystal lattice slightly (about .01 percent), thereby modifying the refractive index.

The source of these electric charges apparently lies in defects in the crystal lattice of the material. The defects can be mechanical flaws in the lattice structure (missing atoms at certain lattice sites), substitutional dopants (a foreign atom at some lattice site) or interstitial dopants (a foreign atom wedged between native atoms). Very small amounts of these defects, on the order of parts per million, can cause the photorefractive effect.

Each crystal defect can be the source of an extra charge, which can be either electrons (particles of negative charge) or holes (regions of positive charge), depending on the particular crystal. In the dark, these charges are trapped; in the presence of light, they are free to roam within the crystal until they eventually become caught again. If light illuminates charges in one region of the crystal, they will diffuse away from that region and accumulate in the dark, in the way cockroaches scurry underneath furniture to avoid light.

Each charge that moves inside the crystal leaves behind an immobile charge of the opposite sign. In the region between these positive and negative charges, the electric field is strongest, and the crystal lattice will distort the most. A beam of light that passes through this region of the crystal will experience a different refractive index from that of the unaffected regions.

The time it takes for light to rearrange charges in a crystal depends on the intensity of the light and also on how fast charges migrate in the crystal. Weak light takes longer than strong light to build up the same electric field. For low-intensity light (about .01 watt per square centimeter), it can take minutes for the charges to reach their equilibrium pattern. For high-intensity light (about a billion watts per square centimeter), the response time can be less than a nanosecond. A photorefractive crystal, like photographic film, requires a certain amount of light to complete its "exposure."

The change in refractive index is linearly proportional to the strength of the electric field if the crystal lattice lacks a certain property called an inversion symmetry. The electric field will remain in the crystal long after the light is removed, just as the mound of sand remains in its new location long after the ants have left.

One of the most useful consequences of the photorefractive effect is the exchange of energy between two laser beams, which is also known as two-beam coupling. If two laser beams of the same frequency intersect, they will interfere and produce a stationary pattern of bright and dark regions—or more specifically a pattern whose intensity varies sinusoidally with position in the crystal. If this sinusoidal pattern of light forms within a photorefractive crystal, electric charges will move to generate an electric field whose strength also varies sinusoidally. The resultant field will distort the

crystal lattice in a similar periodic manner, producing changes in the refractive index. Ultimately a "refractive-index grating" (also called a refractive-index volume hologram) will be formed within the crystal.

The electric field and the refractive-index grating will have the same periodicity as the light pattern, but they will be shifted by one quarter of a period in space from the incident light. This displacement—called a 90-degree phase shift—is the optimal configuration for the exchange of energy between the original two laser beams.

Once the refractive-index grating has been established in the crystal, some of the light of one beam will be deflected, or diffracted, by the grating in the direction of the other beam (and vice versa). Hence, the two deflected beams will interfere with the two original beams—constructively in one case and destructively in the other. In the case of constructive interference, the peaks of the light waves in one of the deflected beams combine with the peaks of one of the original beams, and the beams therefore reinforce each other. In the case of destructive interference, the peaks of the waves from the other deflected beam combine with the valleys from the other original beam, and the light waves diminish each other. The beam formed from constructive interference will emerge from the crystal stronger than when it entered, whereas the beam formed from destructive interference will emerge weaker. Hence, one of the beams will have gained energy from the other. Which beam gains and which beam loses is determined by the orientation of the crystal and whether the charge carriers are holes or electrons.

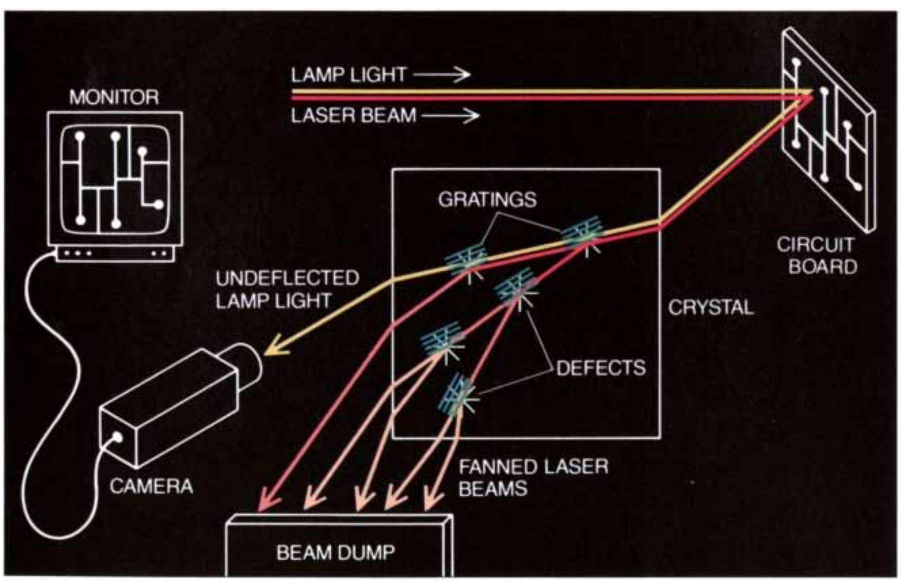
Photorefractive materials exhibit two-beam coupling because the optical pattern and the refractive-index grating are shifted in space. Two-beam coupling is not found in most nonlinear materials, however, because they respond "locally" to optical beams (for example, atomic orbitals are deformed by the intense electric fields of the laser beams). In most nonlinear materials, therefore, the optical pattern and grating precisely overlap. The light deflected by the grating interferes with each of the undeflected beams in exactly the same way. Thus, the two beams exchange an equal amount of energy, so neither grows in intensity.

To enhance the photorefractive effect, investigators have learned to control the flow of charge within a material. The two mechanisms that control the flow of charge in a crystal are diffusion and drift. They are analogous to the diffusion and drift of smoke from a burning ember. Left on its own, smoke will diffuse to regions of low smoke density. If a slight breeze is blowing, however, the smoke will drift in a particular direction. The particles in the smoke behave like mobile charges in a photorefractive material: the charges tend to move toward regions of low charge density, and they drift in response to any electric field.

The simple diffusion of charges from bright regions to dark regions of a crystal does not produce the strongest electric field possible. In 1981 Jean-Pierre Huignard and Abdellatif Marrakchi of Thomson-CSF Laboratories in Orsay, France, applied an electric field externally to a photorefractive crystal to build up a spatially varying field stronger than that produced by diffusion alone. The applied electric field, however, shifted the refractive-index grating away from the optimal quarter-cycle phase shift.

To prevent this nonoptimal spatial shifting of the grating, Sergei I. Stepanov and Mikhail P. Petrov of the A. F. Ioffe Physico-Technical Institute of the Soviet Academy of Sciences in Leningrad developed a clever technique. When they applied an external electric field that rapidly alternated its direction, the charges would preferentially drift in one direction for the first half cycle of the applied field and in the opposite direction for the next half cycle. The process is similar to having two people alternately blow on a burning ember from opposite sides. The resulting smoke pattern is both intensified and spread out farther in space but has the same average location as if no net wind were present. In photorefractive crystals the process yields an internal electric field larger than that produced by diffusion alone, and the refractive-index grating has the same average quarter-cycle phase shift as if no drift field were present.

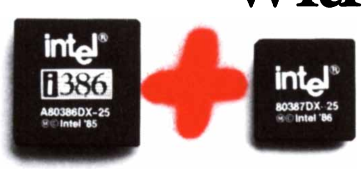
Workers have used this technique to enhance the efficiency of two-beam coupling as well as an effect called beam fanning (technically known as stimulated, forward photorefractive scattering). Discovered in the 1970s, beam fanning is perhaps one of the most intriguing nonlinear optical phenomena. It can be observed, for example, when a pencil-thin, weak beam from a helium-neon laser illuminates a crystal such as barium titanate ($BaTiO_3$). Initially the beam passes through the crystal unaltered. After a second or so (the time depends on the intensity of the light), the beam begins to spread out in the crystal, curving to one side. In the process the curved beam divides into many rays that appear to spray out into a broad fan of light—hence the term "beam fanning." Depending on the choice of photorefractive crystal, the emerging light in cross section can



COHERENT OPTICAL EXCISOR scatters intense light from a laser but transmits weak light from a lamp. Here the excisor protects a camera from the damaging laser rays. The laser beam, which consists of a single frequency of light, scatters off defects in the photorefractive crystal and then interferes with itself. The interference pattern creates a refractive-index grating. As a result, most of the laser light is scattered to one side of the crystal, a process called beam fanning. Because the lamp light consists of many frequencies, neither an interference pattern nor a grating is created, and so most of the lamp light is transmitted through the crystal.

INTEL MATH COPROCESSORS

Multiply 386™ System Performance With This Simple Addition.



Every 386™-based personal computer is powered by an Intel 386 microprocessor. And every one has a socket inside just waiting for you to plug in extra power with one of our Intel 387™ Math CoProcessors.

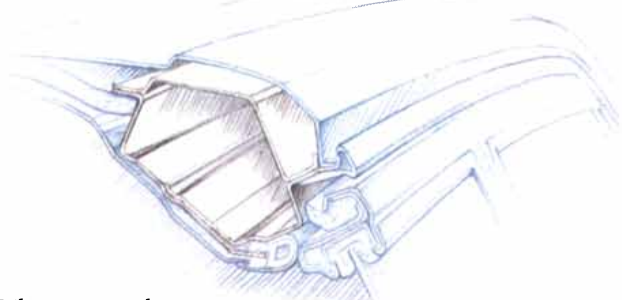
Add one to your 386-based PC to make over 1200 software programs run up to five times faster. That includes spreadsheets, business graphics, accounting and CAD programs.

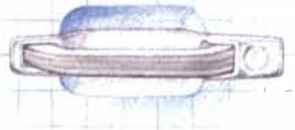
We created Intel Math CoProcessors to be fully compatible with the Intel microprocessor inside your system. So there's total peace of mind that they're fully compatible with your system and your software.

Intel 387 Math CoProcessors. It's a matter of simple addition.

intel®

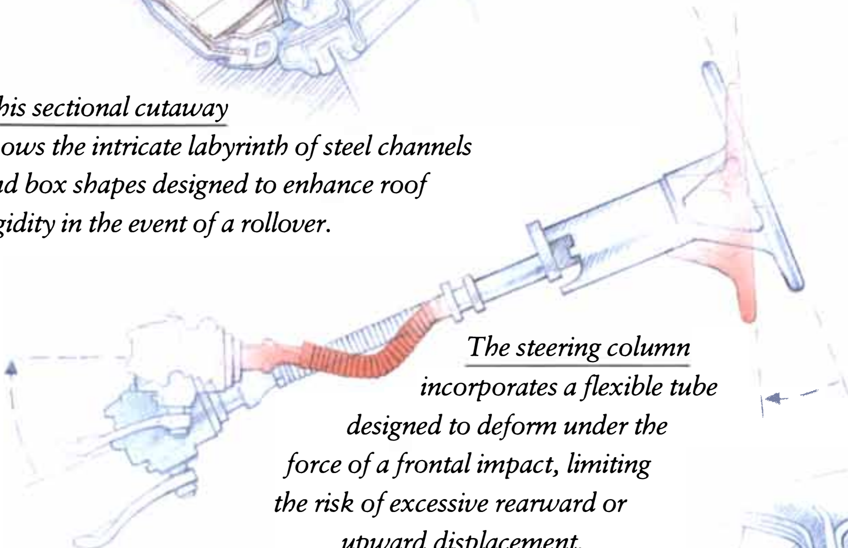
Call 1-800-538-3373 for our new guide with details on how you can boost your software's performance using an Intel Math Coprocessor.

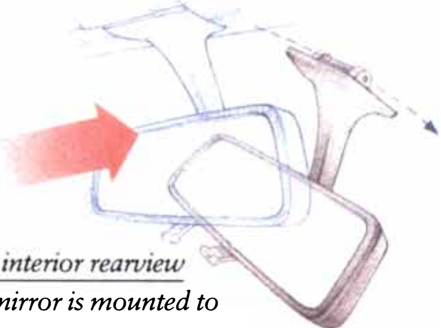




The exterior door handle is not a styled flap but a sturdy grip. Its looped shape is meant to permit maximum pulling force should rescue assistance ever be needed.

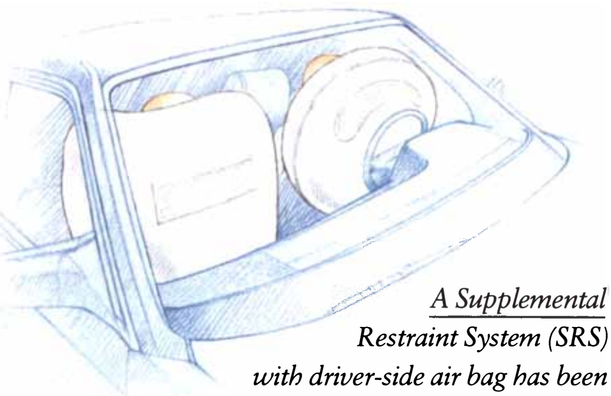
This sectional cutaway shows the intricate labyrinth of steel channels and box shapes designed to enhance roof rigidity in the event of a rollover.





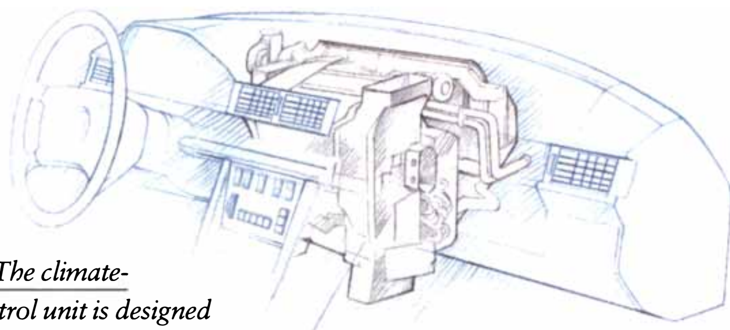
The interior rearview mirror is mounted to break away if struck with moderate force.

The steering column incorporates a flexible tube designed to deform under the force of a frontal impact, limiting the risk of excessive rearward or upward displacement.

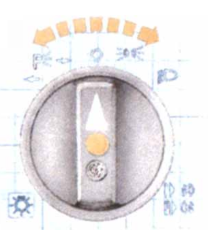


The frontal structure is designed to absorb and channel kinetic energy not only in head-on but also offset frontal impacts—more frequent and more severe. Mercedes-Benz pioneered both the basic energy-absorbing body concept and this offset enhancement.

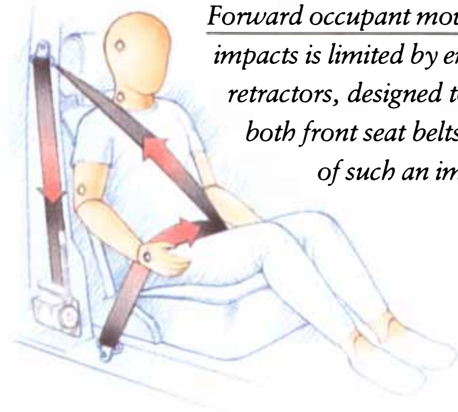
A Supplemental Restraint System (SRS) with driver-side air bag has been standard in every Mercedes-Benz since 1985. On many models, the system now includes both driver and front passenger air bags.



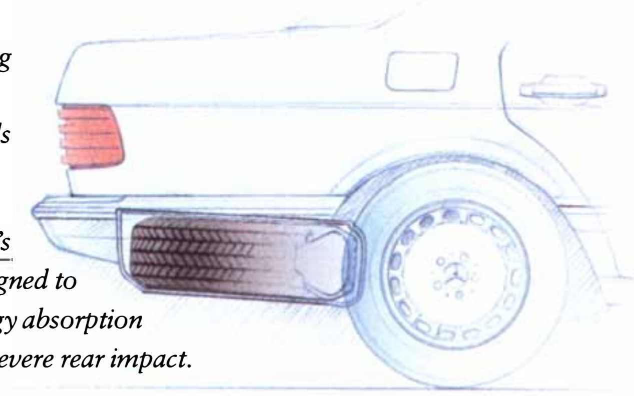
Like every interior control and lever, this recessed headlight-control switch is designed and shaped and placed to help reduce the chance of occupant injury in a severe impact.



The climate-control unit is designed to be crushable in a severe impact, minimizing the risk of its being pushed rearward into the passenger area.

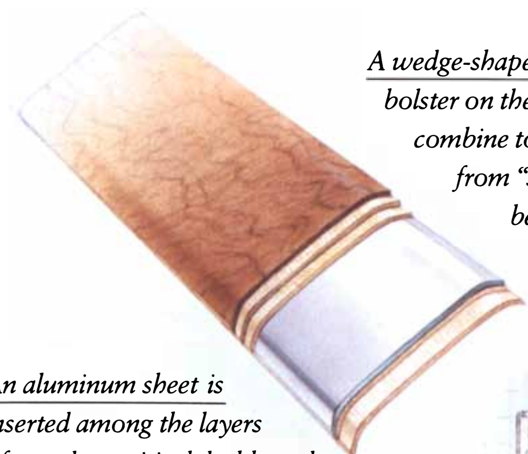


Forward occupant movement in certain impacts is limited by emergency tensioning retractors, designed to tighten slack in both front seat belts within milliseconds of such an impact.

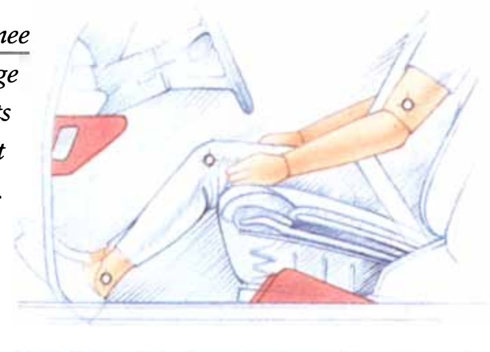


The spare tire's placement is designed to add extra energy absorption in case of a severe rear impact.

A crash course in Mercedes-Benz



A wedge-shaped seat insert and a padded knee bolster on the instrument panel's lower edge combine to help prevent front occupants from "submarining" under their seat belts in a severe frontal impact.

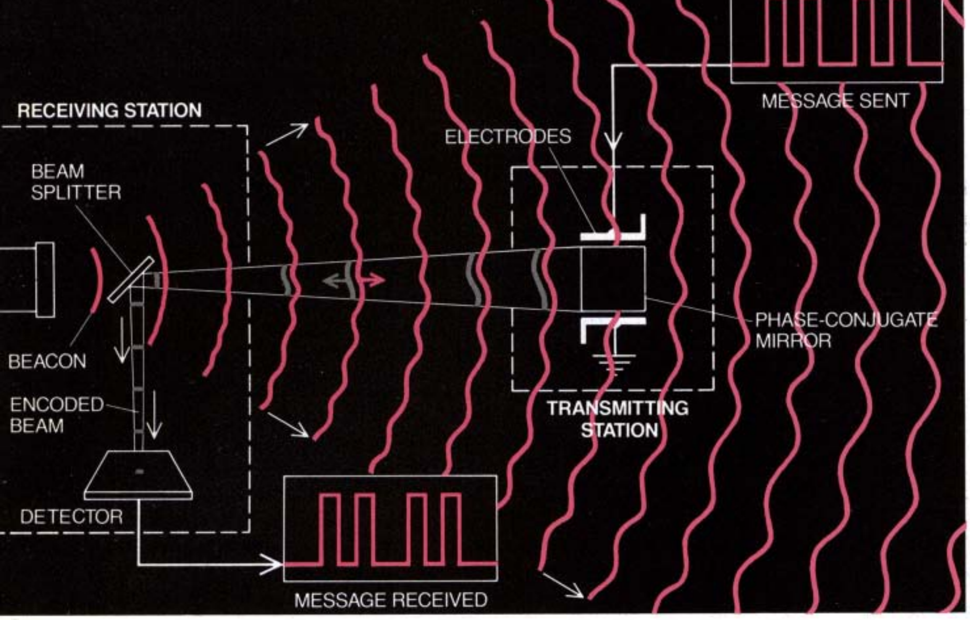


An aluminum sheet is inserted among the layers of wood on critical dashboard areas to help prevent splintering under the force of a direct impact.

Recognizing the importance of keeping occupants inside the car in a severe impact, Mercedes-Benz places extreme importance on door-lock design. This cone-type lock was patented in 1959 and has since been steadily refined.

The most effective single safety element is still the seat belt. So please, buckle up—even if you drive a Mercedes-Benz. For more information about Mercedes-Benz safety, call 1-800-243-9292 or visit your authorized Mercedes-Benz dealer. Some of the safety features depicted vary from model to model.

ENGINEERED LIKE NO OTHER CAR IN THE WORLD



PHASE-CONJUGATE MIRROR made from a photorefractive crystal allows communication through the atmosphere. The receiving station emits a beam of light. As the beam travels through the atmosphere, it spreads out and distorts. At the transmitting station a photorefractive crystal either reflects or transmits the light that hits it, depending on the voltage applied through the electrodes. Because the crystal acts as a phase-conjugate mirror, the reflected light is time-reversed. A message can be encoded by alternately reflecting and transmitting the light. As the time-reversed message beam interacts with the atmosphere, all distortions are removed from the beam, and the message can be decoded at the receiving station.

assume the shape of a beautiful array of elliptical patterns and can have differing polarization components.

Beam fanning results from an energy exchange between the incident beam of light and light randomly scattered by crystal defects. Light scattered by a defect interferes with the unscattered beam, forming a refractive-index grating because of the photorefractive effect. Once this occurs, the unscattered light can transfer additional energy to the scattered beam. It turns out that the energy-exchange mechanism is not isotropic, and so the scattered beams of light become preferentially intensified only over a finite range of angles. This intensified, scattered beam is then randomly scattered by other crystal defects, and the process is repeated. As a result, a large number of scattered beams fan out from the original beam. Depending on the orientation of the crystal and the point at which the incident laser beam enters the crystal, the fanned beam can be made to curve through shallow angles (barely curving through the crystal) or over rather large angles (striking an extreme corner of the crystal).

Beam fanning and other photorefractive effects have been exploited in such applications as coherent beam excisors, optical interconnection elements, phase-conjugate mirrors, novelty filters and edge enhancement of images.

The coherent optical excisor—first described in 1985 by Mark Cronin-Golomb and Amnon Yariv of the California Institute of Technology—can potentially filter out light that scatters from high-intensity laser beams. The excisor has the potential to protect sensitive optical detectors, which may be necessary to monitor high-power lasers during industrial processes such as annealing and welding. The excisor can also prevent damage to video cameras [*see illustration on page 66*].

The key component of the coherent optical excisor is a photorefractive crystal, which is placed in front of the detector or the camera's lens. The intense beam can be diverted away from the detector because the beam-fanning effect essentially guides the coherent laser light to the side as it travels through the crystal. The detector or camera can view the object in the presence of intense light, however, because the background room light scattered from the rest of the object passes through the crystal essentially unaltered.

Beam fanning also plays an important role in phase-conjugate mirrors made from photorefractive materials. These mirrors possess the peculiar property that an optical beam "reflected" from them will travel exactly backward in space as if time were reversed. Because of this property, phase-conjugate mirrors have myriad applications in the fields of optical communications, high-power lasers and optical computing. As an example, they can be incorporated into a system to correct undesirable aberrations that laser beams sometimes acquire during propagation through distorting media or powerful laser amplifiers [see "Applications of Optical Phase Conjugation," by David M. Pepper; SCIENTIFIC AMERICAN, January, 1986].

In 1977 Robert W. Hellwarth of the University of Southern California suggested a basic configuration for phase-conjugate mirrors, and two years later Sergei G. Odulov and one of us (Kukhtarev) and independently Huignard and his co-workers produced such a mirror that incorporated photorefractive materials. In 1982 one of us (Feinberg) serendipitously discovered a class of phase-conjugate mirrors that many investigators use today. Feinberg had focused three laser beams on a crystal of barium titanate. One beam contained the light waves whose time-reversed replica was sought; the two additional "pump" beams were needed to form the phase-conjugate mirror (or so Feinberg and his colleagues thought at the time). To check the experiment, Feinberg blocked the pump beams to ensure that the presumed time-reversed beam did not arise merely from a simple reflection from a crystal face. At first, the time-reversed beam obediently vanished. But after a short time, the time-reversed beam surprisingly reappeared. Feinberg had found a phase-conjugate mirror that required only a single beam. Feinberg's elegant device is an example of a more general class of self-pumped, phase-conjugate mirror, which was pioneered by Jeffrey O. White, Mark Cronin-Golomb, Baruch Fischer and Amnon Yariv of Cal Tech.

Although phase-conjugate mirrors can be made from many classes of nonlinear optical materials, photorefractive elements have several distinct advantages: first, the mirrors require only one input beam—the very beam whose phase-conjugate replica is sought—thus forming the so-called self-pumped phase conjugator; and second, very low laser powers and intensities can initiate the process leading to the time-reversed beam.

Why does nature love the phase-conjugate beam? A partial answer to this question posed by Hellwarth can be advanced at least in the case of barium titanate, as postulated by Kenneth MacDonald, then at the University of Southern California, and one of us (Feinberg). After a short time, beam fanning caus-

"The field cannot well be seen from within the field."

Ralph Waldo Emerson

To know what is possible tomorrow you must be willing to step outside of what is possible today.

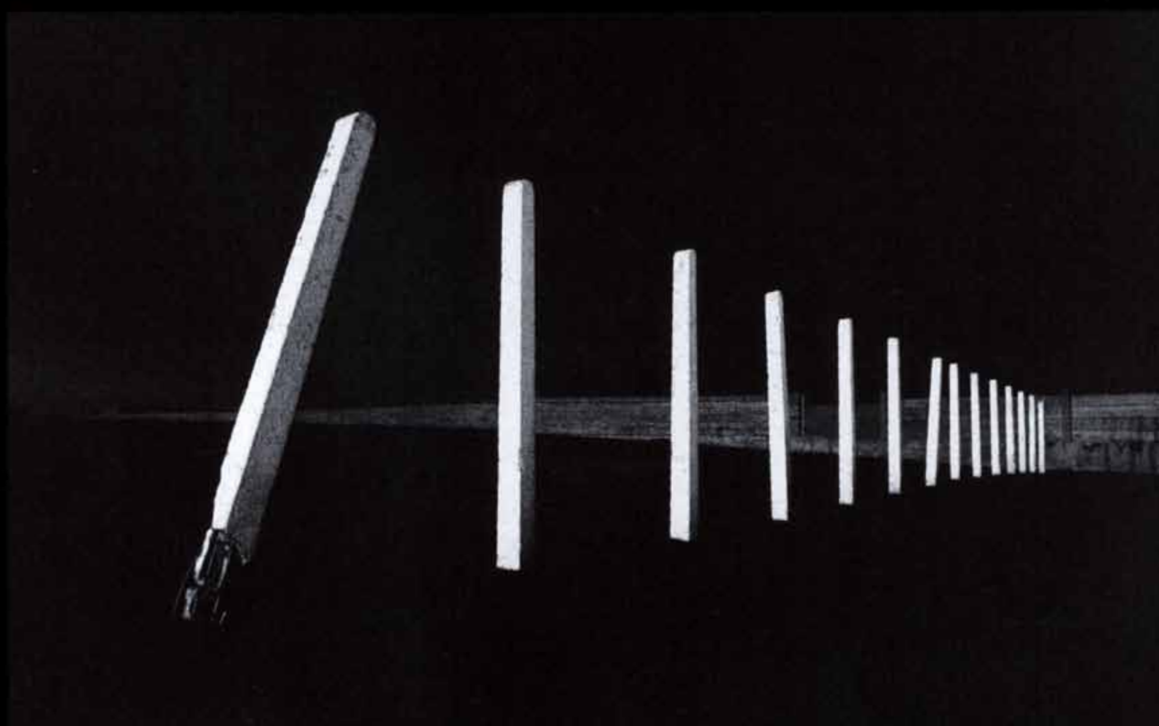
The long view is all.

At Toshiba, we've been taking that view to help create a tangible future for over one hundred years.

Sustaining it now with our commitment to manufacturing in the U.S. And with a wide diversity of products of the highest quality.

All of which have helped make us one of the leading electrical and electronic manufacturers in the world.

A nice perspective to have. When you work in as many fields as we do.



Michael Kenna, *Tilted Poles*

Color TVs & VCRs • Computers & Printers • Copiers • CT Scanners • Digital Instrumentation • Disk Drives • Facsimile Systems • Home Appliances • Industrial Motors & Controls Liquid Crystal Displays • Medical Tubes • MR Imaging • Nuclear Medicine • OPTO Electronic Devices • Picture Tubes • Portable Audio • Professional Video Systems • Room Air Conditioners Semiconductors • Telecommunication Systems • Turbines & Generators • Ultrasound • Uninterruptible Power Supply • Variable Speed Drives • X-Ray Systems

In Touch with Tomorrow

TOSHIBA

so the incident laser light to be swept preferentially to one side of the crystal. If the incident beam and crystal are positioned so that the fanned beam is swept into a far corner of the crystal, the fanned beam undergoes two internal reflections, essentially folding back onto itself. This reflected beam fans again back along the direction of the incident beam. Out of all the scattered beams inside the crystal, the time-reversed beam—by virtue of its backward trajectory—gains more energy than the other scattered beams. This beam-reversal process can be very efficient: 60 percent of the power in the incident beam can emerge as the phase-conjugate beam.

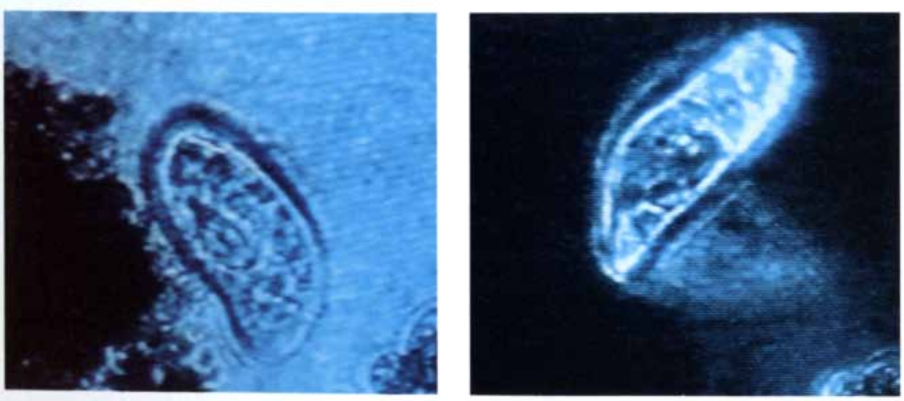
One of us (Pepper) has added an additional twist to this novel phase-conjugate mirror. Pepper attached electrodes to a photorefractive crystal to apply a time-varying electric field across the crystal. When a laser beam strikes this crystal, it not only is time-reversed but also can be pulsed in time like a shuttered mirror. In this manner, pulsed information can be relayed from the conjugate mirror back to the laser source; the time-reversed nature of the beam guarantees that the two communication points remain locked onto each other. This scheme can be used to establish a communications channel between two satellites or to relay information from a remote sensor placed at one end of an optical fiber link [*see illustration on page* 70].

Another application that takes advantage of the energy-exchange mechanism is a device called a novelty filter, which highlights whatever is changing in a highly complex scene. Such devices can pick out moving airplanes against a background of buildings, a sportswoman diving into a still pond or bacteria swimming against a background of motionless algae.

One type of novelty filter demonstrated by Cronin-Golomb (now at Tufts University) involves two beams that illuminate a photorefractive crystal. An image is encoded spatially onto the first laser beam. The crystal is oriented so that this image-bearing beam transfers most of its energy to the second beam. After the image-bearing beam passes through the crystal, it becomes almost completely dark. Whenever something in the image-bearing beam changes, however, the energy-exchange mechanism is disturbed momentarily, and the part of the image beam that has changed will pass through the crystal. This part of the beam can then be viewed on a video monitor. Once the motion has ceased, the image-bearing beam will become dark once again after passing through the crystal.

Photorefractive crystals can also be employed in other applications to enhance the edges of an image. An image is encoded onto an "object" beam, which is directed into a photorefractive crystal along with a "reference" beam. The two beams interfere inside the crystal and produce a hologram of the original image. The image can be recovered by a third "readout" beam, which is aimed in a direction opposite to the reference beam. If the object beam is relatively weak, then the reconstructed image will be a faithful replica of the original picture. If the object beam is more intense than the other two beams, however, then the edges of the reconstructed image will be enhanced.

The intensity of the object beam varies locally in the crystal because the beam contains an image. Hence, its intensity will match that of the reference beam at every edge in the image, because an edge contains a full range of intensities, from dark to bright. Wherever the intensities of the two beams match, the optical interference pattern at that particular region in the crystal will have the largest modulation from bright to dark to bright and so on. The strong interference pattern will then generate a strong refractive-index grating. When the readout beam interacts with the strong grating, it will be most efficiently deflected, or diffracted, at that location, and so the reconstructed image will emerge with all of its edges highlighted.



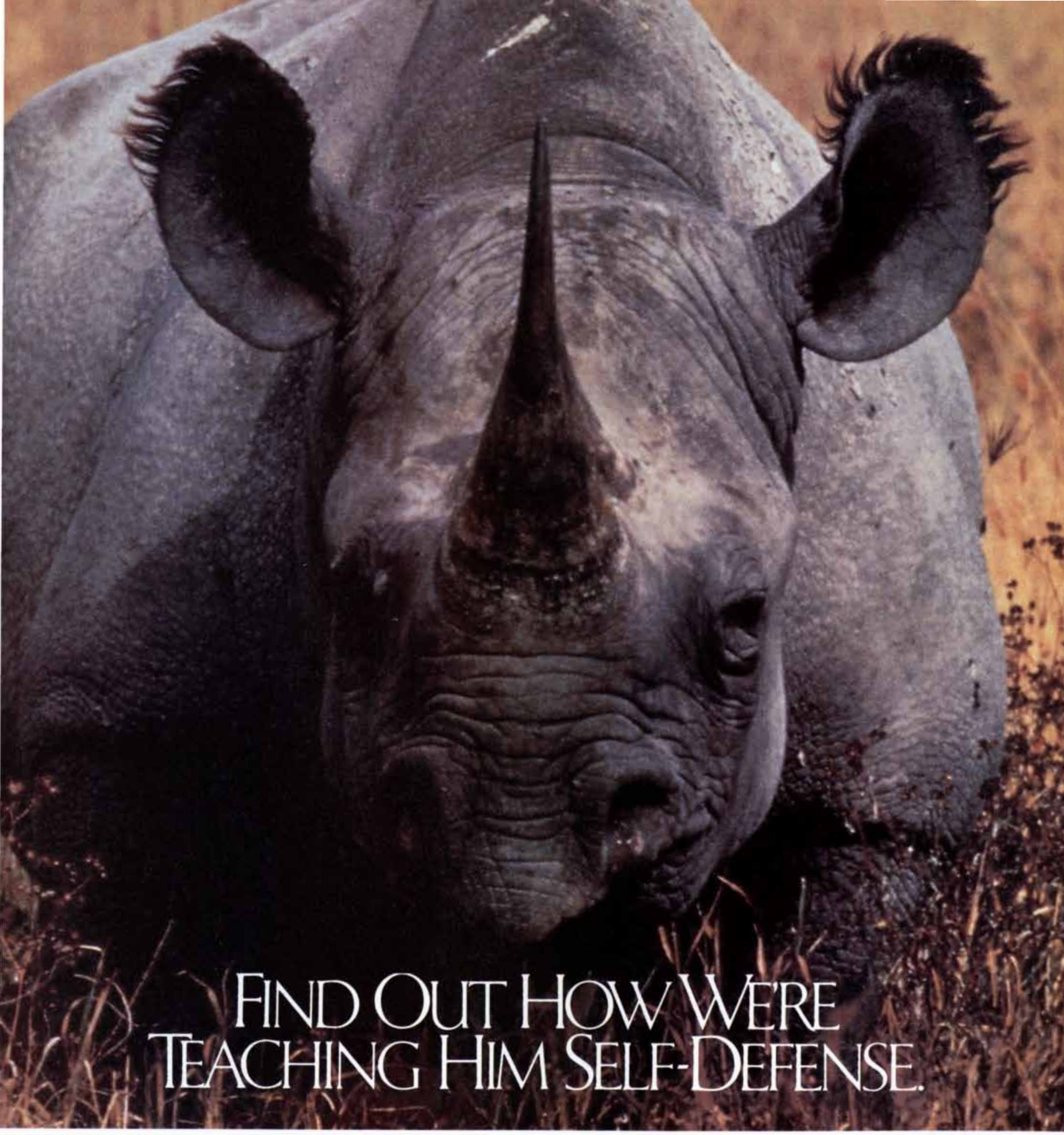
PARAMECIUM is hidden among algae and debris (*left*), but when the scene is viewed through a novelty filter, only the swimming microorganism appears (*right*). Novelty filters can highlight any object that moves against a stationary background. This filter was designed by R. M. Pierce, R. Cudney, G. D. Bacher and J. Feinberg.

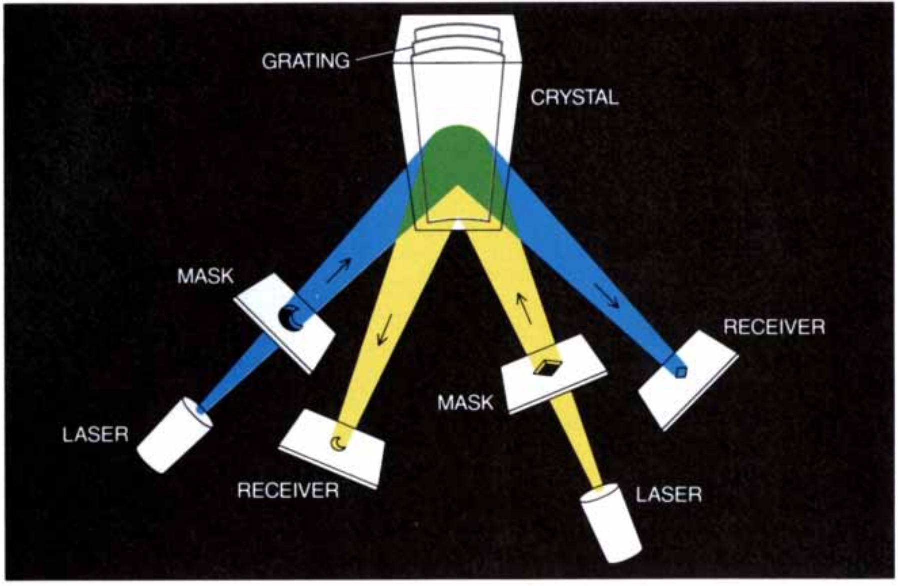
Because photorefractive crystals can act as both energy couplers and phase-conjugate mirrors, they are particularly useful for reconfigurable optical interconnects and frequency-locking of lasers. Photorefractive crystals can relay information from one optical element (say, an optical fiber) to another (a data-processing element), free of complicated optical elements or electronic interconnects. This optical relaying scheme can also force two (or more) separate lasers to "lock" onto each other so that the two lasers oscillate at precisely the same optical wavelength; in this way, the two separate lasers essentially behave as one larger laser.

How do two or more mutually incoherent beams of light become connected in a photorefractive crystal? When the different beams illuminate a photorefractive crystal, each will produce its own armada of scattered beams and random refractive-index gratings (or holograms) within the crystal. If one hologram from one beam exactly matches one of the holograms from the other beam, then that hologram will grow faster in strength than the others. The result is that the two beams will be connected to each other. This device is called a mutually pumped, or doubly pumped, phase-conjugate mirror.

The device—first demonstrated in 1987 by Fischer, now at the Technion–Israel Institute of Technology in Haifa, and his co-workers, by Robert W. Eason and A.M.C. Smout of the University of Essex in Great Britain and subsequently by Mark D. Ewbank of the Rockwell International Science Center in Thousand Oaks, Calif.—can connect any two beams originating from any direction. If the two beams contain images, then each image will be converted into the other as the beams traverse the crystal [*see illustration on page 74*]. If one beam is pulsed rapidly, the temporal information is relayed back toward the other beam.

If two beams come from different lasers that oscillate at slightly different frequencies, then the beam from one laser will be diverted into the approximate conjugate, or time-reversed, di-


FIND OUT HOW WE'RE
TEACHING HIM SELF-DEFENSE.



PHOTOREFRACTIVE CRYSTAL couples two laser beams, a process that may be exploited in communications systems, optical computers or other laser networks.

rection of the beam from the other laser, and vice versa. If two such lasers are optically connected by a crystal, they can, under the proper conditions, be made to oscillate coherently, thereby locking the frequency of the two lasers. With this scheme, workers hope that thousands of semiconductor lasers can be efficiently combined into a high-power source of coherent light.

To construct better optical components and devices, investigators are searching among the many different kinds of photorefractive materials to find the most efficient and reliable crystals. Photorefractive materials vary greatly in their optical, electrical and structural properties—for example, the insulator barium titanate, the semiconductor gallium arsenide and the organic compound 2-cyclooctylamino-5-nitropyridine doped with 7,7,8,8,-tetracyanoquinodimethane.

These seemingly different materials nonetheless exhibit similar photorefractive effects. They all have a crystal lattice that is relatively easy to distort, and they all contain defects, which act as a source of charge carriers and charge traps. Under the influence of an electric field, however, charges move about 10,000 times faster in gallium arsenide than they do in barium titanate. For the same incident light intensity, therefore, the photorefractive effect evolves much more rapidly in gallium arsenide than it does in barium titanate. The resulting refractive-index grating is not as strong in gallium arsenide, however, as it is in barium titanate. The optical properties of the crystals in the different materials vary greatly as well: gallium arsenide and other semiconductors are typically photosensitive in the near infrared part of the optical spectrum, whereas most insulators and organic compounds are sensitive in the visible spectrum.

A flurry of experiments currently in progress aim to identify and possibly control the defects responsible for the photorefractive effect in various crystals. As an example, the source of the photorefractive effect in barium titanate is an open question: many researchers attribute the effect to the presence of various ionization states of transition metal impurities such as iron, cobalt and manganese; oxygen vacancies in the lattice may also contribute to the charge carriers. In gallium arsenide, on the other hand, the photorefractive effect is attributed to an intrinsic lattice defect thought to be formed by a combination of an arsenic atom replacing a gallium atom in the lattice and an additional arsenic atom placed in the same lattice cell, forming the so-called EL2 center.

Regardless of the character of the crystalline defects, the properties of most photorefractive crystals can be altered by doping them with impurities or by drawing atoms out of the lattice. For example, one of us (Feinberg) and Stephen Ducharme of U.S.C. found that when a crystal of barium titanate is heated in an oxygen-free environment to remove some oxygen from the crystal lattice, its photorefractive properties are altered markedly because the dominant charge carriers are changed from holes to electrons.

As workers increase the concentration of the defects in a photorefractive material, they have found that the number of available charges increases, thereby enhancing the strength of the internal electric field and the refractive-index gratings. On the other hand, defects scatter and absorb light from the incident beams. Defect concentrations of from one to 100 parts per million appear to be optimal in terms of providing a reasonable number of charges without appreciably attenuating the input light.

Perhaps the best photorefractive materials will be those painstakingly fashioned out of semiconductor materials. Stephen E. Ralph, David D. Nolte and Alastair M. Glass of AT&T Bell Laboratories have recently shown that layers of gallium aluminum arsenide alloys only a few atoms thick can be assembled into structures—called superlattices and quantum wells—that exhibit a measurable photorefractive effect. Such materials are currently being studied for their novel electrical properties and high speed of response. These crystals may also lead to a new class of integrated processors based on optics and electronics.

What makes photorefractive materials so promising for optical computing is their high sensitivity to light, coupled with their ability to connect light beams from different lasers and to change one information pattern into another. The future of photorefractive materials will depend on whether their optical properties can be as meticulously tailored as semiconductors are today. Ideally, photorefractive devices would be integrated with semiconductor lasers and detectors to form a single, compact device capable of processing millions of bits of data simultaneously in each microsecond, yielding a total data-processing rate of trillions of bits per second.

FURTHER READING

OPTICAL PHASE CONJUGATION IN PHOTOREFRACTIVE MATERIALS. Jack Feinberg in *Optical Phase Conjugation.* Edited by R. A. Fisher. Academic Press, 1983.

PHOTOREFRACTIVE NONLINEAR OPTICS. Jack Feinberg in *Physics Today,* Vol. 41, No. 10, pages 46–52; October, 1988.

PHOTOREFRACTIVE MATERIALS AND THEIR APPLICATIONS I & II: TOPICS IN APPLIED PHYSICS. Volumes 61 and 62. Edited by P. Günter and J.-P. Huignard. Springer-Verlag, 1988.

NONLINEAR OPTICAL PHASE CONJUGATION. Special Issue. Guest edited by David M. Pepper. *IEEE Journal of Quantum Electronics,* Vol. 25, No. 3, pages 312–647; March, 1989.

What Caused the Mass Extinction?

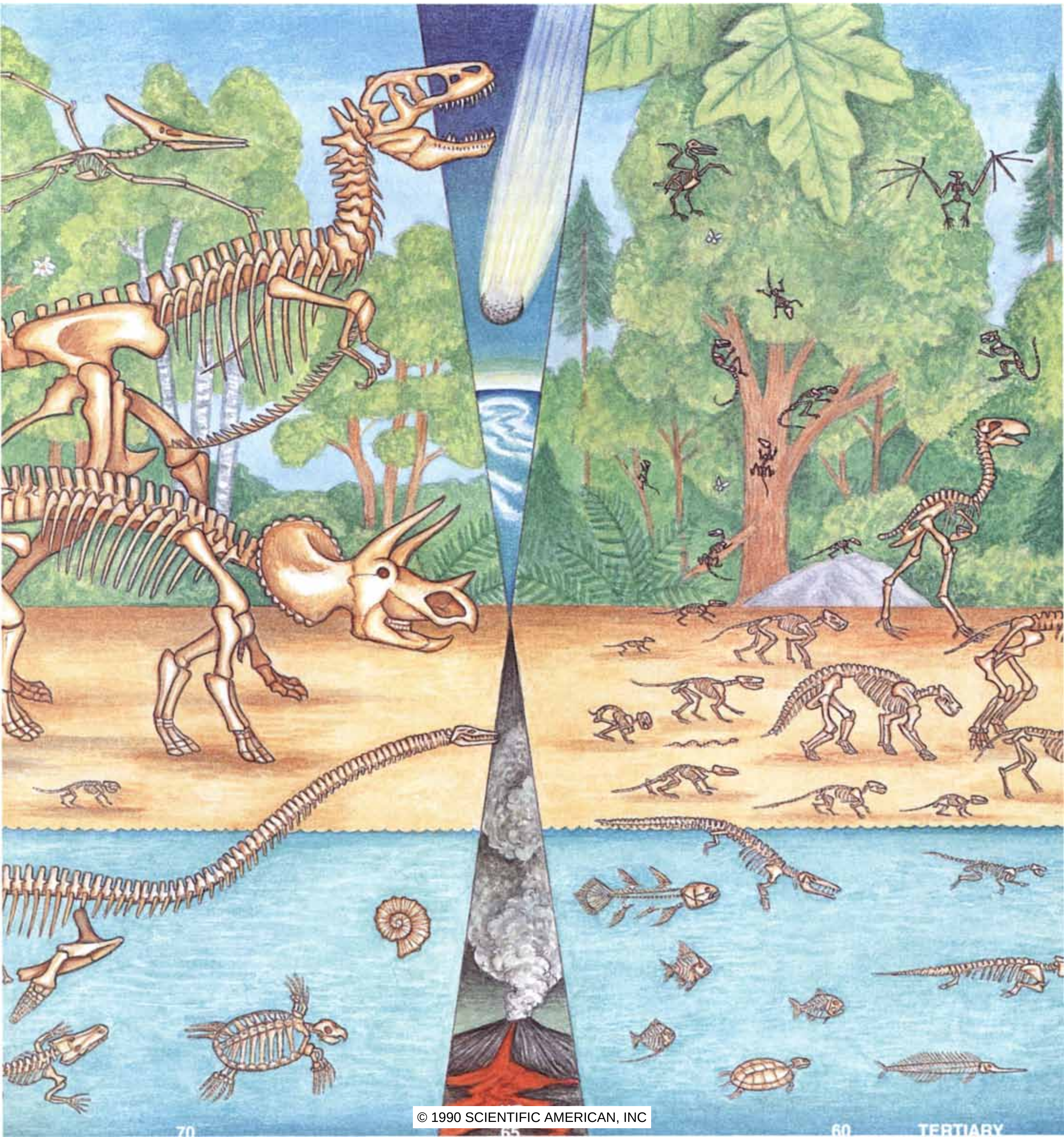
Sixty-five million years ago the world changed abruptly. The reptiles that had dominated the landscape vanished. So did more than half of all species of plants and land and marine animals. Because a species can survive even if its numbers are greatly reduced, the fact that so many species died out at the end of the Cretaceous period (the KT boundary) suggests that for a while the earth was an extremely inhospitable place.

Mammals somehow survived the hard times.

Spreading into ecological niches that the dinosaurs had vacated, they became the dominant kind of large animal. One species eventually began to investigate the fossil record of its distant origins.

That record is rich in clues of all kinds. In addition to dinosaur bones—relatively rare—there are traces of many other organisms, including abundant microscopic fossils of foraminifera and other marine life. The thickness and chemical composition of the strata also have a tale to tell.

The substance of the tale, however, seems to depend on who is reading it. Walter Alvarez and Frank Asaro of the University of California at Berkeley see traces of a huge impact from space hidden in layers of sedimentary rock in the Apennines of Italy and elsewhere around the globe. Vincent E. Courtillot of the Institute of Physics of the Earth in Paris sees evidence of a world wracked and polluted by the enormous volcanic eruptions that formed the Deccan Traps in India. On the following pages they present their arguments. The vanished saurians, meanwhile, keep their silence.



An Extraterrestrial Impact

Accumulating evidence suggests an asteroid or comet caused the Cretaceous extinction

by Walter Alvarez and Frank Asaro

About 65 million years ago something killed half of all the life on the earth. This sensational crime wiped out the dinosaurs, until then undisputed masters of the animal kingdom, and left the humble mammals to inherit their estate. Human beings, descended from those survivors, cannot avoid asking who or what committed the mass murder and what permitted our distant ancestors to survive.

For the past dozen years researchers from around the world, in disciplines ranging from paleontology to astrophysics, have mustered their observational skills, experimental ingenuity and theoretical imagination in an effort to answer these questions. Those of us involved in it have lived through long months of painstaking measurement, periods of bewilderment, flashes of insight and episodes of great excitement when parts of the puzzle finally fell into place.

We now believe that we have solved the mystery. Some 65 million years ago a giant asteroid or comet plunged out of the sky, striking the earth at a velocity of more than 10 kilometers per second. The enormous energy liberated by that impact touched off a nightmare of environmental disasters, including storms, tsunamis, cold and darkness, greenhouse warming, acid rains and global fires. When quiet returned at last, half the flora and fauna had become extinct. The history of the earth had taken a new and unexpected path.

Other suspects in the dinosaur murder mystery, such as sea level changes, climatic shifts and volcanic eruptions, have alibis that appear to rule them out. Some issues, however, are still unclear: Where was the impact site? Was it a single or multiple impact? Have such impacts occurred on a regular, periodic timetable? What is the role of such catastrophes in evolution?

The puzzle presented by a mass extinction is both like and unlike that of a more recent murder. There is evidence—chemical anomalies, mineral grains and isotopic ratios instead of blood or fingerprints or torn matchbooks—scattered throughout the world. No witnesses remain, however, and no chance exists of obtaining a confession. The passage of millions of years has destroyed or degraded most of the evidence in the case, leaving only the subtlest clues.

Indeed, it is difficult even to be sure which of the individual fossils that survive are those of victims killed by the impact. But paleontologists know there must have been victims because fossil-bearing sedimentary rocks show a great discontinuity 65 million years ago. Creatures such as dinosaurs and ammonites, abundant for tens of millions of years, suddenly disappeared forever. Many other groups of animals and plants were decimated.

This discontinuity defines the boundary between the Cretaceous period, during which dinosaurs reigned supreme, and the Tertiary, which saw the rise of the mammals. (It is known as the KT boundary after *Kreide*, the German word for "Cretaceous.")

When we began to study the KT boundary, we wanted to find out just how long the extinction had taken to occur. Was it sudden—a few years or centuries—or was it a gradual event that took place over millions of years? Most geologists and paleontologists had always assumed that the extinction had been slow. (These fields have a long tradition of gradualism and are uncomfortable with invoking catastrophes.) Because dinosaur fossils are relatively rare, their age provides little detailed information on the duration of the extinction. It was possible to view the extinction of dinosaurs as gradual.

When paleontologists looked at the fossils of pollen or single-celled marine animals called foraminifera, however, they found the extinction to be very abrupt. In general, smaller organisms produce more abundant fossils and so yield a sharper temporal picture.

The extinction also appears more sudden as paleontologists study closely the fossil record for medium-size animals such as marine invertebrates. Among these are the ammonites (relatives of the modern chambered nautilus), which died out at the end of the Cretaceous period. The best record of their extinction is found in the coastal outcrops of the Bay of Biscay on the border between Spain and France.

In 1986 Peter L. Ward and his colleagues at the University of Washington made detailed studies of these outcrops at Zumaya in Spain. Ward found that the ammonites appeared to die out gradually—one species disappearing after another over an interval of about 170 meters, representing about five million years. But in 1988 Ward studied two nearby sections in France and found evidence that these ammonite species actually survived right up to the KT boundary. The apparent gradual extinction at Zumaya was merely the artifact of an incomplete fossil record. If organisms whose fossils are well preserved died out abruptly, then it is likely that others that perished about the same time, such as dinosaurs, whose remains are more sparsely preserved, did so as well.

This establishes that the extinction was abrupt in geologic terms, but it does not establish how many years this extinction took, because it is a major accomplishment to date a rock to an accuracy of a million years. Intervals in the geologic records can be determined with precision only to within 10,000 years (.01 Myr), a period longer than the entire span of human civilization.

The duration of the mass extinction that marks the KT boundary can be estimated more precisely than this. In the deep-water limestones at Gubbio in Italy, a thin layer of clay separates Cretaceous and Tertiary sediments. The

WALTER ALVAREZ and FRANK ASARO are longtime collaborators on the impact hypothesis of mass extinction. Alvarez is a geologist at the University of California, Berkeley; his laboratory is the Apennine mountains of Italy, where he is a honorary citizen of the town of Piobbico. Asaro is a nuclear chemist at the Lawrence Berkeley Laboratory. Before his work with Alvarez, he made chemical analyses of archaeological materials; he determined that one of Berkeley's prized possessions, an inscribed brass plate attributed to Sir Francis Drake, was not authentic.

layer, discovered by Isabella Premoli Silva of the University of Milan, is typically about one centimeter thick. In the 1970s one of us (Alvarez) was part of a group that found the clay falls within a six-meter thickness of limestone deposited during the .5-Myr period of reversed geomagnetic polarity designated 29R. On the face of it, this suggests that the clay layer, and the mass extinction it marks, represents a span of no more than .001 Myr, about 1,000 years.

Jan Smit of the University of Amsterdam did a similar study of sediments at Caravaca in southern Spain, where the stratigraphic record is even more precise, and estimated the extinction lasted no more than 50 years. By geologic standards this is blindingly fast!

Our work on the KT boundary began in the late 1970s when we and our Berkeley colleagues Luis W. Alvarez and Helen V. Michel tried to develop a more accurate way to determine how long the Gubbio KT clay layer took to be deposited. Our efforts failed, but they did provide a crucial first clue to the identity of the mass killer. (That is what detectives and scientists need: a lot of hard work and an occasional lucky break.)

The method depended on the rarity of iridium in the earth's crust—about .03 part per billion as compared with 500 parts per billion, for example, in the primitive stony meteorites known as carbonaceous chondrites. Iridium is rare in the earth's crust because most of the planet's allotment is alloyed with iron in the core.

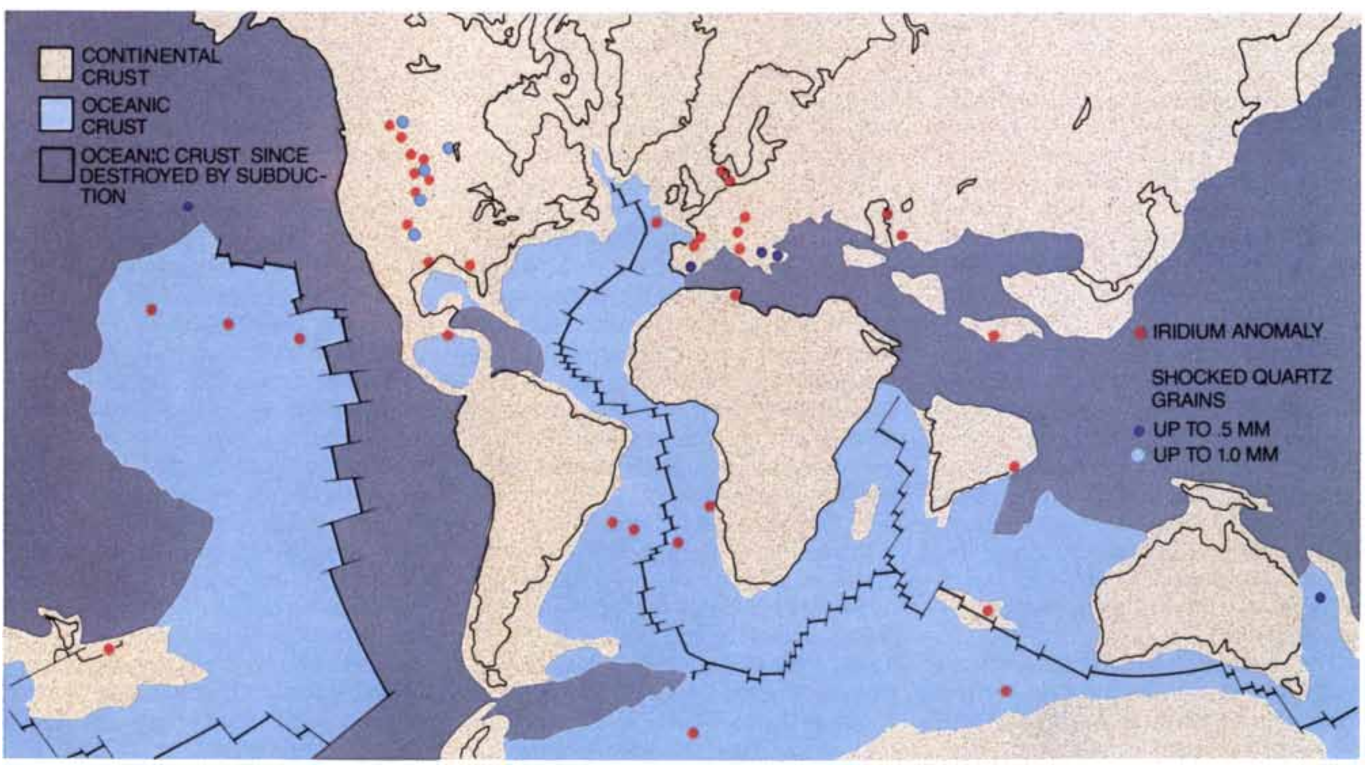
We suspected that iridium would enter deep-sea sediments, such as those at Gubbio, predominantly through the continual rain of micrometeorites, sometimes called cosmic dust. This constant infall would provide a clock: the more iridium in a sedimentary layer, the longer it must have taken to lay down. Moreover, iridium could be measured at very low concentrations by means of neutron-activation analysis, a technique in which neutron bombardment converts the metal into a radioactive and hence detectable form.

One scenario we considered was that the KT boundary clay layer formed over a period of about 10,000 years when organisms that secrete calcareous shells died out, and so no calcium carbonate (which makes up most of the limestone) was deposited. Most layers at Gubbio contain about 95 percent calcium carbonate and 5 percent clay; the boundary layer contains 50 percent clay. If this scenario was correct, the ratio of iridium to clay would be the same in the boundary clay as in higher and lower layers. If clay deposition had slowed at the same time as calcium carbonate deposition, the ratio would be higher than that in adjacent rocks.

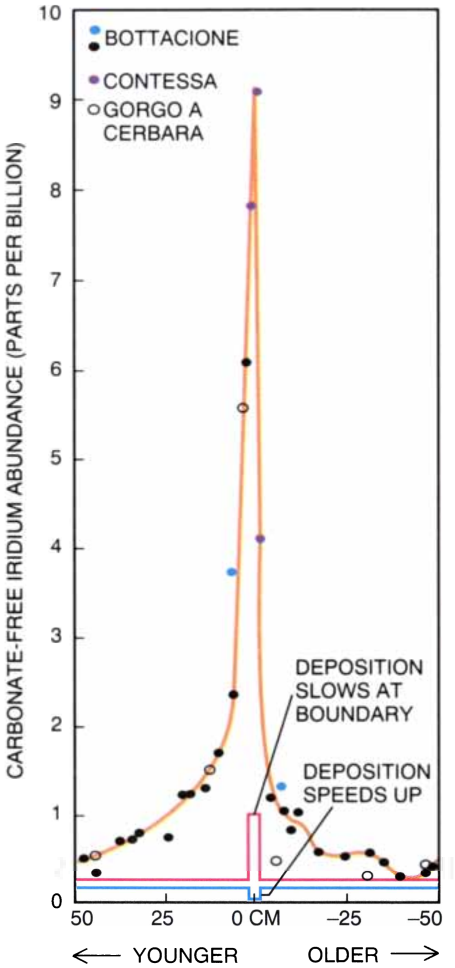
In June of 1978 our first Gubbio iridium analyses were ready. Imagine our astonishment and confusion when we saw that the boundary clay and the immediately adjacent limestone contained far more iridium than any of our scenarios predicted—an amount comparable to that in all the rest of the rock deposited during the 500,000 years of interval 29R.

Clearly, this concentration could not have come from the usual sprinkling of cosmic dust. For a year we debated possible sources, testing and rejecting one idea after another. Then in 1979 we proposed the one solution that had survived our testing: a large comet or asteroid about 10 kilometers in diameter had struck the earth and dumped an enormous quantity of iridium into the atmosphere.

Since we first proposed the impact hypothesis, so much confirming evidence has come to light that most scientists working in the field are persuaded that a great impact occurred. More than 100 scientists in 21 laborato-



EVIDENCE OF GIANT IMPACT that caused the mass extinction at the end of the Cretaceous period is scattered around the globe. This map shows continents as they were at the time of the extinction 65 million years ago. Anomalous levels of iridium, which suggest a large impact, have been found in sediments from the Cretaceous-Tertiary (KT) boundary, as have impact-generated quartz crystals. The 150-kilometer crater that an impact would have made has not been found.



IRIDIUM ANOMALY in limestone sediments is a clear sign of extraterrestrial impact. Iridium is rare in the earth's crust; only an extraterrestrial body could have deposited the amount found in the KT boundary. Other extinction scenarios have been suggested (*colored lines*) in which the rate of limestone deposition varied while cosmic iridium infall remained constant, but they cannot explain such high levels of iridium.

ries in 13 countries have found anomalously high levels of iridium at the KT boundary at about 95 sites throughout the world. The anomaly has been found in marine and nonmarine sediments, at outcrops on land and in oceanic sediment cores. Further, we have analyzed enough other sediments to know that iridium anomalies are very rare. As far as we know, the one at the KT boundary is unique.

The iridium anomaly is well explained by impact because the ratio of iridium to elements with similar chemical behavior, such as platinum, osmium, ruthenium, rhodium and gold, is the same in the boundary layer as it is in meteorites. Miriam Kastner of the Scripps Institution of Oceanography, working with our group, has determined that the gold-iridium ratio in the carefully studied KT boundary at Stevns Klint in Denmark agrees to within 5 percent with the ratio in the most primitive meteorites (type I carbonaceous chondrites).

Indeed, the ratios of all the platinum-group elements found in the KT boundary give evidence of extraterrestrial origin. George Bekov of the Institute of Spectroscopy in Moscow and one of us (Asaro) have found that the relative abundances of ruthenium, rhodium and iridium can distinguish stony meteorites from terrestrial samples. Analysis of KT boundary samples from Stevns Klint, Turkmenia in the Soviet Union and elsewhere support the impact hypothesis.

So do ratios of isotopes. Jean-Marc Luck, then at the Institute of Physics of the Earth in Paris, and Karl K. Turekian of Yale University found that most of the osmium in KT boundary samples from Denmark and New Mexico could not have come from a continental source, because the abundance of osmium 187 is too low. The ratio of osmium 187 to osmium 186 is higher in continental rocks than in meteorites or in the earth's mantle because those rocks are relatively enriched in rhenium, whose radioactive isotope, rhenium 187, decays to osmium 187. The osmium in KT samples must be extraterrestrial or from the earth's mantle.

Not only does the composition of rocks at the KT boundary suggest impact, but so does their mineralogy. In 1981 Smit discovered another telltale clue: mineral spherules as large as a millimeter in diameter in the Caravaca KT clay. (Alessandro Montanari of Berkeley confirmed their presence in Italy as well.) The spherules originated as droplets of basaltic rock, shock-melted by impact and rapidly cooled during ballistic flight outside the atmosphere, then chemically altered in the boundary clay. They are the basaltic equivalent of the more silica-rich glassy tektites and microtektites that are the known result of smaller impacts. The basaltic chemistry suggests that the impact took place on oceanic crust.

In addition to the spherules, shocked grains of quartz have been discovered by Bruce F. Bohor of the U.S. Geological Survey in Denver and Donald M. Triplehorn of the University of Alaska. Painstaking studies by E. E. Foord, Peter J. Modreski and Glen A. Izett of the USGS show that the grains carry the multiple intersecting planar "lamellae"—bands of deformation—symptomatic of hypervelocity shock. Such grains are found only in known impact craters, at nuclear test sites, in materials subjected to extreme shock in the laboratory—and in the KT boundary.

There is in fact a candidate crater beneath the glacial drift at Manson, Iowa; it lies in a quartz-rich bedrock, and its location is suitable to explain the size and abundance distribution of the shocked quartz grains. At 32 kilometers in diameter, the crater is too small to have been formed by the single body posited as having caused the extinction. Nevertheless, detailed studies of the crater show it to have an age indistinguishable from that of the KT boundary, and so it probably played a part in the mystery.

How would an impact disperse shocked and molten materials around the globe? A 10-kilometer asteroid moving at more than 10 kilometers per second would ram a huge hole in the atmosphere. When it hit the ground, its kinetic energy would be converted to heat in a nonnuclear explosion 10,000 times as strong as the total world arsenal of nuclear weapons. Some vaporized remains of the asteroid and rock from the ground near the impact point would then be ejected through the hole before the air had time to rush back in.

The fireball of incandescent gas created by the explosion would also propel material out of the atmosphere. The fireball of an atmospheric nuclear explosion expands until it reaches the same pressure as the surrounding atmosphere, then rises to an altitude where its density matches that of the surrounding air. At that point, usually around 10 kilometers high, the gas spreads laterally to form the head of the familiar mushroom cloud.

Computer models of explosions with energies of 1,000 megatons—about 20 times the energy of the largest nuclear bombs but only 1/100,000 the energy of the KT impact—have shown that the fireball never reaches pressure equilibrium with the surrounding atmosphere. Instead, as the fireball expands to altitudes where the density of the atmosphere declines significantly, its rise accelerates and the gas leaves the atmosphere at velocities fast enough to escape the earth's gravitational field. The fireball from an even greater asteroid impact would simply burst out the top of the atmosphere, carrying any entrained ejecta with it, sending the material into orbits that could carry it anywhere on the earth.

The impact of a comet-size body on the earth, creating a crater 150 kilometers in diameter, would clearly kill everything within sight of the fireball. Researchers are refining their understanding of the means by which an impact would also trigger extinction worldwide. Mechanisms proposed include darkness, cold, fire, acid rain and greenhouse heat.

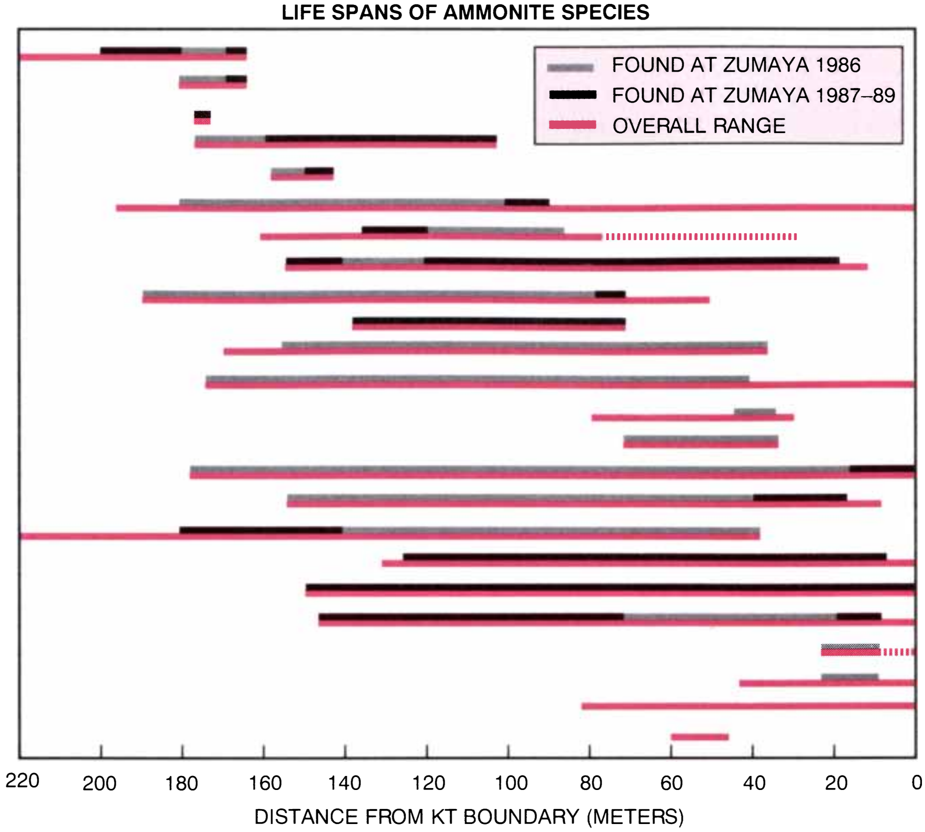
In our original paper, we proposed that impact-generated dust caused global darkness that resulted in extinctions. According to computer simulations made in 1980 by Richard P. Turco of R&D Associates, O. Brian Toon of the National Aeronautics and Space Administration and their colleagues, dust lofted into the atmosphere by the impact of a 10-kilometer object would block so much light that for months you would literally be unable to see your hand in front of your face.

Without sunlight, plant photosynthesis would stop. Food chains everywhere would collapse. The darkness would also produce extremely cold temperatures, a condition termed impact winter. (After considering the effects of the impact, Turco, Toon and their colleagues went on to study nuclear winter, a related phenomenon as capable of producing mass extinctions today as impact winter was 65 million years ago.)

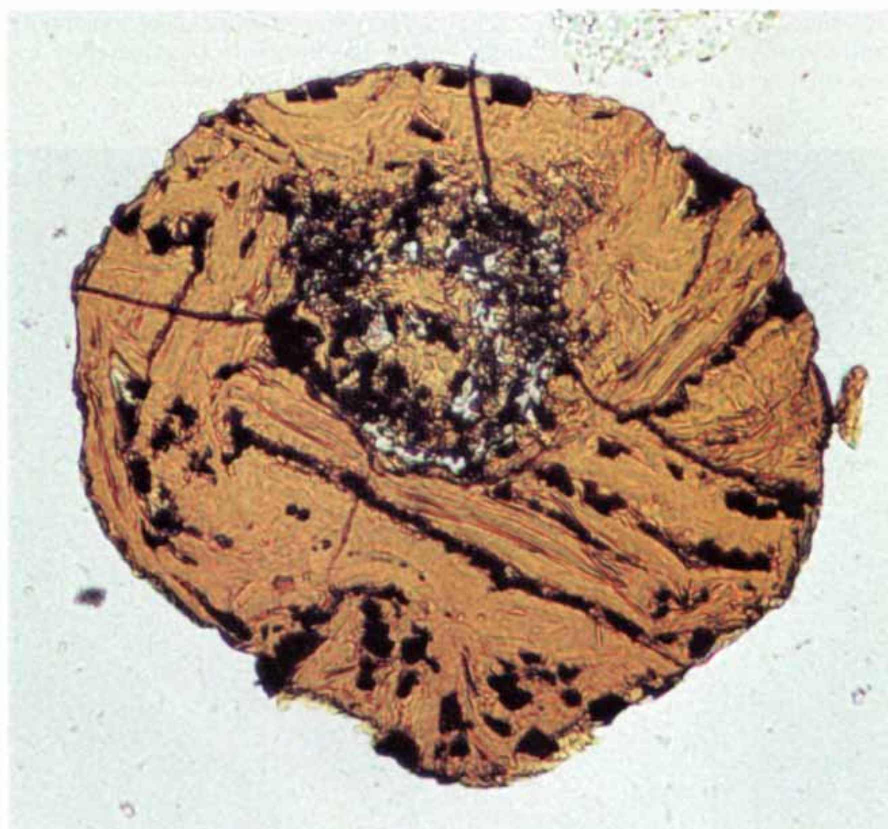
In 1981 Cesare Emillíani of the University of Miami, Eric Krause of the University of Colorado and Eugene M. Shoemaker of the USGS pointed out that an oceanic impact would loft not only rock dust but also water vapor into the atmosphere. The vapor, trapping the earth's heat, would stay aloft much longer than the dust, and so the impact winter would be followed by greenhouse warming. More recently John D. O'Keefe and Thomas J. Ahrens of the California Institute of Technology have suggested that the impact might have occurred in a limestone area, releasing large volumes of carbon dioxide, another greenhouse gas. Many plants and animals that survived the extreme cold of impact winter could well have been killed by a subsequent period of extreme heat.

Meanwhile John S. Lewis, G. Hampton Watkins, Hyman Hartman and Ronald G. Prinn of the Massachusetts Institute of Technology have calculated that shock heating of the atmosphere during impact would raise temperatures high enough for the oxygen and nitrogen in the air to combine. The resulting nitrous oxide would eventually rain out of the air as nitric acid—an acid rain with a vengeance. This mechanism may well explain the widespread extinction of marine invertebrate plants and animals, whose calcium carbonate shells are soluble in acidic water.

Another killing mechanism came to



MARINE FOSSIL RECORD testifies to the suddenness of the Cretaceous extinction. Although examination of some rocks suggests a gradual extinction, scrutiny of the complete fossil record shows that virtually all species died out simultaneously.



BASALTIC SPHERULES have been found embedded in clays at the KT boundary at several locations. Now chemically altered, the spherules originated as molten droplets of ocean subfloor that were deposited globally by the KT impact.

light when Wendy Wolbach, Ian Gilmore and Edward Anders of the University of Chicago discovered large amounts of soot in the KT boundary clay. If the clay had been laid down in a few years or less, the amount of soot in the boundary would indicate a sudden burning of vegetation equivalent to half of the world's current forests. Jay Meos of the University of Arizona and his colleagues have calculated that infrared radiation from ejecta heated to incandescence while reentering the atmosphere could have ignited fires around the globe.

Detailed studies of the KT boundary sediments may eventually provide evidence supporting a particular killing mechanism. For example, dissolution patterns in the Italian limestone show that bottom waters were acidic immediately after the extinction. And work we have done with William Lowrie of ETH-Zurich shows that those waters also changed briefly from their normal oxidizing state to a reducing condition, possibly because of the massive death of marine organisms.

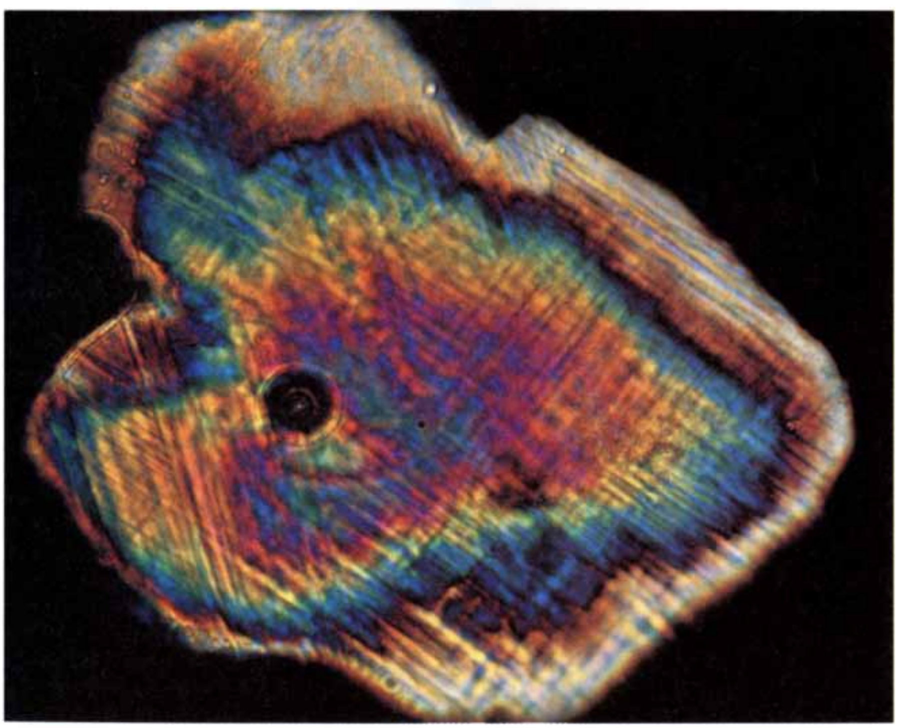
It has always been a major disappointment that no one has found the 150-kilometer crater a 10-kilometer impacting object should have produced. The crater might be hidden under the Antarctic ice sheet, or it might have been on the 20 percent of the earth's surface that has subsequently been consumed in subduction zones at the edges of oceanic plates. The evidence regarding the location is contradictory: the basaltic spherules in the boundary clay point to an impact on the ocean floor, but the shocked quartz grains argue for a continental hit.

A newly emerging point of view suggests, unlikely as it may seem, that the KT extinction may have been caused by two or more nearly simultaneous impacts. Shoemaker and Piet Hut of the Institute for Advanced Study in Princeton, N.J., have identified a number of mechanisms that could yield multiple impacts, either on the same day or over the course of many years. Double or multiple craters have been found on the earth, the moon and other planets, suggesting that some asteroids may consist of two or more objects mutually orbiting one another. Alternatively, the earth may have been struck by two or more large fragments of a comet nucleus in the process of breaking up.

Multiple impacts over longer periods could have occurred if a dispersing comet nucleus left several large fragments in an earth-crossing orbit. Such impacts could also occur randomly if some other factor increased the average number of comets in the inner solar system. Although not one scenario has won out, collectively they indicate that multiple impacts are not as improbable as might be thought.

The comet theory gains credibility from the discovery of apparently extraterrestrial materials near the KT boundary. Meixun Zhao and Jeffrey L. Bada of the University of California at San Diego analyzed chalk layers just above and below the KT boundary in Denmark. They found amino acids that are not used by life on the earth but do occur in carbonaceous chondrite meteorites. It seems unlikely that amino acids could survive the heat of a large impact, and they in fact do not appear in the KT boundary itself.

Kevin Zahnle and David Grinspoon of NASA have proposed that dust from a disintegrating comet entered the earth's atmosphere over an extended period and carried these extraterrestrial amino acids with it. During that interval the impact of a large fragment of the comet would have caused the KT extinction.



QUARTZ GRAIN shows laminar deformations generated by hypervelocity shock. Such quartz is found in laboratories, at nuclear test sites and near impact craters.

An apparently unrelated line of inquiry, based on statistical rather than chemical analyses, has yielded a hypothesis explaining how comets could hit the earth periodically. In 1984 David M. Raup and John J. Sepkoski, Jr., of the University of Chicago published an analysis of the fossil record, which seemed to indicate that mass extinctions have occurred at 32-million-year intervals. Like most scientists working with the KT boundary, we were very skeptical of their results. But astrophysicist Richard A. Muller of the University of California at Berkeley reexamined Raup and Sepkoski's data and convinced himself that the periodicity was real.

Muller, Marc Davis of Berkeley and Hut hypothesized that a dim, unrecognized companion star orbiting the sun every 32 million years (which they provisionally dubbed Nemesis) might regularly disturb the orbits of comets on the outer fringe of the solar system. The disturbance would send a million-year storm of comets into the inner solar system, greatly increasing the chance of a large impact (or multiple impacts) on the earth. Daniel Whitmire of the University of Southwestern Louisiana and Albert Jackson of Computer Sciences Corporation independently proposed the same hypothesis.

When Muller showed one of us (Alvarez) the paper proposing Nemesis, I was very skeptical. I remember telling him that I thought it was "an ingenious solution to a nonproblem" because I was not convinced of Raup and Sepkoski's evidence for periodic mass extinctions. If the hypothesis was correct, I pointed out, terrestrial impact craters should show the same periodicity in their ages. Muller and I found, to his delight and to my surprise, that crater ages do show essentially the same peri-

TWIN CRATERS of Clearwater Lakes in Canada show that multiple impacts on the earth are not as unlikely as they might seem. Evidence suggests that at least two impacts within a short time were responsible for the Cretaceous extinction.

odicity as mass extinctions. Since then I have felt that the hypothesis must at least be taken seriously.

It turns out, however, that it is very difficult to find a dim red star close to the sun when one has no idea where to look. Muller and Saul Perlmutter of Berkeley are now about halfway through a computerized telescopic search for a star with the characteristics of Nemesis; they expect to finish in a couple of years. Meanwhile new analysis of crater ages and extinction dates has raised questions about whether they actually are periodic. The small numbers of events and the sketchy information available make the question difficult to answer unequivocally.

Murder suspects typically must have means, motive and opportunity. An impact certainly had the means to cause the Cretaceous extinction, and the evidence that an impact occurred at exactly the right time points to opportunity. The impact hypothesis provides, if not motive, then at least a mechanism behind the crime. How do other suspects in the killing of the dinosaurs fare?

Some have an air-tight alibi: they could not have killed all the different organisms that died at the KT boundary. The venerable notion that mammals ate the dinosaurs' eggs, for example, does not explain the simultaneous extinction of marine foraminifera and ammonites.

Stefan Gartner of Texas A&M University once suggested that marine life was killed by a sudden huge flood of fresh water from the Arctic Ocean, which apparently was isolated from other oceans during the late Cretaceous and filled with fresh water. Yet this ingenious mechanism cannot account for the extinction of the dinosaurs or the loss of many species of land plants.

Other suspects might have had the ability to kill, but they have alibis based on timing. Some scientific detectives have tried to pin the blame for mass extinction on changes in climate or sea level, for example. Such changes, however, take much longer to occur than did the extinction; moreover, they do not seem to have coincided with the extinction, and they have occurred repeatedly throughout the earth's history without accompanying extinctions.

Others consider volcanism a prime suspect. The strongest evidence implicating volcanoes is the Deccan Traps, an enormous outpouring of basaltic lava in India that occurred approximately 65 million years ago. Recent paleomagnetic work by Vincent E. Courtillot [see "A Volcanic Eruption," on page 85] and his colleagues in Paris confirms previous studies. They show that most of the Deccan Traps erupted during a single period of reversed geomagnetic polarity, with slight overlaps into the preceding and succeeding periods of normal polarity. The Paris team has found that the interval in question is probably 29R, during which the KT extinction occurred, although it might be the reversed-polarity interval immediately before or after 29R as well.

Because the outpouring of the Deccan Traps began in one normal interval and ended in the next, the eruptions that gave rise to them must have taken place over at least .5 Myr. Most workers interested in mass extinction therefore have not considered volcanism a serious suspect in a killing that evidently took place over .001 Myr or less.

Some researchers have argued that, contrary to the fossil record, the KT extinctions took place over many thousands of years and that volcanism can account for quartz grains, spherules and the iridium anomaly.

In 1983 William H. Zoller and his colleagues at the University of Maryland at College Park discovered high concentrations of iridium in aerosols from Kilauea volcano in Hawaii collected on filters 50 kilometers away; however, the ratio between iridium and other rare elements in the volcanic aerosols does not match the ratio found at the KT boundary. The ratio of gold to iridium in the Kilauea aerosols is more than 35 times that in the KT boundary at Stevns Klint.

There has also been debate as to whether an explosive volcanic eruption might produce shocked quartz. It now seems agreed, however, that volcanic explosions can produce some deformation but that the distinctive multiple lamellae seen in the KT boundary quartz can only be formed by impact shocks. In addition, John McHone of Arizona State University has found that they contain stishovite, a form of quartz produced only at pressures far greater than those of volcanic eruptions. And Mark H. Anders of Columbia University and Michael R. Owen of St. Lawrence University have used a technique known as cathode luminescence, in which an electric field causes quartz to glow, to determine the origin of the KT grains. The colors produced by the grains are not volcanic; they argue instead for impact on an ordinary sedimentary sandstone.

Moreover, basaltic spherules in the KT boundary argue against explosive volcanism in any case; spherules might be generated by quieter forms of volcanism, but then they could not be transported worldwide.

The apparent global distribution of the iridium anomaly, shocked quartz and basaltic spherules is strong evidence exonerating volcanism and pointing to impact. Eruptions take place at the bottom of the atmosphere; they send material into the high stratosphere at best. Spherules and quartz grains, if they came from an eruption, would quickly be slowed by atmospheric drag and fall to the ground.

Nevertheless, the enormous eruptions that created the Deccan Traps did occur during a period spanning the KT extinction. Further, they represent the greatest outpouring of lava on land in the past quarter of a billion years (although greater volumes flow continually out of mid-ocean ridges). No investigator can afford to ignore that kind of coincidence.

It seems possible that impact triggered the Deccan Traps volcanism. A few minutes after a large body hit the earth the initial crater would be 40 kilometers deep, and the release of pressure might cause the hot rock of the underlying mantle to melt. Authorities on the origin of volcanic provinces, however, find it very difficult to explain in detail how an impact could trigger large-scale basaltic volcanism.

In the past few years the debate between supporters of each scenario has become polarized: impact proponents have tended to ignore the Deccan Traps as irrelevant, while volcano backers have tried to explain away evidence for impact by suggesting that it is also compatible with volcanism. Our sense is that the argument is a Hegelian one, with an impact thesis and a volcanic antithesis in search of a synthesis whose outlines are as yet unclear.

Even in its present incompletely solved form, the mystery of the KT mass killing carries a number of lessons. The late 18th and early 19th centuries, when the study of the earth was first becoming a science, was a period marked by a long battle between catastrophists—who thought that sudden great events were crucial to the evolution of the planet—and uniformitarians—who explained all history in terms of gradual change.

Steven J. Gould of Harvard University has shown how the uniformitarians so thoroughly won this battle that generations of geology students have been taught catastrophism is unscientific. The universe, however, is a violent place, as astronomy has shown, and it is now becoming clear that the earth has also had its violent episodes.

Evidence that a giant impact was responsible for the extinctions at the end of the Cretaceous has finally rendered the catastrophic viewpoint respectable. Future geologists, with the intellectual freedom to think in both uniformitarian and catastrophic terms, have a better chance of truly understanding the processes and history of the planet than did their predecessors.

Catastrophes have an important role to play in evolutionary thinking as well. If a chance impact 65 million years ago wiped out half the life on the earth, then survival of the fittest is not the only factor that drives evolution. Species must not only be well adapted, they must also be lucky.

If chance disaster occasionally wipes out whole arrays of well-adapted organisms, then the history of life is not preordained. There is no inevitable progress leading inexorably to intelligent life—to human beings. Indeed, Norman Sleep of Stanford University and his colleagues have suggested that in the very early history of the earth, when impacts were more frequent, incipient life may have been extinguished more than once.

Impact catastrophes may also prevent evolution from bogging down. The fossil record indicates that in normal times each species becomes increasingly well adapted to its particular ecological niche. Thus, it becomes ever more difficult for another species to evolve into that niche.

As a result, the rate of evolution slows. Wholesale removal of species by impact, however, provides a great opportunity for the survivors to evolve into newly vacant niches. (We have heard graduate students compare this situation with the excellent job prospects they would face if half of all tenured professors were suddenly fired.) Indeed, the fossil record shows that the rate of evolution accelerated immediately after the end of the Cretaceous.

Among the happy survivors of the KT extinction were the early Tertiary mammals, our ancestors. When dinosaurs dominated the earth, mammals seem always to have been small and insignificant. Warm-blooded metabolism, small size, large number or other traits may have suited them to endure the harsh conditions imposed by impact—or they may just have been lucky. And with the removal of the huge reptiles from the scene, mammals began an explosive phase of evolution that eventually produced human intelligence. As detectives attempting to unravel this 65-million-year-old mystery, we find ourselves pausing from time to time and reflecting that we owe our very existence as thinking beings to the impact that destroyed the dinosaurs.

FURTHER READING

EXTRATERRESTRIAL CAUSE FOR THE CRETACEOUS-TERTIARY EXTINCTION. Luis W. Alvarez, Walter Alvarez, Frank Asaro and Helen V. Michel in *Science*, Vol. 208, No. 4448, pages 1095–1108; June 6, 1980.

GEOLOGICAL IMPLICATIONS OF IMPACTS OF LARGE ASTEROIDS AND COMETS ON THE EARTH. Edited by Leon T. Silver and Peter H. Schultz. *Geological Society of America Special Paper*, Vol. 190, 1982.

TOWARD A THEORY OF IMPACT CRISES. W. Alvarez in *Eos*, Vol. 67, No. 35, pages 649–658; September 2, 1986.

NEMESIS AFFAIR: A STORY OF THE DEATH OF DINOSAURS AND THE WAYS OF SCIENCE. David M. Raup. W. W. Norton & Co., 1986.

NEMESIS. R. A. Muller. Weidenfeld and Nicolson, 1988.

GLOBAL CATASTROPHES IN EARTH HISTORY: AN INTERDISCIPLINARY CONFERENCE ON IMPACTS, VOLCANISM, AND MASS MORTALITY. Edited by V. L. Sharpton and P. D. Ward. *Geological Society of America Special Paper*, Vol. 247 (in press).

A Volcanic Eruption

What dramatic event 65 million years ago killed most species of life on the earth? The author argues it was a massive volcanic eruption

by Vincent E. Courtillot

The mysterious mass extinction that took place 65 million years ago has been attributed to either the impact of a large asteroid or a massive volcanic eruption. Both hypotheses presume that clouds of dust and chemical changes in the atmosphere and oceans created an ecological domino effect that eradicated large numbers of animal and plant families. The geologic record generally is consistent with either scenario; the central issue has been how rapid the event was. New evidence implies that the mass extinction occurred over tens or even hundreds of thousands of years. Such a duration closely corresponds to an episode of violent volcanic eruptions in India that occurred at the time of the mass extinction. Moreover, other extinction events also appear to be roughly simultaneous with periods of major volcanic activity.

The conventional divisions of geologic history reflect times of significant geologic and biological change. The mass extinction 65 million years ago defined the end of the Mesozoic era, when reptiles enjoyed great evolutionary success, and the beginning of the Cenozoic era, when mammals became extremely prevalent. Because the last period of the Mesozoic is the Cretaceous and the first period of the Cenozoic the Tertiary, the time of the most recent mass extinction is called the Cretaceous-Tertiary, or KT, boundary.

At this boundary the dinosaurs met their demise and, even more remarkable, 90 percent of all genera of protozoans and algae disappeared. John J. Sepkoski, Jr., and David M. Raup of the University of Chicago conclude that from 60 to 75 percent of all species vanished then. Equally important, many species, among them the ancestors of human beings, survived.

In 1980 Luis W. and Walter Alvarez (father and son) of the University of California at Berkeley, along with their colleagues Frank Asaro and Helen V. Michel, discovered unusually high concentrations of the metal iridium—from 10 to 100 times the normal levels—in rocks dating from the KT boundary in Italy, Denmark and New Zealand. Iridium is rare in the earth's crust but can be relatively abundant in other parts of the solar system. The Berkeley group therefore concluded that the iridium came from outer space, and thus the asteroid hypothesis was born.

A large asteroid impact would have cloaked the earth with a cloud of dust, resulting in darkness, suppression of photosynthesis, the collapse of food chains and, ultimately, mass extinction. The iridium is contained in a thin layer of clay whose chemical composition differs from that of the layers both above and below the boundary. Alvarez's group interpreted the clay as being the altered remains of the dust thrown up by the impact. In this view the boundary layer was laid down in less than one year, a flickering instant in geologic time. Other unusual findings at the KT boundary, most notably quartz crystals that appear to have been subjected to extremely powerful physical shocks, also could be explained by an asteroid impact.

An alternative to the asteroid hypothesis had already been brewing for some time. As early as 1972 Peter R. Vogt of the Naval Research Laboratory in Washington, D.C., pointed out that extensive volcanism had taken place at roughly the time of the KT boundary, principally in India. The volcanism produced extensive lava flows, known as the Deccan Traps (*deccan* means "southern" in Sanskrit, and *trap* means "staircase" in Dutch). Vogt suggested that the traps might be connected to the many changes that took place at the end of the Cretaceous period.

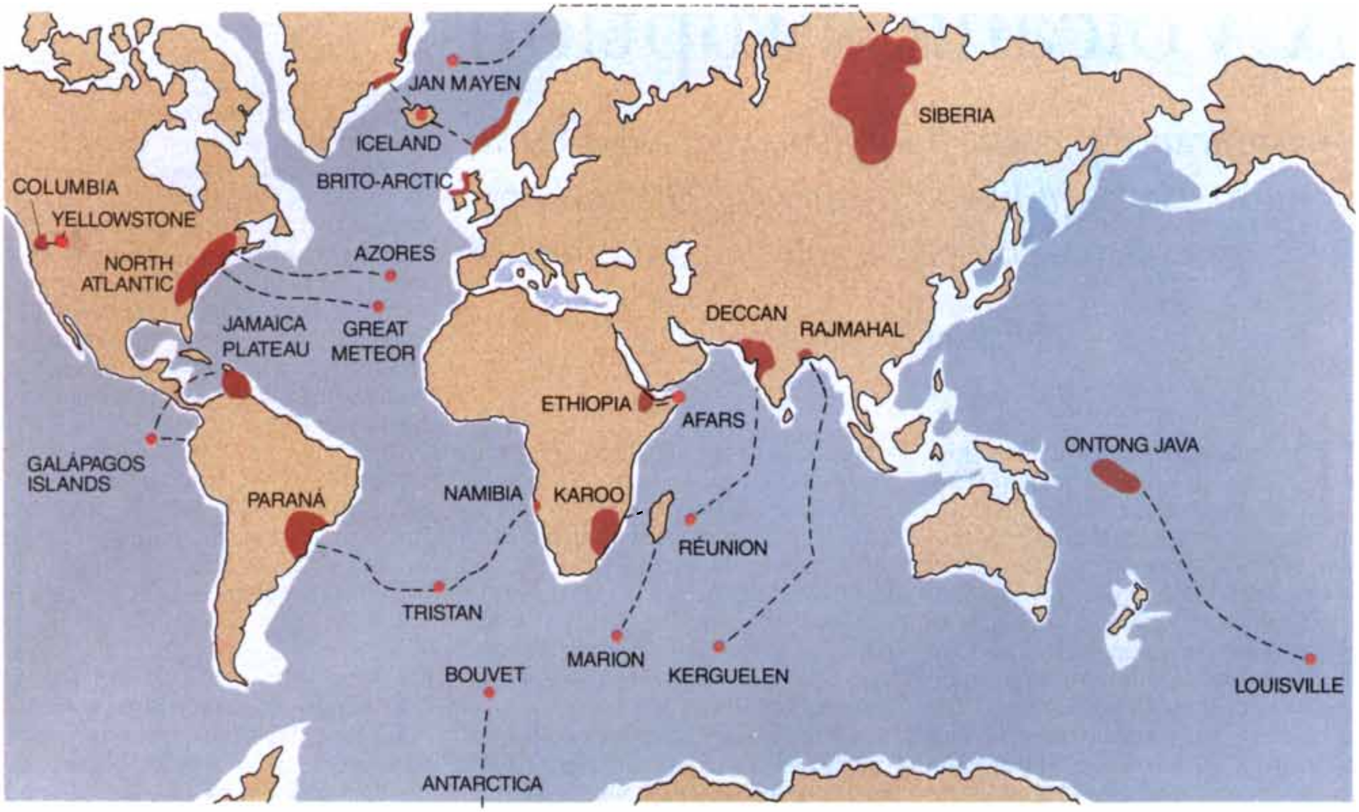
In the mid-1970s Dewey M. McLean of the Virginia Polytechnic Institute proposed that volcanoes could produce mass extinctions by injecting vast amounts of carbon dioxide into the atmosphere that would trigger abrupt climate changes and alter ocean chemistry. Charles B. Officer and Charles L. Drake of Dartmouth College analyzed sediments from KT boundary sections and concluded that the iridium enrichment and other chemical anomalies at the boundary were not deposited instantaneously but rather over a period of 10,000 to 100,000 years. They also argued that the anomalies were more consistent with a volcanic rather than meteoritic origin.

The amount of time represented by the clay layer at the KT boundary emerged as a major point of contention. Dating a 100-million-year-old rock with a precision of one part in 1,000 (that is, to within 100,000 years) is not yet possible. Yet much of the debate focuses on whether the boundary clay was deposited in less than one year (as would be expected from an impact) or in 10,000 (from an extended period of volcanism).

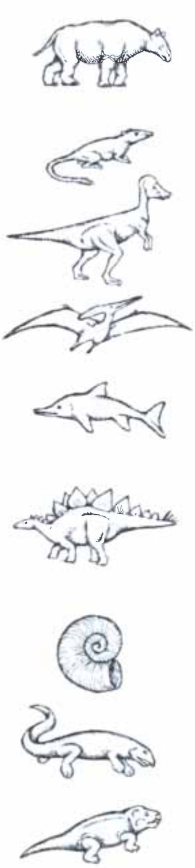
The sheer size of the Deccan Traps suggests that their formation must have been an important event in the earth's history. Individual lava flows extend well over 10,000 square kilometers and have a volume exceeding 10,000 cubic kilometers. The thickness of the flows averages from 10 to 50 meters and sometimes reaches 150 meters. In western India the accumulation of lava flows is 2,400 meters thick (more than a quarter the height of Mount Everest). The flows may have originally covered more than two million square kilometers, and the total volume may have exceeded two million cubic kilometers.

An important, unresolved question was whether the date and duration of Deccan volcanism are compatible with the age and thickness of the KT boundary. Until recently the lava samples from the Deccan Traps were thought to range in age from 80 to 30 million years (estimated by measuring the decay of the radioactive isotope potassium 40 in rocks). Whether this range was real or just reflected an error in measurement was unknown. So in 1985 I joined forces with a number of

VINCENT E. COURTILLOT has studied a range of geologic phenomena, including variation of the earth's magnetic field, continental rifting and deformation, and polar wander. He is professor of geophysics at the Institute of Physics of the Earth in Paris and director of the institute's Paleomagnetism and Geodynamics Laboratory. Courtillot is also director of research and graduate studies for the French university system in the Ministry of National Education.



FLOOD BASALT PROVINCES are tremendous lava flows that record sites of past geologic upheaval. Each province is linked with a hot spot, a locus of ongoing volcanic activity that remains relatively fixed as lithospheric plates slide over it.



MASS EXTINCTION (Approximate age in millions of years)	FLOOD BASALT	RELATED HOT SPOT
MID MIOCENE (14±3)	COLUMBIA RIVER (16±1)	YELLOWSTONE
LATE EOCENE (36±2)	ETHIOPIA (35±2)	AFARS
MAASTRICHTIAN (65±1) KT BOUNDARY	BRITO-ARCTIC (62±2)	ICELAND
	DECCAN TRAPS (66±2)	REUNION
CENOMANIAN (91±2)	WEST PACIFIC SEAMOUNTS (92±3)	PACIFIC SUPERSWELL
TITHONIAN (137±7)	PARANA (130±5)	TRISTAN da CUNHA
	NAMIBIA (135±5)	
PLIENSBACHIAN (191±3)	KAROO (190±5)	MARION
NORIAN (211±8)	NORTH ATLANTIC (200±5)	AZORES/GREAT METEOR
DZHULFIAN (249±4) PT BOUNDARY	SIBERIA (250±10)	JAN MAYEN

colleagues to try to clarify the picture.

One important clue emerged from the fact that the Deccan rocks are basalts, volcanic rocks rich in magnesium, titanium and iron that are rather strongly magnetic. When basaltic lava cools, the magnetization of tiny crystals of iron-titanium oxides in the rock becomes frozen, aligned with the earth's magnetic field. The polarity of the field occasionally reverses, so that the magnetic north pole becomes south and vice versa. These brief reversals—about 10,000 years long—occur in random fashion at a rate that has varied from about one reversal every million years at the end of the Cretaceous to roughly four every million years in recent times.

Jean Besse and Didier Vandamme at the Institute of Physics of the Earth in Paris and I found that more than 80 percent of the rock samples from the Deccan Traps had the same, reversed polarity. Had the volcanism truly continued from 80 to 30 million years ago, we would have expected to find approximately equal numbers of normal- and reverse-magnetized samples, because tens of reversals took place during that 50-million-year stretch.

In fact, the thickest (1,000-meter-thick) exposed sections of the traps record only one or two reversals. We therefore concluded in 1986 that Deccan volcanism began during an interval of normal magnetic activity, climaxed in the next, reversed interval, then waned in a final, normal interval. Judging from the usual frequency of reversals, our results implied that the volcanism could not have lasted much more than one million years.

If so, the spread of ages found by potassium 40 dating must have been wrong. My colleagues Henri Maluski of the University of Montpellier and Gilbert Féraud of the University of Nice and other researchers used a newer, more reliable technique—argon-argon dating—to determine how much potassium 40 had decayed during the lifetime of the rock samples. Their results confirmed that the Deccan flows were laid down over a relatively brief period. Age estimates for the Deccan lavas now cluster between 64 and 68 million years, and much of the remaining scatter in ages may result from alteration of the samples or differing laboratory standards.

Although accurate dating of sedimentary rock is difficult, recent findings by Ashok Sahni of the University of Chandigarh, J. J. Jaeger of the University of Montpellier and their colleagues further narrow estimates of the age of the Deccan Traps. Sediments immediately below the Deccan flows contain dinosaur fossil fragments that seem to date from the Maastrichtian stage, the last eight million years of the Cretaceous. Dinosaur and mammalian teeth and dinosaur egg fragments that appear to be of Maastrichtian age have also been found in layers of sediment between the flows. This implies that Deccan volcanism began during the very last stage of the Cretaceous.

More precise data come from oil-exploration wells on the east coast of India, which crossed three thin trap flows, each separated by a layer of sedimentary rock. The lowest level of lava rests on sedimentary layers that contain fossils of a plankton called *Abatomphalus mayaroensis*, which thrived during the last one million years of the Cretaceous and became extinct shortly thereafter. The sedimentary rock layers between the lava flows also contain fossils from the exact same time, but the layers above the flows do not.

A. mayaroensis fossils appear in strata with normal magnetic polarity that lie below (before) the KT boundary and disappear at the boundary itself, which is located in the next, magnetically reversed set of strata [*see illustration on page 90*].

The most reasonable conclusion from the various evidence is that Deccan volcanism began during the last normal magnetic interval of the Cretaceous, climaxed during the following reversed interval (at or very near the Cretaceous-Tertiary boundary) and ended in the first normal magnetic interval of the Cenozoic era.

DECCAN TRAPS in India are layered flows of basaltic lava that were laid down at the Cretaceous-Tertiary boundary, when the last major mass extinction occurred.

Magnetic and fossil studies together reduce the estimated duration of Deccan volcanism to about 500,000 years, the best time resolution that can be obtained using present techniques. The fact that Deccan volcanism—one of the largest and fastest episodes of lava flow of the past 250 million years—coincided with the KT boundary to within the best time accuracy now attainable made it hard for us to escape the conclusion that a link existed between the Deccan Traps and the mass extinction.

Having established that the Deccan Traps erupted roughly simultaneously with the extinction at the end of the Cretaceous period, we next sought to determine whether a volcanic eruption could explain the observed features of the KT boundary layers. In general, either a huge volcanic eruption or an asteroid impact could plausibly have produced these features.

The unusual iridium-rich deposit that appears to have been laid down simultaneously around the earth need not have come from outer space. William H. Zoller, Ilhan Olmez and their colleagues at the University of Maryland at College Park discovered unusual iridium enhancements in particles emitted by the Kilauea volcano in Hawaii. J. P. Toutain and G. Meyer of the Institute of Physics of the Earth found iridium in particles emitted by another volcano, the Piton de la Fournaise on the island of Réunion, which (as discussed below) is related to the Deccan volcanism. Iridium-rich volcanic dust has been found embedded in the Antarctic ice sheet, thousands of kilometers from the source volcanoes.

The composition of the clay at the boundary layer differs from that of the clays above and below the layer. The usual mineral in clay, illite, is replaced by smectite, which can be created when basaltic rock is altered. Recent studies of the mineralogy of the KT boundary clay at Stevns Klint in Denmark led W. Crawford Elliott of Case Western Reserve University and his co-workers and Birger Schmitz of the University of Göteborg to conclude that the clay consists of a distinctive kind of smectite that in fact is altered volcanic ash.

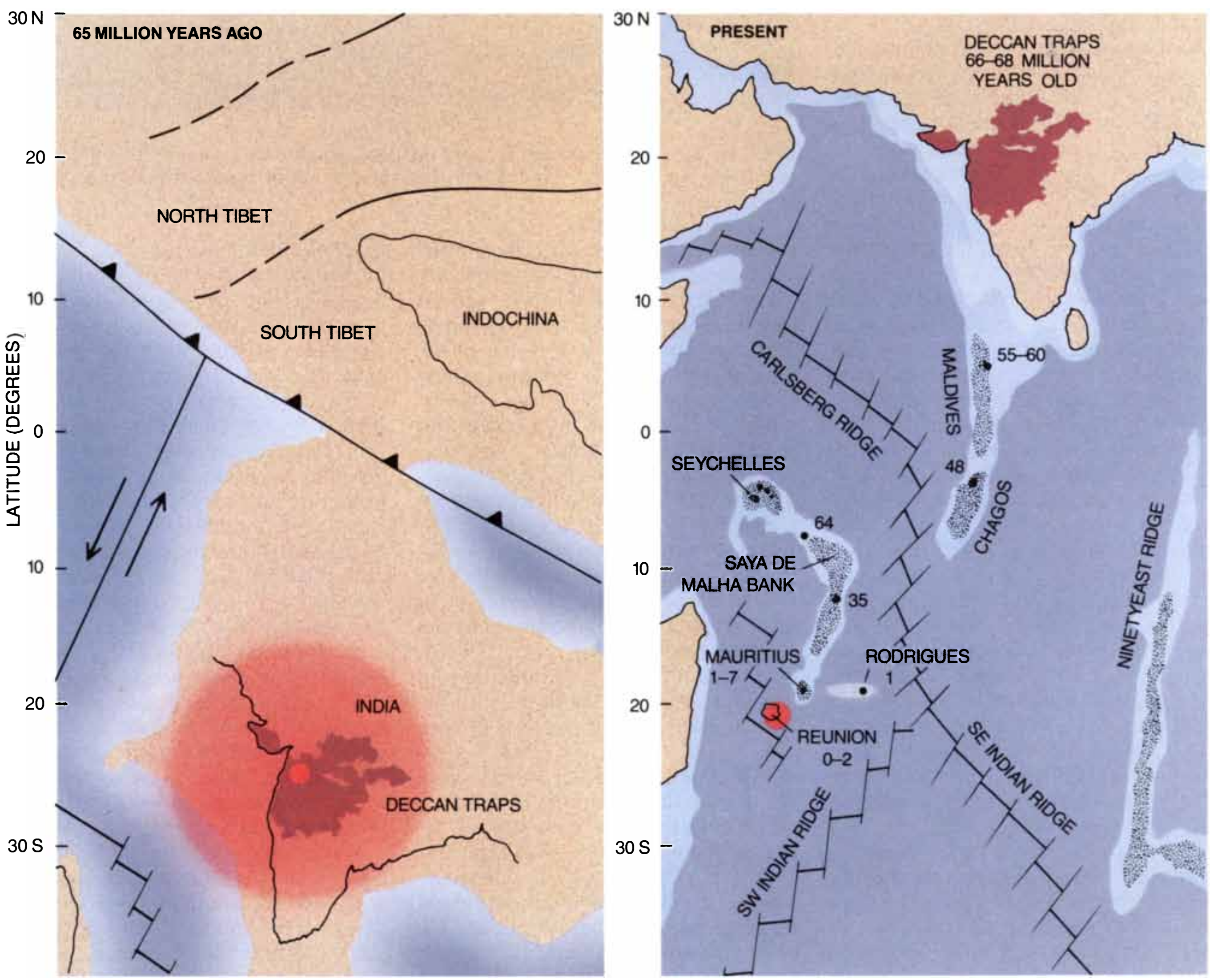
The KT boundary clay can be simulated by mixing 10 parts of material from the earth's crust with one part of material from common stony meteorites. The earth's mantle (the layer below the crust), however, has a composition similar to that of stony meteorites and so could generate the same chemical anomalies. Karl K. Turekian of Yale University and Jean-Marc Luck, then at the Institute of Physics of the Earth, found that the relative abundance of the elements rhenium and osmium in the clay resembles the ratio in both meteorites and in the earth's mantle.

Peculiar physical features in material from the KT boundary also can be explained by either hypothesis. Boundary layers contain large numbers of tiny spherules. Some spherules consist of clay minerals that appear to be altered remains of molten basaltic droplets, but it is impossible to say whether they originated as volcanic ejecta or from oceanic crust melted by an asteroid impact. Matters are somewhat confused by the fact that at least some of the spheres turned out to be round fossil algae or even recent insect eggs that contaminated the material.

The discovery of shocked, deformed grains of quartz crystal in KT boundary layers, first made by Bruce F. Bohor and Glen A. Izett of the U.S. Geological Survey in Denver, is often considered the strongest evidence in favor of the impact hypothesis. Such shocked grains had been found previously only from known impact craters (such as Meteor Crater in Arizona) or from sites of underground nuclear explosions. They are produced by dynamic shock stress at more than 100,000 times atmospheric pressure, but shocked structures can be produced at much lower pressures if the rock is heated before the shock occurs, as would be the case in a volcanic eruption.

As magma rises to the earth's surface, it decompresses and releases dissolved gases. At the same time, the magma often cools and thickens. If it cools particularly quickly, it becomes so stiff that the gases cannot escape. Pressure therefore builds up, possibly leading to an explosion and powerful shock waves. Such stresses might be sufficient to shock quartz crystals if the temperatures and duration were great enough.

Magma that is rich in silicate material is viscous and especially prone to provoke explosive eruptions; examples of silicic volcanism include Vesuvius and Mount St. Helens. In 1986 Neville L. Carter of Texas A&M University and his associates discovered evidence of shock features similar to those at the KT boundary in rocks from some geologically recent silicic volcanic explosions, such as the large Toba, Sumatra, eruption of 75,000 years ago. Us-

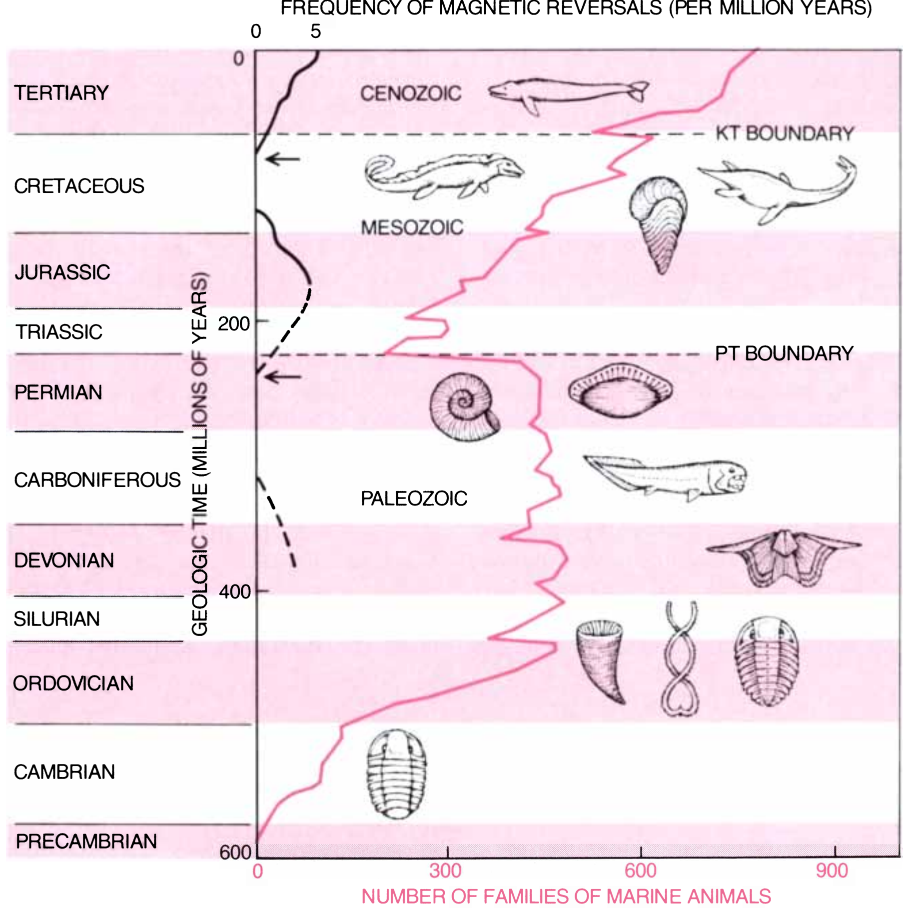


VOLCANIC HOT SPOT now beneath the volcano on the island of Réunion (the Piton de la Fournaise) was under India when the Deccan Traps erupted, according to research by the author (*left*). Ages of the seamounts south of India increase steadily from Réunion to the Deccan Traps, where the line of volcanic activity made its first appearance (*right*).

ing transmission electron microscopy, Jean-Claude Doukhan of the University of Lille recently found that shock features produced by laboratory impact, meteorite impact and those observed in samples from the KT boundary are all different from one another in some respects and that the similarity between laboratory and meteorite features has been overstated. Shock features from KT samples are decorated with microscopic bubbles that are not observed in samples from meteorite impacts and that seem to indicate a higher formation temperature, compatible with a volcanic origin.

Explosive silicic volcanism commonly precedes periods of relatively quiet, Deccan-type (flood basaltic) volcanism, during which basaltic lava flows freely and copiously. Ten to 15 percent of the volume of lava from known Deccan-type flows erupts in episodes of explosive silicic volcanism. A rising plume of hot magma would melt its way through the continental crust, producing the viscous silicic (acidic) magmas that lead to explosive volcanism.

The unusual chemical and physical features in the KT boundary layers are present worldwide. An asteroid impact could have propelled material into the stratosphere, where it would have been transported around the globe. On the other hand, Richard B. Stothers and his co-workers at the National Aeronautics and Space Administration's Goddard Space Flight Center in Greenbelt, Md., modeled the manner in which fountains of lava, such as those from Kilauea in Hawaii, expel dust and ejecta. When scaled up to the dimensions of the Deccan volcanism, their models predict that large amounts of material should also be lofted into the stratosphere. Atmospheric circulation would distribute material rather evenly between the two hemispheres, no matter where it was originally emitted.



MARINE ANIMAL DIVERSITY is plotted alongside the rate of reversals of the earth's magnetic field (data prior to 165 million years ago are approximate). Two long periods with no reversals stand out: one before the Permian-Triassic extinction of 250 million years ago, the other before the Cretaceous-Tertiary extinction of 65 million years ago. The correlations suggest a causal relation between the behavior of the earth's core, where the magnetic field is generated, and mass extinctions.

The appalling consequences of an asteroid impact and a massive volcanism would be quite similar. The first effect would have been darkness resulting from large amounts of dust (either impact ejecta or volcanic ash) into the atmosphere. The darkness would have halted photosynthesis, causing food chains to collapse. Such environmental trauma appears to be reflected in the fossil record. Freshwater creatures were much less affected than land- or sea-based ones, perhaps because freshwater animals did not feed on vascular plants (as do many land-dwelling animals) or on photosynthetic plankton (an important food source for marine vertebrates that was devastated at the end of the Cretaceous).

Life would also have been confronted by large-scale toxic acid rain. The heat of a large impact would have triggered chemical reactions in the atmosphere that would in turn produce nitric acid. Alternatively, volcanic eruptions would have emitted sulfur that would form sulfuric acid in the air. The environmental effects of sulfur-rich volcanism can be significant even in the case of fairly moderate eruptions. The 1783 eruption at Laki, Iceland, killed 75 percent of all livestock and eventually 24 percent of the country's population, even though it released only 12 cubic kilometers of basaltic lava. The event was followed by strange dry fogs and an unusually cold winter in the Northern Hemisphere.

Using the Kilauea eruption as a model, Terrence M. Gerlach of Sandia National Laboratory in Albuquerque estimated that the Deccan Traps injected up to 30 trillion tons of carbon dioxide, six trillion tons of sulfur and 60 billion tons of halogens (reactive elements such as chlorine and fluorine) into the lower atmosphere over a few hundred years. The emissions from the Laki eruption seem to have been far greater than would be expected from simply scaling up the figures for Kilauea, so the estimates may represent a lower limit. Airborne sulfur and dust from a 1,000-cubic-kilometer lava flow could decrease average global temperatures by three to five degrees Celsius (five to nine degrees Fahrenheit).

Other factors could contribute to an opposite effect, however. Marc Javoy and Gil Michard, both of the Institute of Physics of the Earth and the University of Paris, propose that sulfur dioxide from Deccan volcanoes turned the ocean surface acidic, killing the algae that normally extract carbon dioxide from the atmosphere and then carry it to the ocean bottom when they die. Acidic ocean waters also would have dissolved carbonate sediments at the bottom, releasing trapped carbon dioxide. Altogether atmospheric carbon dioxide levels would shoot up to about eight times the present concentration, producing a rise in temperature of five degrees C (nine degrees F). The interaction between cooling from dust and warming from carbon dioxide (which may occur on widely different time scales) is unclear, but the resulting climate gyrations probably would have been especially traumatic for the global ecosystem. Both the asteroid and volcanic hypotheses predict overlapping cooling and warming effects.

So far the evidence discussed has been equally consistent with both hypotheses. But many details suggest that the mass extinction and odd physical processes that occurred at the end of the Cretaceous took place over hundreds of thousands of years. This period is comparable to the duration of Deccan volcanism but incompatible with a sudden asteroid impact.

A number of paleontologists have pointed out that the extinction at the end of the Cretaceous was not a single, instantaneous event. Extinction rates appear to have started to increase up to a million years before the KT boundary. Even near the boundary, the pattern is not uniform: for instance, planktonic foraminifera and nanoplankton (microscopic calcareous algae) species exhibit different patterns of extinction and recovery. This ragged sequence is known as stepwise mass extinction.

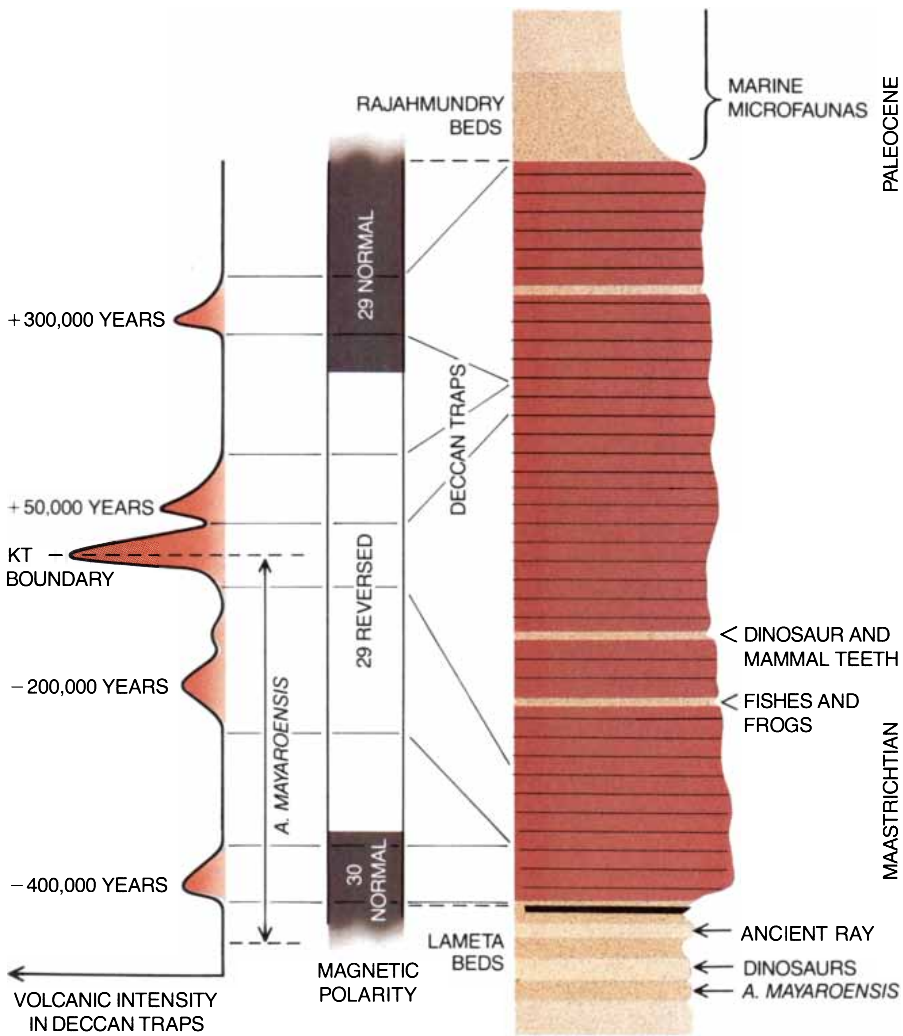
One of the most thorough recent studies of the pattern of extinctions was conducted by Gerta Keller of Princeton University. When she analyzed the well-preserved sections of the KT boundary in Tunisia and Texas, Keller found evidence for a first phase of extinction (also seen in the macrofossil record) that began 300,000 years before the KT iridium event and for another extinction event that took place 50,000 years after the boundary. Keller attributes the first event to falling sea levels and global cooling.

Other evidence confirms that the earth experienced not one but many disruptions at the end of the Cretaceous. Abrupt change occurred, for example, in the abundance of carbon 13 and oxygen 18 (respectively, light and heavy versions of these elements, whose concentrations vary according to the ocean temperature and acidity and to the number of living creatures present). Extinctions and carbon 13 fluctuations observed in strata in Spain occur in magnetic intervals that fit the same normal-reversed-normal polarity pattern found in the Deccan Traps.

Even the iridium appears to display a number of fine fluctuations near the KT boundary. Robert Rocchia and his colleagues at the Atomic Energy Commission and National Center for Scientific Research in Gif-sur-Yvette and Saclay, France, found secondary iridium peaks above and below the primary iridium layer (corresponding to time intervals of about 10,000 years) in KT boundary clay in Spain and Denmark. Rocchia, I and our colleagues found that the layer of iridium enrichment in Gubbio, Italy, seems spread over about 500,000 years. The much discussed shocked quartz crystals exhibit a similar pattern of distribution. Officer and Carter discovered that shocked minerals extend through four meters of the Gubbio section, again corresponding to a time span of about 500,000 years.

James C. Zachos of the University of Rhode Island and his co-workers measured the chemical composition of microscopic fossils from the North Pacific seafloor and found that the productivity of open-sea marine life was suppressed at the time of the KT boundary and for about 500,000 years thereafter. They also concluded that significant environmental changes, including cooling, began at least 200,000 years before the boundary.

Some proponents of the impact theory, most prominently Piet Hut of the Institute for Advanced Study in Princeton, N.J., and his colleagues, quickly substituted a series of comet impacts for the single asteroid impact to explain these findings. The search for an all-encompassing answer also led to the suggestion that the Deccan Traps might mark the site of the asteroid impact, but there are many difficulties with that idea. No traces of an impact have been found in India. Robert S. White of the University of Cambridge has shown that large impacts cannot trigger massive volcanism, because the section of the mantle just below the lithosphere (the relatively rigid crust and upper



SEDIMENTARY ROCKS sandwiched between Deccan lava flows contain dinosaur fragments and other Cretaceous fossils that are absent in layers above the flows. The mass extinction occurred while the flows were in progress (*right side*). An inferred volcanic history (*left side*) can explain the multiple extinctions and chemical anomalies seen at the end of the Cretaceous period. Magnetic reversals help synchronize the observed layers with the hypothesized volcanic behavior.

mantle) does not normally contain large reserves of molten rock. Moreover, Deccan volcanism started during a normal geomagnetic interval, a few hundred thousand years before the reversed magnetic interval containing the KT iridium anomaly and the clay layer.

During the Cretaceous period, volcanism increased, the sea level rose and fell drastically and the global mantle shifted significantly. The Cretaceous period and the one that preceded it, the Jurassic, were also times of major continental breakups. Between 120 and 85 million years ago, the earth's magnetic field did not undergo a single magnetic reversal, but 15 to 20 million years before the KT boundary, the field started reversing again. Reversal frequency, which indicates activity in the earth's core and at the core-mantle boundary, has increased regularly since then to about once every 250,000 years at present.

All these features can be related to an episode of energetic mantle convection that began tens of millions of years before the KT boundary. To me, the existence of overlapping short- and long-term geodynamic, geologic and paleontological anomalies points to a common internal cause.

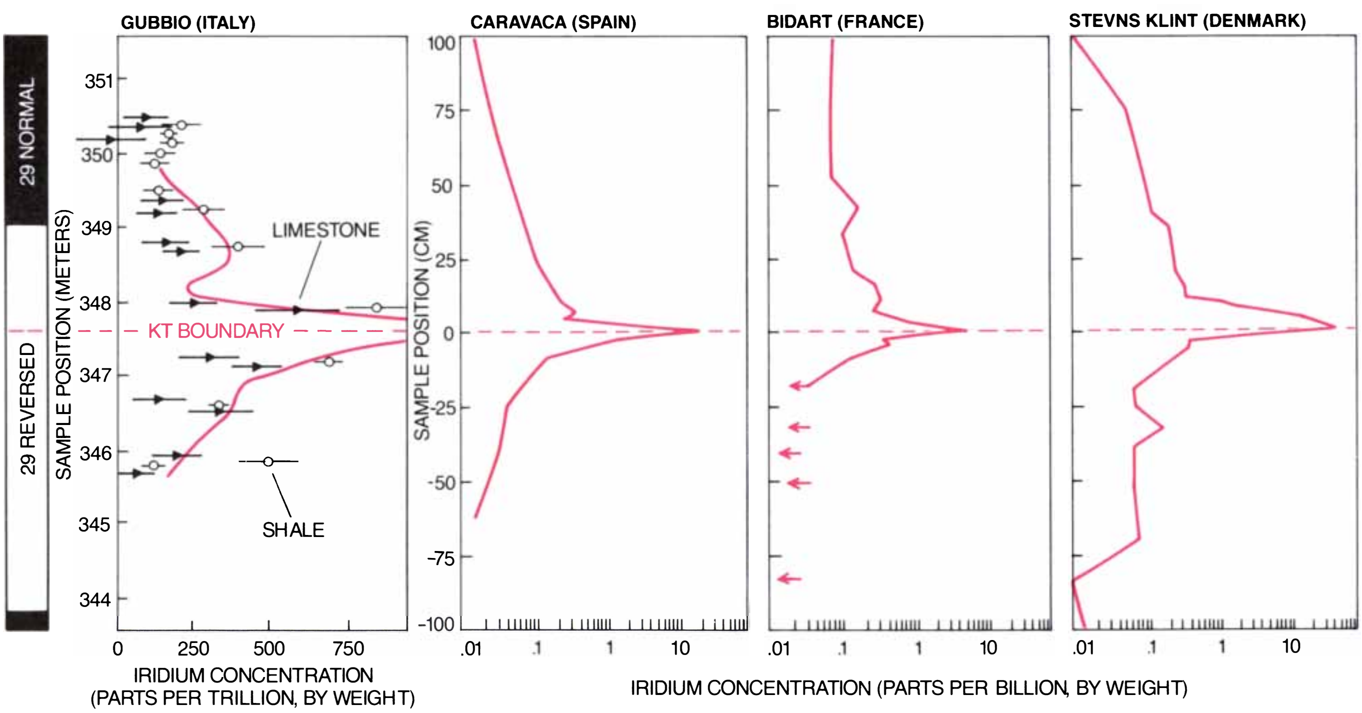
What might that cause be? A likely answer comes from the theory of mantle hot spots, developed most prominently by W. Jason Morgan of Princeton University and others. Peter L. Olson and Harvey Singer of Johns Hopkins University developed a model that may explain these regions of persistent volcanic activity. A plume of hot, low-density and low-viscosity material rises from the lowermost parts of the mantle, forming a quasi-spherical head as it pushes its way through cooler, thicker mantle. The head keeps growing as long as it is fed by a conduit of molten rock rising from below.

White and Dan P. McKenzie, also of Cambridge, along with Mark Richards and Robert A. Duncan of Oregon State University and myself, think that as a hot mantle plume rises, the crust above the plume lifts and stretches, leading to continental rifting [see "Volcanism at Rifts," by Robert S. White and Dan P. McKenzie; SCIENTIFIC AMERICAN, July, 1989]. The plume material decompresses as it reaches the surface and so melts rapidly (in less than one million years). The head of the plume would elevate a large area of crust, so that when the magma finally breaks through to the surface, it runs rapidly downhill, producing extensive flows.

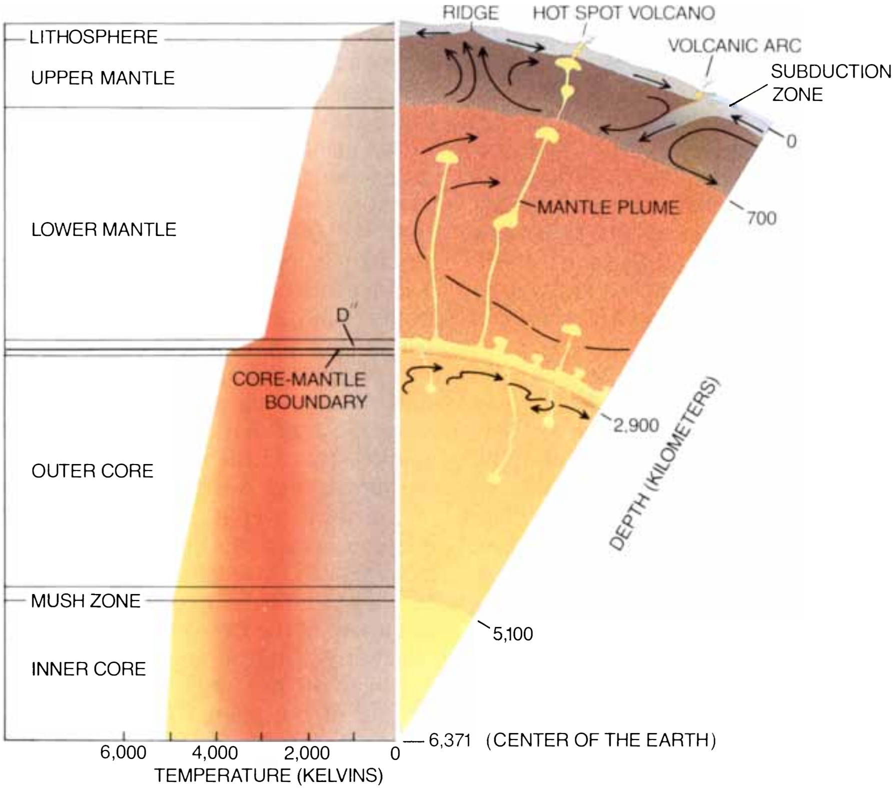
The Deccan eruptions could have followed the arrival of such a head at the base of the lithosphere. Volcanism from a hot plume would be rapid and highly episodic. Individual flows would be extruded in days or weeks; the next flow would follow years to thousands of years later. The far-reaching ecological consequences of each flow could explain the stepwise mass extinctions.

The giant mantle plume that produced the Deccan Traps should have left structural and dynamic relics. In 1987 the Ocean Drilling Program, led by Duncan, explored and dated an undersea chain of volcanoes that extends from southwest India, near the Deccan Traps, to Réunion, the active volcano east of Madagascar. Réunion is a hot spot volcano—one powered by a deep, rising flow of hot magma from the mantle—that burned its way through the Indian and African continents as they drifted over it. The ages of the Réunion seamounts increase steadily from zero to two million years around Réunion itself to 55 to 60 million years just south of the Deccan Traps.

Richards, Duncan and I believe that the Réunion hot spot may represent the tail of hot magma that would be expected to follow in the wake of the plume that produced the traps. Besse, Vandamme and I verified that the mantle hot spot now beneath Réunion was located precisely under the Deccan Traps at the end of the Cretaceous. There is no trace of the hot spot from before the KT boundary; the episode of violent Deccan volcanism appears to mark the appearance of



IRIDIUM could have been released either by an asteroid impact or by a massive volcanic eruption. The elevated iridium concentration in strata from the Cretaceous-Tertiary boundary in Italy, Spain, France and Denmark extends over as much as 500,000 years. An asteroid impact would have produced a sudden, well-defined rise in iridium. The Deccan eruptions, in contrast, appear to have persisted for just the right length of time to account for the observed iridium distribution.



MANTLE PLUME MODEL is depicted in this schematic cross section of the earth. Heat from the outer core thickens the lowermost layer of the mantle, the D´´, suppressing the core convection that produces magnetic field reversals. The D´´ eventually breaks up into huge, rising plumes. The lack of magnetic reversals during the Cretaceous may indicate a growing D´´ layer, which then shed hot plumes that produced catastrophic volcanism—the Deccan Traps—when they reached the surface 65 million years ago. The Piton de la Fournaise may represent the plume's tail.

the hot spot at the surface of the earth.

The internal geologic activity associated with a rising mantle plume fits the behavior of the earth's magnetic field at the time of the KT boundary. Slow convection of the molten iron in the earth's outer core—10 kilometers per year—is thought to produce the earth's magnetic field. Instabilities at the boundary between the core and the mantle above it may cause magnetic reversals.

Heat escaping from the core raises the temperature and so lowers the density of material in the deepest layer of the mantle (called the D´´), which grows thicker until it becomes unstable and forms rising plumes of magma. Long durations with few or no magnetic reversals, such as the span from 120 to 85 million years ago, indicate a lack of outer core activity and the growth of the D´´ layer.

About 80 million years ago the layer broke up, sending enormous hot magma plumes upward. At this point, flow of heat from the core to the mantle would have increased, and magnetic reversals would have resumed. At typical mantle velocities of about one meter a year, the plumes would have traveled a few million years before reaching the surface, where their sudden decompression of the plumes would have led to explosive volcanism followed by large lava flows. Smaller, secondary plumes would not have reached the surface but could have accelerated mantle convection, seafloor spreading, sea level changes and other geologic disruptions that took place during the Cretaceous.

This kind of geologic upheaval may be a natural consequence of the fact that the earth is an active, complex heat engine composed of layers that have vastly different physical and chemical properties. Smooth, well-regulated mantle convection and brutal, plumelike instabilities are perhaps just two extremes of the ways in which the earth's internal heat escapes to the outside.

If this is indeed the way the earth functions, similar catastrophes should have taken place. In fact, most major, relatively recent extinction events (those since the Mesozoic era began 250 million years ago) seem to correlate in time with a large flood basalt eruption. Interestingly, the longest known period during which the earth's magnetic field did not reverse also ended with the largest mass extinction, the one that marked the dawn of the Mesozoic era. More than 95 percent of marine species disappeared at that time. The 250-million-year-old Siberian Traps are a prime candidate for having caused this extinction.

Both the asteroid impact and volcanic hypotheses imply that short-term catastrophes are of great importance in shaping the evolution of life. This view would seem to contradict the concept of uniformitarianism, a guiding principle of geology that holds that the present state of the world can be explained by invoking currently occurring geologic processes over long intervals. On a qualitative level, volcanic eruptions and meteorite impacts happen all the time and are not unusual. On a quantitative level, however, the event witnessed by the dinosaurs is unlike any other of at least the past 250 million years.

Magnetic reversals in the earth's core and eruptions of large plumes in the mantle may be manifestations of the fact that the earth is a chaotic system. Variations in the frequency of magnetic reversals and breakup of continents over the past few hundred million years hint that the system may be quasi-periodic: catastrophic volcanic episodes seem to occur at intervals of 200 million years, with lesser events spaced some 30 million years apart.

It is tempting to speculate that the dawn of the Paleozoic era 570 million years ago, when multicellular life first appeared, might have coincided with one such episode. Large extinctions abruptly open broad swaths of ecological space that permit new organisms to develop. Events that at first seem to have been disasters may in fact have been agents essential in the evolution of complex life.

FURTHER READING

LATE CRETACEOUS AND PAROXYSMAL CRETACEOUS/TERTIARY EXTINCTIONS. Charles B. Officer, Anthony Hallam, Charles L. Drake and Joseph D. Devine in *Nature*, Vol. 326, No. 6109, pages 143–149; March 12, 1987.

MAGNETIC FIELD REVERSALS, POLAR WIND AND CORE-MANTLE COUPLING. V. Courtillot and J. Besse in *Science*, Vol. 237, No. 4819, pages 1140–1147; September 4, 1987.

THE DECCAN FLOOD BASALTS AND THE CRETACEOUS/TERTIARY BOUNDARY. V. Courtillot, G. Féraud, H. Maluski, D. Vandamme, M. G. Moreau and J. Besse in *Nature*, Vol. 333, No. 6176, pages 843–846; June 30, 1988.

Ocean Acoustic Tomography

Oceanographers are using principles analogous to medical X-ray tomography to obtain three-dimensional images of ocean currents and temperatures

by Robert C. Spindel and Peter F. Worcester

In January of 1981 we set sail from Woods Hole, Mass., heading to an area just southwest of Bermuda, centered on 26 degrees north latitude by 70 degrees west longitude. On board the *Oceanus* we had some new scientific instruments that worked well in the laboratory but needed to be tested in the corrosive, buffeting seas. Much as a physician might use sound waves to make a snapshot of, say, a fetus, we were going to use acoustic waves to generate three-dimensional snapshots that would enhance our understanding of the ocean's properties and its fluctuations over time.

The ocean has long been known to moderate land temperatures, to affect seasonal climate changes and to transfer dissolved chemicals, nutrients, pollutants and even plants and animals from one part of the earth to another. This enormous fluid fluctuates in a highly complex manner. The turbulence varies from constant, small-scale mixing and internal waves to seasonal and longer-term currents and basin-scale circulations, such as the Gulf Stream. Making accurate measurements of ocean structure by means of conventional instruments is therefore difficult.

So during the past decade we and our colleagues have been developing a technique called ocean acoustic tomography. Like the computed tomography (CT) used in medicine and seismic tomography used in geology, ocean tomography employs beams of energy to create a three-dimensional image of the area traversed. In CT scans, these energy beams consist of X rays; in seismic tomography, shock waves from earthquakes; and in ocean tomography, low-frequency sound waves.

The tomographic technique has already invited reexamination of many previous notions about the seas. Before 1965 many scientists pictured the ocean's general circulation as consisting of large, stately, slow-moving currents, such as the Gulf Stream and the Kuroshio (which flows northward to the east of Japan).

That view, based on 100 years of sporadic observations made around the globe, produced only a rough average of the true circulation. But in the 1950s and 1960s, researchers began using recently developed techniques and equipment, including subsurface floats that moved with ocean currents and emitted identification signals (and thus could be followed by ships) and ocean-current meters that recorded data for several months at fixed locations in the ocean.

These instruments disclosed an unexpected level of variability in the deep ocean. Rather than being characterized by smooth, basin-scale currents that change seasonally (if at all), the seas are in fact dominated by what oceanographers call mesoscale fields: fluctuating, highly energetic flows whose velocity at times reaches 10 times the mean velocity of the major currents.

Mesoscale phenomena—the oceanic analogue of weather systems—extend 100 kilometers and persist for 100 days (weather systems generally extend about 1,000 kilometers and last three to five days). More than 90 percent of the kinetic energy of the entire ocean may be characterized by mesoscale variability rather than by large-scale currents; the mesoscale may play a significant role in oceanic mixing, air-sea interactions and occasional, but important, climatic events such as El Niño, the atmospheric-oceanic disturbance in the equatorial Pacific that affects global weather patterns.

Conventional measurement techniques are, unfortunately, not up to the task. To sample the mesoscale field properly, monitoring equipment would have to be laid out on a grid at intervals of at least 50 kilometers. At each grid point, sensors would have to be lowered to the average ocean depth of about five kilometers and remain in place for many months.

Such detailed mapping would be prohibitively expensive and impractical. To cover an area 1,000 by 1,000 kilometers, about 400 instrumented moorings would be needed. Today the entire international research community deploys fewer than 50 moorings of this type each year, and the total number of these moorings in the world is considerably less than the 2,000 that would be required.

Ships are another possibility, but they would need to stop every 50 kilometers to lower measuring instruments. In addition, a comprehensive view of the one million square kilometers would require 10 ships devoted solely to gathering data. Satellite-borne devices provide extremely useful information, but they cannot obtain data from the ocean's interior.

To overcome the difficulties of conventional techniques, Walter H. Munk of the Scripps Institution of Oceanography and Carl Wunsch of the Massachusetts Institute of Technology proposed in 1979 that tomography be adapted to measuring the physical properties of the ocean. In medical tomography, or CT scanning, X rays map the density variations (and hence internal organs) of the body; the information from the X rays, transmitted

ROBERT C. SPINDEL and PETER F. WORCESTER have participated in ocean acoustic tomography since its inception in 1979. Spindel is director of the applied physics laboratory at the University of Washington. He received his Ph.D. in electrical engineering from Yale University and did his postdoctoral work at the Woods Hole Oceanographic Institution, where he chaired the department of ocean engineering from 1982 to 1987. Worcester is research oceanographer at the Scripps Institution of Oceanography, where he received his Ph.D. in 1977. His thesis was one of the first efforts to explore the possible use of acoustic techniques to sense remotely the ocean interior.

through the body along many different paths, is mathematically recombined to form a three-dimensional image. (Conventional X rays follow single paths and cannot distinguish organs with similar densities.)

Munk and Wunsch reasoned that by transmitting acoustic signals over many paths, they could deduce the properties of the ocean's interior—its temperatures, salinities, densities and current speeds—on the basis of how the ocean altered the signals. Ocean acoustic tomography was born.

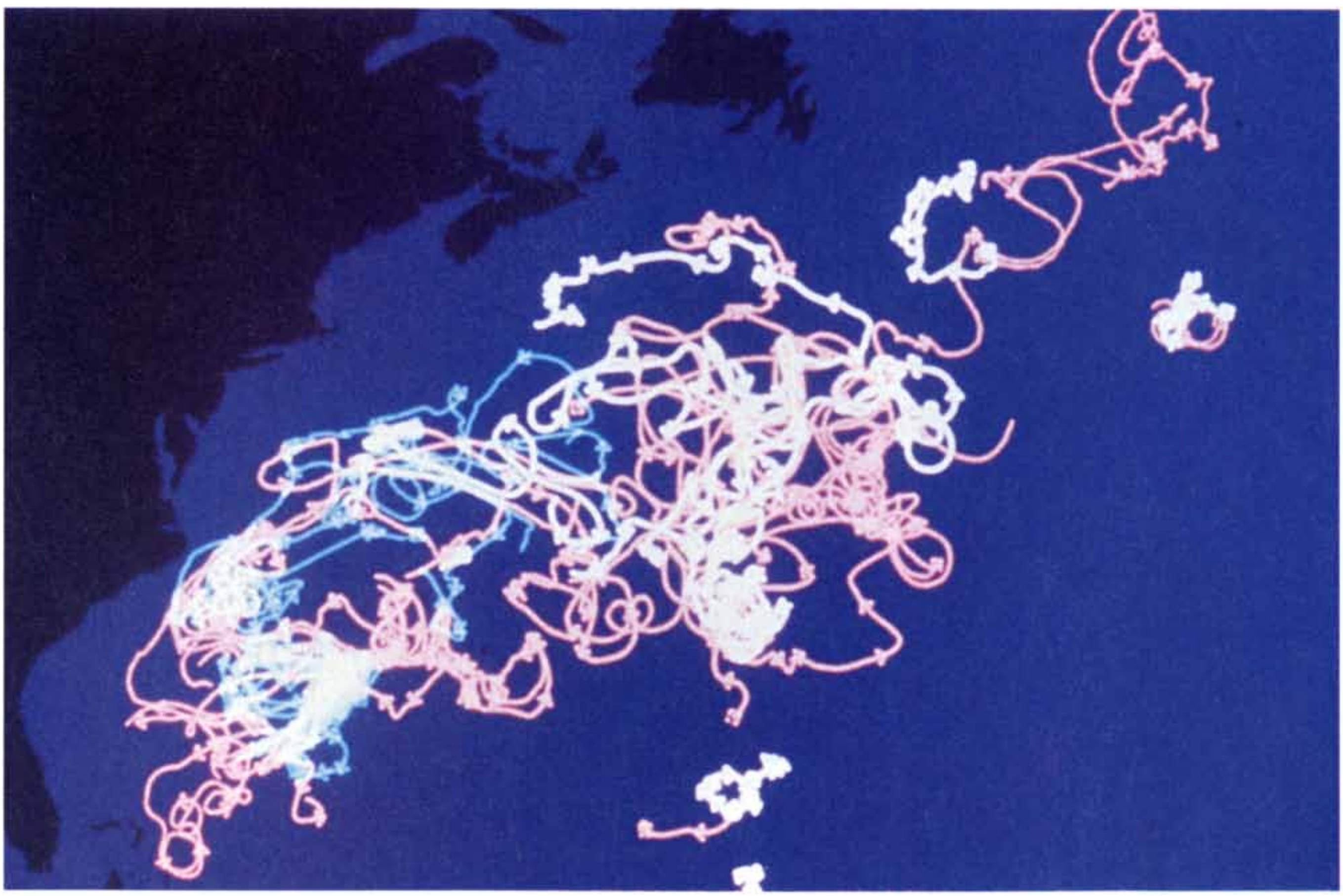
To understand how ocean acoustic tomography works, it is necessary to know how sound travels in the sea. The mathematical rules that govern other wave phenomena, such as the electromagnetic radiation from a light bulb or the vibration of a violin string, also determine the properties of acoustic waves in water: how these waves are refracted and attenuated, how they diminish in intensity as they spread from their source, how they accelerate or slow down while traveling within ocean currents. Because acoustic waves are so well described mathematically, perturbations in the sound field provide the information needed to determine the ocean's temperatures and currents.

The most significant information scientists gain from sound waves moving through water results from the changes in their velocity. The speed of sound increases primarily as temperature and pressure increase. Because the sun heats the surface of the ocean, sound tends to travel faster there than it does in deeper, cooler waters. Its speed decreases as the depth increases, eventually reaching a minimum (at about 1,000 meters in temperate zones). Below this depth, the increased pressure offsets the decreased temperature, so the speed starts to pick up. The result is a profile in which sound travels quickly near the surface, then more slowly with increasing depth and finally more rapidly again toward the bottom, where the speed may exceed that at the surface.

These variations, however, amount to only about 2 to 3 percent of the average speed of sound in water, which is about 1,500 meters per second, or roughly five times that in air. Yet such slight deviations have profound effects: they cause propagating sound waves to refract. A moving wave front can be regarded as an inflexible rod. If the upper part of the rod begins to travel faster than the lower part, the direction of the rod's motion begins to curve toward the lower part. Sound waves above the depth of minimum speed are thus refracted downward because the upper part of the wave front is in an area of faster speed; sound below the minimum-speed depth bends upward. The net effect is to trap sound waves in a "duct," where they cycle vertically through a horizontal column of water—without touching the surface or bottom—as they propagate [*see illustration on page 98*].

This duct, called the SOFAR (sound fixing and ranging) channel, was discovered during World War II and used by the U.S. Navy to locate downed aircraft pilots. The aviator released a small explosive device that sank to the SOFAR channel before detonating; hydrophones cabled to shore received the signals and determined the pilot's position by triangulation.

In the SOFAR channel, commonly referred to now as the axis of the deep sound channel, the minimum-speed depth is greatest at equatorial lati-



SUBSURFACE CURRENT TRAJECTORIES in the Atlantic Ocean, mapped by free-drifting, subsurface floats that remain at a predetermined depth, reveal the rapidly fluctuating, highly energetic current flows near the eastern seaboard of the U.S.

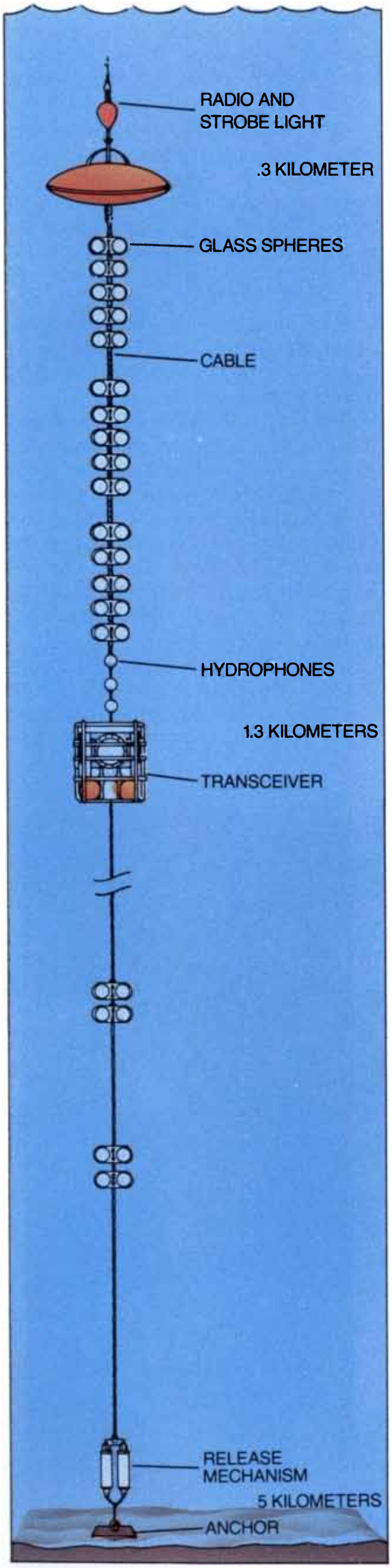
tudes, at about 1,300 meters deep. There the effect of the very warm surface water on the speed of sound counterbalances the effect of the increasing pressure at greater depths. At 30 degrees latitude, where surface waters are not as warm as they are in equatorial zones, the axis is located some 1,000 meters deep. At high latitudes, where solar heating is negligible and waters are almost uniformly cold from surface to bottom, the speed of sound is determined almost completely by pressure. Propagating sound waves are, then, refracted upward only; they reflect off the surface, scatter and lose energy. For this reason, sound transmissions are more difficult near the Arctic than in the temperate seas.

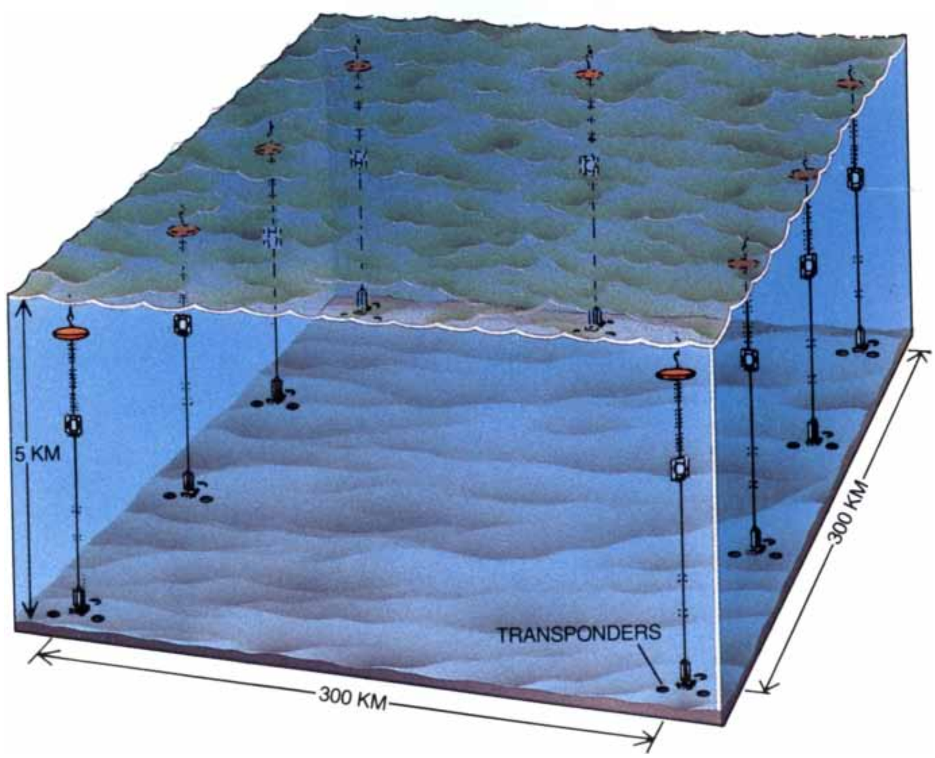
But even for signals that refract and remain trapped in the duct, oceanographers must use low-frequency sound waves to keep the acoustic energy from attenuating over the distances between transmitters and receivers. High frequencies—those of several kilohertz—are rapidly attenuated by chemical processes and molecular motions in the water. Sounds above 10 kilohertz can barely be detected at 10 kilometers; in contrast, those below 100 hertz can easily cover thousands of kilometers without significant energy losses. In 1960 low-frequency sounds from depth charges fired by the research vessel *Vema* off Australia reached receivers near Bermuda 3.5 hours later, having traveled 19,000 kilometers, halfway around the globe.

The deep sound channel not only produces ducted propagation—the basis for long-range sound transmission and reception—but also causes sound waves that radiate at different angles to travel along distinct, separate paths. This multipath propagation—the basis for tomography—samples the vertical plane between each transmitter-receiver pair.

Sound rays launched nearly parallel to the axis of the deep sound channel travel close to it, in the region where sound travels slowly. Rays that are directed at steep upward or downward angles travel high near the top and low near the bottom of the ocean, tracing approximately sinusoidal paths and spending most of their time in regions of high speeds.

Although the rays launched at steep angles travel greater total distances than those launched nearly parallel to the axis, in most areas of the ocean the steep rays arrive at the receiver first because they travel faster. A single sharp sound, such as that produced by an explosion, is thus received at a distant point in the form of replicas of the original sound. The difference in arrival times between the first and last signals, or the multipath spread, amounts





TOMOGRAPHIC MOORING (*left*) uses a radio and strobe light to identify itself for later retrieval. Large glass spheres provide buoyancy to counteract the weight of the cable. An array of moorings mapped an area near Bermuda (*right*). Three acoustic transponders surround each anchor and measure the slight movements that result from subsurface currents.

to about one second for each 100 kilometers between transmitter and receiver. Although dozens of multipaths exist for a given transmitter-receiver pair, only about 10 to 15 can usually be resolved with any clarity.

Each signal, having traveled along a separate path from the others, moves through a different part of the water column; for example, the paths close to the deep sound channel axis collected data only from the water features at that depth. The path each pulse traverses determines its unique arrival time, making the portion of the water column sampled easy to identify.

Every source transmits to every receiver, forming a network of crossing ray paths in a horizontal area [*see top illustration at right*]. Consequently, as more transmitters or receivers are added, the number of possible ray paths increases at a rate much greater than that achieved with conventional instruments, which do not collect data from crossing ray paths. The multiplicative increase in data is one of the chief attractions of tomography: it enables researchers to measure vast areas with relatively few instruments.

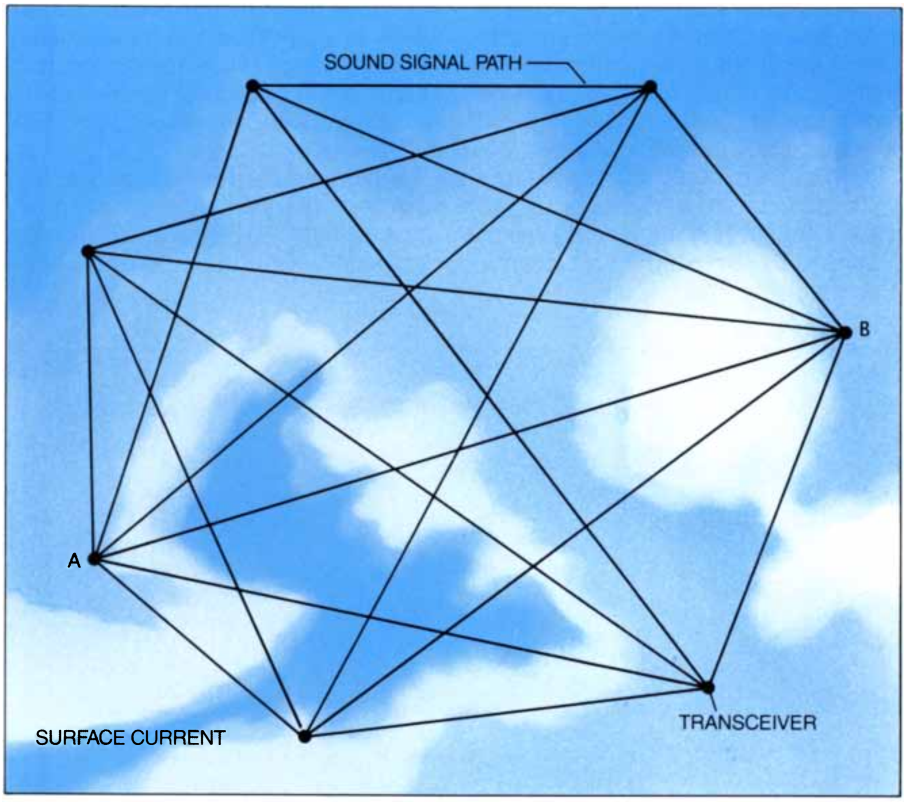
Tomography researchers can also increase the amount of data by using transceivers on each mooring instead of individual transmitters and receivers on separate moorings. Signals sent from A and received at B will only reveal temperature variations between them. But if B also sends signals to A, then by comparing both sets of signals, workers can determine current velocities.

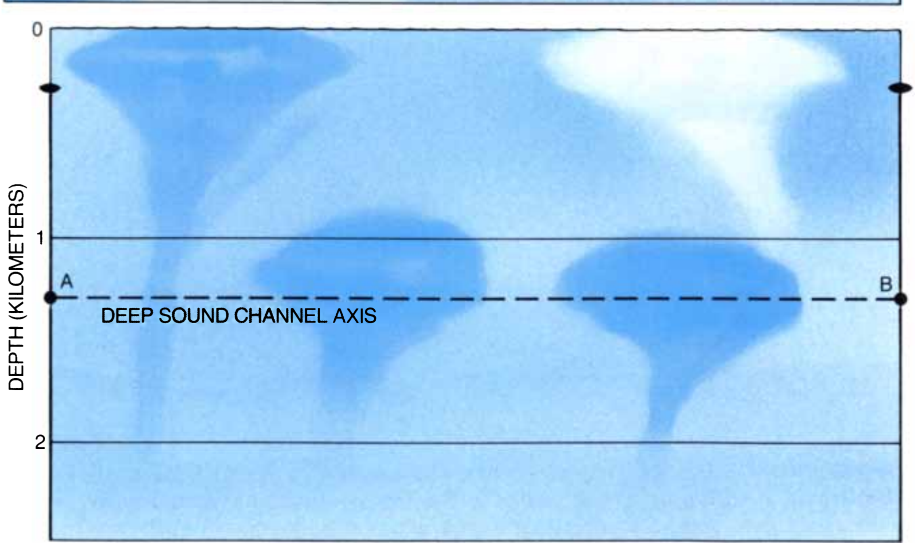
To create detailed tomographic maps of the ocean, researchers therefore moor sound transmitters and receivers (or transceivers) near the axis of the deep sound channel. Each transmitter periodically emits a pulse of low-frequency sound that is received as a series of replica pulses at each receiver. Enough transmitters and receivers are scattered over the ocean being measured to sample the horizontal area adequately, and the sound-transmission schedule is adjusted to sample the ocean's changes over time.

At about the same time Munk and Wunsch proposed their idea, the technology necessary to implement ocean acoustic tomography was just beginning to become available. In 1979 engineers at the Woods Hole Oceanographic Institution, the University of Michigan at Ann Arbor and Scripps rushed to develop instruments to conduct tomography experiments. To generate the low-frequency sounds, they modified the SOFAR float, which had originally been invented to track ocean currents in the deep sound channel. The device used in tomography emits a specially coded signal, easily distinguishable from background noise; the coded signals are similar to those used by the *Voyager* spacecraft to broadcast weak signals to the earth. Receivers process and store the codes. The SOFAR float was also redesigned to be moored at a fixed site rather than allowed to drift freely.

Timekeeping in acoustic tomography must be precise, because the changes in travel time caused by mesoscale fluctuations are slight. For example, sound traveling at 1,500 meters per second takes about 670 seconds to cover 1,000 kilometers in the ocean; a mesoscale eddy that has a temperature anomaly of one degree Celsius produces only a 50- to 100-millisecond change on top of the 670-second gross travel time. Tomography consequently requires timekeeping that is accurate to one or two milliseconds.

But even a small atomic clock that uses about 10 watts continuously consumes too much electricity; powering one for a year would require 8,000 al-





TRANSCEIVERS used in tomography are scattered over the ocean and measure currents and temperatures by sending signals to one another *(top)*. A vertical slice between A and B reveals eddies in the ocean not visible at the surface (*bottom*).

kaline flashlight batteries, a quantity too heavy and too large to be useful. Even battery packs with the highest energy density available would be impractically large and expensive. So oceanographers turned instead to less accurate but less power-hungry oscillators. To maintain timekeeping accuracy, the microprocessors in the transceiver automatically turn on a rubidium-frequency standard once a day to recalibrate the low-power oscillator. Such a timekeeping system requires the energy equivalent of only 100 alkaline batteries.

The final difficulty lay in the conventional method used to deploy instruments at sea. Traditionally, a single cable with a subsurface float is anchored to the seafloor; instruments are connected along the cable at the desired depths. The entire mooring must be kept below the ocean surface to avoid the effects of wind-induced wave motion, which not only alter measurements but also stress the mooring and cause rapid wear, fatigue and eventual failure. To retrieve the instruments, a mechanism located near the bottom of the mooring releases the anchor (when it receives an acoustic command from a ship or a signal from a timed device), allowing the mooring to rise to the surface.

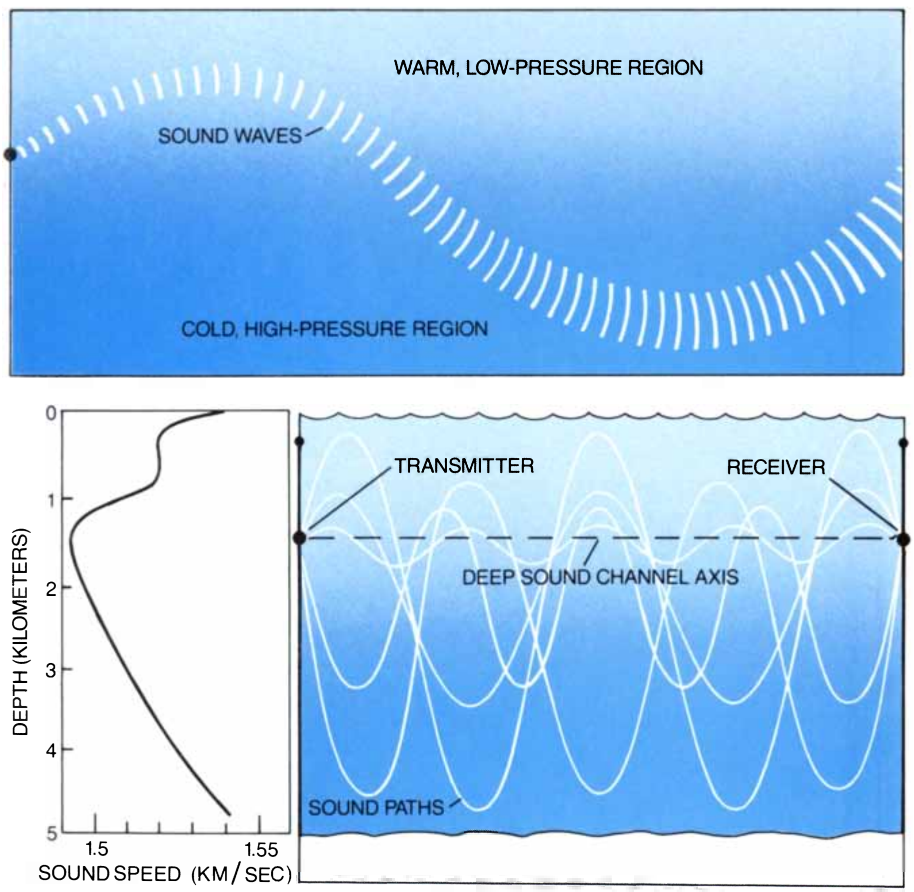
The difficulty for tomography is that such moorings move in response to subsurface currents, much the same way an anchored boat drifts in a "watch circle." The top ends of the moorings may move hundreds of meters horizontally across the ocean even with several thousand pounds of tension in the mooring line. The distance between each transceiver can therefore change by several hundred meters, altering the net travel times of transmitted sound waves by more than 100 milliseconds. These alterations show up as noise and obscure the tomographic signals, whose travel times vary only about 50 to 100 milliseconds.

To compensate for this motion, researchers devised an acoustic position-keeping system to measure the excursions. Three acoustic transponders (instruments that transmit a pulse in response to an acoustic signal) are deployed on the seafloor to form an equilateral triangle that surrounds the mooring anchor and thus each transceiver. A device housed in each transceiver unit emits a high-frequency pulse at about 10 kilohertz and measures the time between the emission and the reception of a reply from each transponder, which "answers" the 10-kilohertz tone. This process fixes the transceiver position with respect to the transponders to within about one meter, helping oceanographers to minimize the errors that result from the movements.

The mathematical methods used by oceanographers to convert the measured data into maps of the ocean temperature and current fields are similar to those used by seismologists to construct maps of the earth's interior from seismic waves generated by earthquakes [see "Seismic Tomography," by Don L. Anderson and Adam M. Dziewonski; SCIENTIFIC AMERICAN, October, 1984]. But, given essential differences between the two fields, oceanographers cannot simply borrow the techniques without modification. For example, seismologists have never been able to obtain direct measurements at depths greater than a few kilometers below the earth's surface. Instead they have had to rely on earthquakes as their source of information about the interior of the earth. Consequently, the models seismologists constructed with their data offered only limited descriptions.

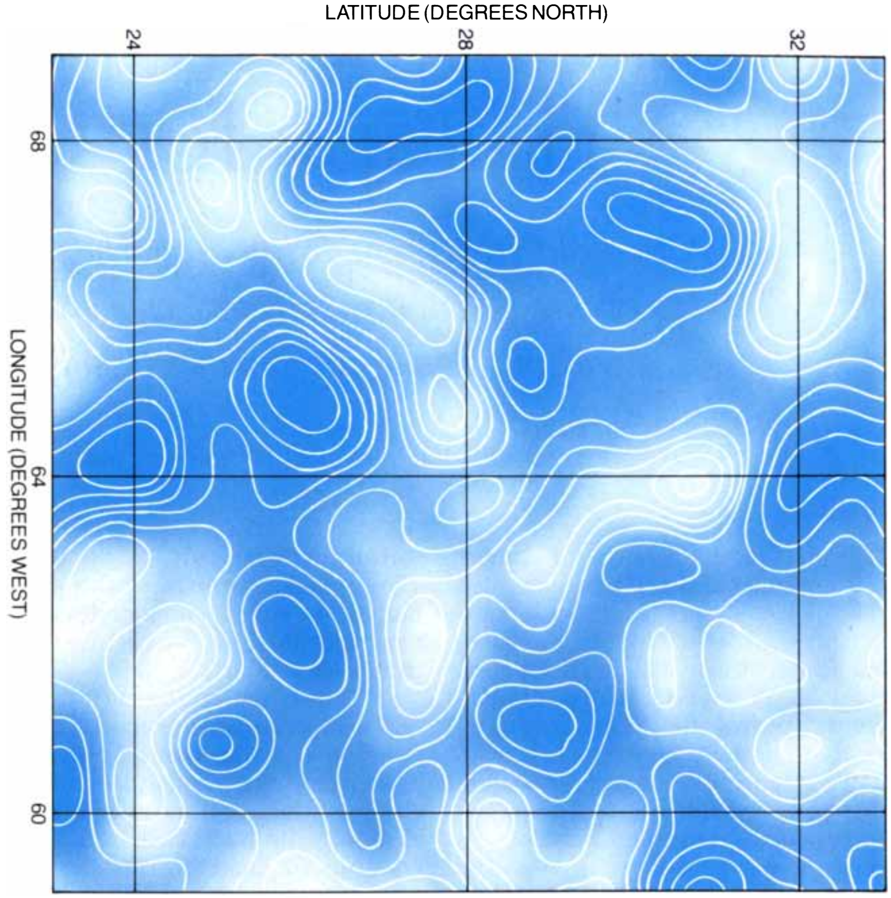
Oceanographers, in contrast, have a great advantage: direct measurements of the ocean have enabled them to form concise theoretical models and to compute statistics from these descriptions. They can constrain their data with the models and use them as guides for further observation. Ironically, the oceanographer's ability to make direct measurements of the ocean's interior probably delayed the development of remote-sensing techniques such as ocean acoustic tomography.

Another difference between seismic tomography and ocean acoustic tomography is that the earth is fairly stable over human lifetimes, whereas the ocean is constantly changing. The mathematics used to construct oceanic maps must thus consider time-dependent data from the ocean, much as weather-forecasting models must cope with rapid weather changes. Although enough information can be extracted to determine how the ocean will evolve over the course of a few weeks, techniques to insert data into time-depen-



REFRACTION OF SOUND WAVES occurs when different parts of the same wave front travel at different speeds (*top*). In the ocean, the sound speed varies with depth (*bottom left*), so that signals refract whenever they encounter the warm waters at the surface or the high pressures of the deep sea; they remain trapped in a duct and can propagate thousands of kilometers. Sound launched at different angles travels over different areas of the seas, providing information in the entire vertical plane (*bottom right*). Only a few of the 10 to 15 resolvable paths are shown.

dent models of the ocean are still in their infancy.



SIMULATED TOMOGRAPHIC "SLICE" constructed from theoretical models shows an example of the mesoscale temperature field about 700 meters deep. The patterns of warm (*light*) and cold (*dark*) waters are the oceanic analogues of weather systems.

By early 1981 the necessary equipment had been assembled, the mathematical calculations to convert data into maps defined and the test area chosen. We and our colleagues deployed nine moorings in a 300-kilometer square centered about 600 kilometers southwest of Bermuda. Because this area was thoroughly surveyed by the Mid-Ocean Dynamics Experiment (MODE), the general oceanography of this site was known, permitting accurate a priori modeling and providing a basis for the tomographic results. For further comparison, conventional instruments and survey ships collected data during the experiments.

Our colleague Bruce D. Cornuelle of Scripps extracted images of the 90,000-square-kilometer area from the tomographic data. We checked the validity of the test results by comparing them to the ship surveys, the data from the moored instruments and the statistics obtained by MODE in 1972. Ocean acoustic tomography produced the first snapshots of a volume of ocean this size in less than 24 hours. In contrast, the survey ship, which had to stop for four to five hours every 50 kilometers to obtain readings, took three weeks to construct the picture, which, because of the long "exposure" time, was blurred by the changing mesoscale field.

Further experiments followed. We and our colleagues moored a transceiver array in the Greenland Sea to study the effects of wind on currents and deep-water mixing there. Wunsch and Paola Malanotte-Rizzoli of M.I.T. and their colleagues studied the variability of the Gulf Stream extension in the central North Atlantic. Harry A. DeFerrari of the University of Miami used tomography to measure the vortex energy in the Florida Straits, a narrow region through which most of the Gulf Stream passes.

In addition to creating maps of local currents and eddies, ocean acoustic tomography can be used to measure spatially averaged currents and temperatures over regions extending thousands of kilometers. With Bruce M. Howe of Scripps, we successfully measured 300-kilometer-long averages of ocean temperature and currents in the Atlantic as a function of depth in 1983. Three years later we deployed an array of transceivers in a triangle having 1,000 kilometers to a side to measure long-term spatial averages of temperatures and currents in the North Pacific. Because such long-range study can be accomplished, ocean acoustic tomography in principle should also be able to measure the temperatures of most of the world's waters simultaneously. Such a feat may provide clues to global warming.

Researchers are currently studying ways of conveniently integrating tomographic data with measurements from free-drifting sensors, survey ships and conventional instruments moored at sea. Workers are also trying to eliminate the need for many fixed moorings by using a single "mooring" suspended from a moving ship, hence creating a synthetic tomographic array. We are particularly eager to combine tomographic data with those from the *Ocean Topography Experiment* satellite. Scheduled for launch in 1992, the satellite will obtain data at the sea surface rather than in the ocean interior, as tomography does.

Acoustic tomography has begun to take its place as an instrument in the ocean scientist's tool kit. The original collaboration of scientists has expanded to include researchers from several organizations, including the University of California at Santa Cruz, the University of Washington, the Naval Research Laboratory, the National Oceanic and Atmospheric Administration, research institutes in France, Japan and West Germany, and many private companies. In the decade ahead we expect others to find widespread applications for tomography that will result in important new understanding of the ocean.

FURTHER READING

OCEAN ACOUSTIC TOMOGRAPHY: A SCHEME FOR LARGE SCALE MONITORING. Walter H. Munk and Carl Wunsch in *Deep-Sea Research*, Vol. 26, No. 2A, pages 123–161; February, 1979.

TOMOGRAPHIC MAPS OF THE OCEAN MESOSCALE. PART I: PURE ACOUSTICS. Bruce Cornuelle et al. in *Journal of Physical Oceanography*, Vol. 15, No. 2, pages 133–152; February, 1985.

OCEAN ACOUSTIC TOMOGRAPHY: MESOSCALE VELOCITY. B. M. Howe, P. F. Worcester and R. C. Spindel in *Journal of Geophysical Research*, Vol. 92, No. C4, pages 3785–3805; April, 1987.

OCEAN ACOUSTIC TOMOGRAPHY: A PRIMER. Robert A. Knox in *Oceanic Circulation Models: Combining Data and Dynamics*. Edited by D.L.T. Anderson and J. Willebrand. Kluwer Academic Publishers, 1989.

Boron Neutron Capture Therapy for Cancer

A twofold treatment brings together boron and neutrons to create a lethal radiation that can attack tumor cells without damaging normal tissues in the process

by Rolf F. Barth, Albert H. Soloway and Ralph G. Fairchild

The ideal therapy for cancer would consist of a regimen that kills tumor cells without seriously damaging normal tissues. Complete eradication is essential for cure: unless every cancer cell is destroyed, the danger exists that the tumor might reestablish itself. Although today's standard treatments—surgery, radiation therapy and chemotherapy—have been successful in curing some kinds of cancers, there are many exceptions. The possibility that progress toward another powerful therapy might be made has led the authors of this article and their colleagues at other centers to work on an approach called boron neutron capture therapy (BNCT).

BNCT brings together two components that when kept separate have only minor effects on normal cells. The first component is a stable isotope of boron (boron 10) that can be concentrated in tumor cells. The second is a beam of low-energy neutrons that produces short-range radiation when absorbed, or "captured," by the boron. The combination of these two conditions at the site of a tumor releases intense radiation that can destroy malignant tissues.

BNCT is based on the nuclear reaction that occurs when boron 10, a nonradioactive isotope that makes up approximately 20 percent of naturally occurring boron, is irradiated with and absorbs neutrons. The neutrons that it takes up are called thermal, or slow, neutrons. They are of such low energy (about .025 electron volt) that they cause little tissue damage as compared with other forms of radiation such as protons, gamma rays and fast neutrons. When an atom of boron 10 captures a neutron, an unstable isotope, boron 11, forms. The boron 11 instantly fissions, yielding lithium 7 nuclei and energetic alpha particles. These heavy particles, which carry 2.79 million electron volts of energy, are a highly lethal form of radiation.

If the treatment proceeds as intended, the destructive effects of the capture reaction would occur primarily in those cancer cells that have accumulated boron 10. Normal cells with low concentrations of boron would be spared. One limiting factor, however, is that capture reactions with boron also occur with the nitrogen and hydrogen in normal cells. Fortunately, the neutron-capture cross sections of these atoms, measured in a unit called a barn (10^{-24} square centimeter), are small compared with boron 10, and so the amount of radiation generated would be quite negligible. (The cross section refers to the likelihood of each atom to capture an incoming particle, in this case a neutron.)

Still, normal tissues contain large amounts of nitrogen and hydrogen (3 and 10 percent, respectively, by weight). Therefore, the upper limit of the neutron dose that can be delivered to a tumor is ultimately determined by the tolerance of the surrounding tissue to the radiation (protons and gamma rays) resulting from capture reactions with nitrogen and hydrogen.

A major appeal of BNCT is that the predominant type of radiation produced by the capture reactions is alpha particles. These are nuclei of helium atoms that contain two protons and two neutrons but lack orbiting electrons. Alpha particles move rather slowly, have a pathlength of about one cell diameter (10 microns) and give rise to closely spaced ionizing radiation events. Moreover, it takes only a few alpha particles releasing their energy within a cancer cell to destroy it. These characteristics make BNCT a theoretically ideal system for the destruction of cancer cells.

Alpha particles have other biological advantages. Unlike some forms of ionizing radiation, such as X rays, alpha particles do not require oxygen to enhance their biological effectiveness. A rapidly expanding tumor frequently outgrows its blood supply, so that some regions receive less oxygen than normal tissues do. As a result of this oxygen depletion, the tumor can become more resistant to the effects of conventional radiation therapy. Tumor sensitivity to alpha particles is retained, however, because the particles, not needing oxygen, can attack all parts of a tumor with equal vigor.

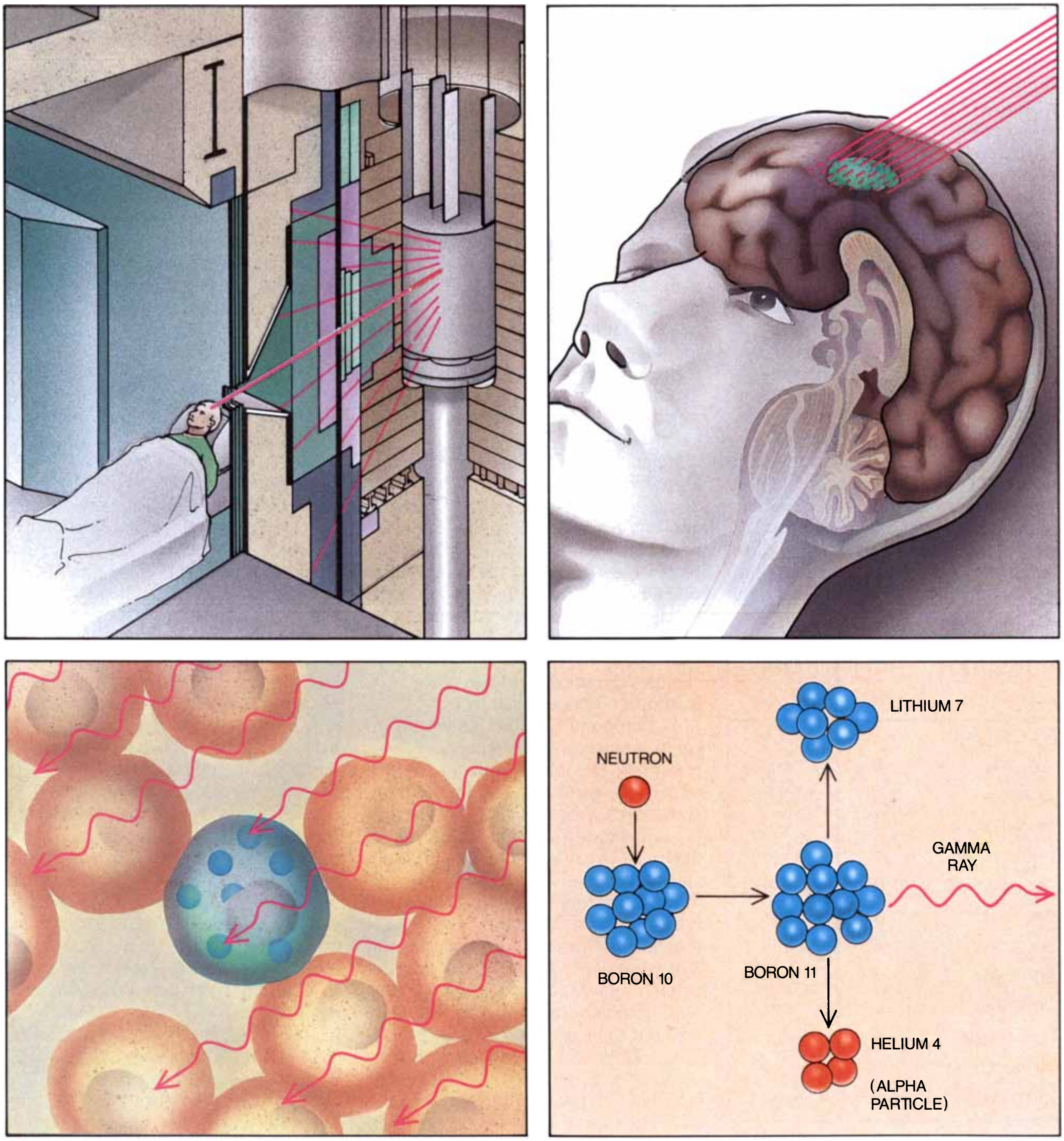
Another advantage of alpha particles is that they can kill dividing and nondividing tumor cells alike—tumors are known to have a large number of viable but inactive cells. Other forms of radiation treatment and chemotherapy tend to work best only on the cells that are dividing.

BNCT is not a new idea. It was first suggested in 1936 by Gordon L. Loch-

ROLF F. BARTH, ALBERT H. SOLOWAY and RALPH G. FAIRCHILD collaborate on a boron neutron capture therapy project funded by the Department of Energy. Barth, professor of pathology at Ohio State University, Columbus, received his medical degree from Columbia University. He trained at the Karolinska Institute in Sweden and at the National Institutes of Health. Soloway is professor of medicinal chemistry at Ohio State University. He received his Ph.D. from the University of Rochester and worked at the Sloan-Kettering Institute. Fairchild works in the medical department at Brookhaven National Laboratory and is professor of radiology and oncology at the State University of New York, Stony Brook. He studied engineering at Cornell University and received his Ph.D. in physics from Adelphi University.

er, then at a division of the Franklin Institute in Swarthmore, Pa. Nothing much came of it until recently because of two major problems: how to deliver a sufficient number of boron 10 atoms selectively to cancer cells and how to enable an adequate number of neutrons to trigger the neutron capture reaction. These problems are a continuing challenge, but researchers have made advances in both areas. Several laboratories in the U.S., Europe, Japan and Australia are now working intensively on their resolution.

Investigators are exploring such potential boron carriers as drugs, monoclonal antibodies and derivatives of naturally occurring compounds. Nuclear reactors and other sources can provide adequate fluxes of neutrons. The radiation field from a reactor is a mixture of gamma rays, fast neutrons, epithermal neutrons (with energies ranging from one to 10,000 electron volts) and thermal neutrons. The challenge for physicists and engineers is to develop neutron beams that con-



BORON NEUTRON CAPTURE THERAPY (BNCT) can be performed at a facility with a nuclear reactor or at hospitals that have developed alternative neutron sources. A beam of epithermal neutrons penetrates the brain tissue, reaching the malignancy. Once there these low-energy neutrons combine with boron 10 (delivered beforehand to the cancer cells by drugs or antibodies) to form boron 11, releasing a lethal radiation (alpha particles) that can kill the tumor.

sist primarily of epithermal and thermal neutrons.

The first clinical studies of BNCT took place in the 1950s and early 1960s. Among the early experimenters were Lee E. Farr, James S. Robertson and Elmer E. Stickley of Brookhaven National Laboratory and William H. Sweet, Gordon L. Brownell and one of us (Soloway) of the Massachusetts General Hospital and the Massachusetts Institute of Technology. Their work focused on treating glioblastoma multiforme, the most malignant of all brain tumors. This cancer was chosen in part because malignant cells that escape surgical removal almost always lead to the formation of new tumors in the surrounding normal brain tissue, resulting in the patient's death.

BNCT treatment failed to prolong the lives of these patients, and in some instances, complications resulted from the irradiation. The experimenters concluded that the boron compounds did not concentrate and persist in the tumor cells and that the thermal neutrons did not penetrate to adequate depths. The result was excessive damage to normal tissue and inadequate destruction of deep-seated tumors.

Soloway and Sweet began a search for boron compounds that would concentrate more selectively in glioblastoma multiforme cells. From among many compounds screened, they chose one (called polyhedral borane anion), which had higher concentrations in the tumor than in the surrounding normal brain. Yet this compound, too, gave disappointing results. All the patients treated in this way showed varying degrees of damage to normal brain tissue, specifically to the walls of blood vessels. The concentration of boron in tumors was still too low when compared with that in normal brain tissue and the blood.

The search for boron compounds that would selectively localize in tumors continued, using mice that were carrying transplantable brain tumors. These studies turned up two promising boron compounds containing sulfhydryl groups that can bind to protein. Their protein-binding characteristics may explain why they concentrate in brain tumors.

Another area of special interest is the development of compounds that might serve in the treatment of malignant melanoma, a pigmented form of skin cancer. The incidence of this tumor and the number of deaths attributable to it have increased tenfold during the past 50 years. The latest American Cancer Society statistics predict 27,600 new cases and 6,300 deaths attributable to melanoma in the U.S. each year.

Melanomas can incorporate elevated amounts of phenylalanine, an essential amino acid, into the pigment melanin and its precursors as well as into proteins. Hence, experimenters have directed their efforts to the synthesis of phenylalanine derivatives that contain boron. The hope is that such fradulent amino acids might "fool" the melanoma cells, which would use them as biochemical building blocks in place of or in addition to phenylalanine.

A compound that has shown promise in animals with melanoma is *p*-boronophenylalanine, a boron analogue of phenylalanine that was first synthesized in 1957 by Harold R. Snyder and his collaborators at the University of Illinois at Urbana-Champaign. The substance accumulates significantly not only in melanomas but also in other types of tumors. Bernard F. Spielvogel together with his associates at Boron Biologicals in Raleigh, N.C., has described other amino acid analogues that contain boron, and they also may turn out to be useful in BNCT.

A related line of attack seeks to exploit the propensity that malignant cells have to take up nucleosides—the building blocks of DNA—faster than normal cells generally do. (A nucleotide is made up of nucleosides.) One key advantage of concentrating the boron in the nucleus of a cell is that such a location may greatly enhance the destructive effects of alpha particles.

Raymond F. Schinazi, now at Emory University, synthesized the first boron-containing nucleoside in collaboration with William H. Prusoff. Other chemists in the U.S. and Japan are formulating related compounds that may concentrate better in DNA and nucleic acids. The hope is that if cell division proceeds at a higher rate in tumors than in the surrounding normal cells, the tumor cells will incorporate larger amounts of nucleosides containing boron and therefore will become more susceptible to the capture reaction.

Two other classes of simple compounds that might be useful as capture agents for BNCT are porphyrins and promazines. Although still in the developmental stages, several compounds have been synthesized and have been shown to be effective in cell culture and animal tumors.

Another approach to the delivery of boron 10 seeks to use antibodies as guided missiles. Antibodies are proteins that have the potential to recognize the unique structures called antigens on the surfaces of tumor cells. Twenty years ago one of us (Soloway) and his colleagues at the Massachusetts General Hospital succeeded in linking compounds containing boron to protein molecules. At that time, the only antibodies available were not specific enough to deliver boron reliably to tumor cells.

Then, about 15 years ago, Cesar Milstein and Heinz Kohler of the University of Cambridge found a way to make highly specific antibodies, which are called monoclonals. They hybridized normal antibody-producing cells with continuously dividing malignant melanoma cells to yield progeny called hybridomas. The hybrids produced monoclonal antibodies that were identical to one another.

Researchers quickly recognized that these monoclonal antibodies might be used to deliver a variety of tumor-killing agents to cancer cells. The same potential would apply to antibodies carrying boron. During the past eight years we and others have developed methods of linking boron compounds to antibody molecules. By linking boron to a polymer of the amino acid lysine, for example, we have attached more than 1,000 boron atoms to a molecule of antibody. For reasons that are as yet unclear, the resulting structures failed to find their antigenic target when injected into tumor-bearing mice. Other approaches, which produce less alteration of the antibody molecule, are currently under investigation.

Antibodies alone probably will not

NUCLIDE		BARNS
BORON 10	^{10}B	3838.00
ELEMENT		
OXYGEN	O	0.0002
CARBON	C	0.0037
MAGNESIUM	Mg	0.069
PHOSPHORUS	P	0.19
HYDROGEN	H	0.332
CALCIUM	Ca	0.44
SULFUR	S	0.52
SODIUM	Na	0.536
NITROGEN	N	1.75
POTASSIUM	K	2.07
IRON	Fe	2.62
CHLORINE	Cl	33.8

ELEMENTS in the body capture neutrons to produce radiation. But this propensity—measured in barns, or 10^{-24} square centimeter—is low in comparison to boron's capacity to absorb neutrons.

be able to deliver the required amount of boron atoms. They most likely will be used in combination with one or another of the boron compounds already known to concentrate in tumor cells. Effective chemotherapy in cancer is based on using a combination of drugs that attack tumor cells in varying ways. Similarly, successful BNCT would probably have to depend on several means of delivering the necessary amount of boron 10 to cancer cells.

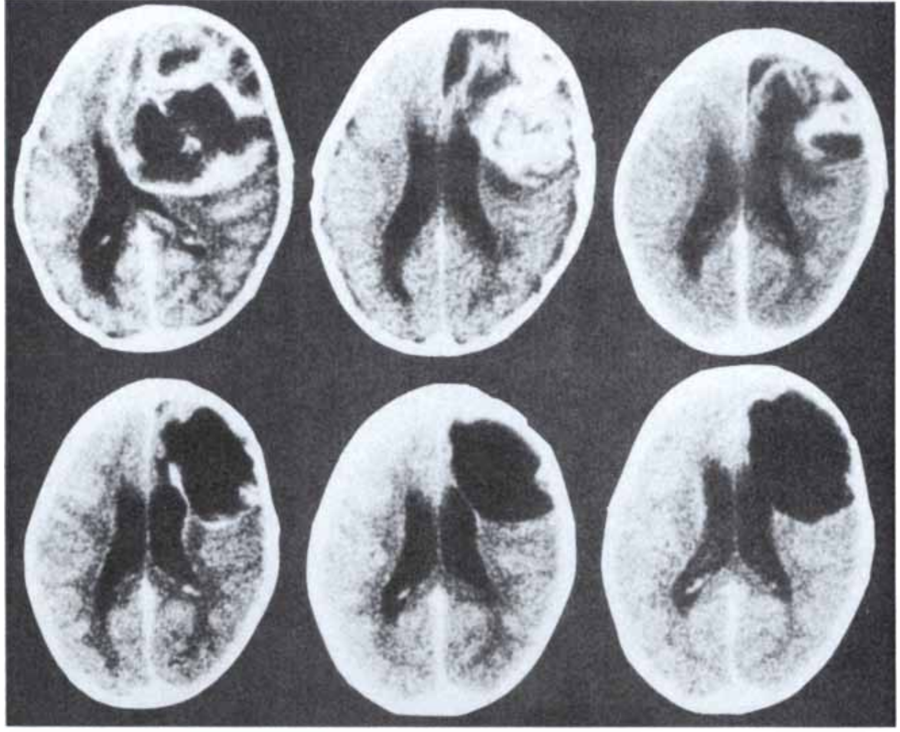
The work did yield an immediate positive benefit, however. Together with Moïse Bendayan at the University of Montreal in Canada, we found that boron can be used in conjunction with a special type of electron microscope, which can detect atoms of elements. Boron is an ideal probe for identifying cell structures and their chemical components.

The requirements for successful BNCT depend on several other factors, including the location of the tumor and its depth in the body, the relative concentrations of boron 10 in the tumor and in normal tissues as well as the number and energy of the neutrons reaching the target site. Calculations have shown a need for a billion boron atoms per cell and a flow of thermal neutrons of from 10^{12} to 10^{13} neutrons per square centimeter. The reason for the large number of boron atoms is the need to ensure that the radiation dose from alpha particles exceeds the dose from the background radiation caused by neutron capture reactions with nitrogen and hydrogen.

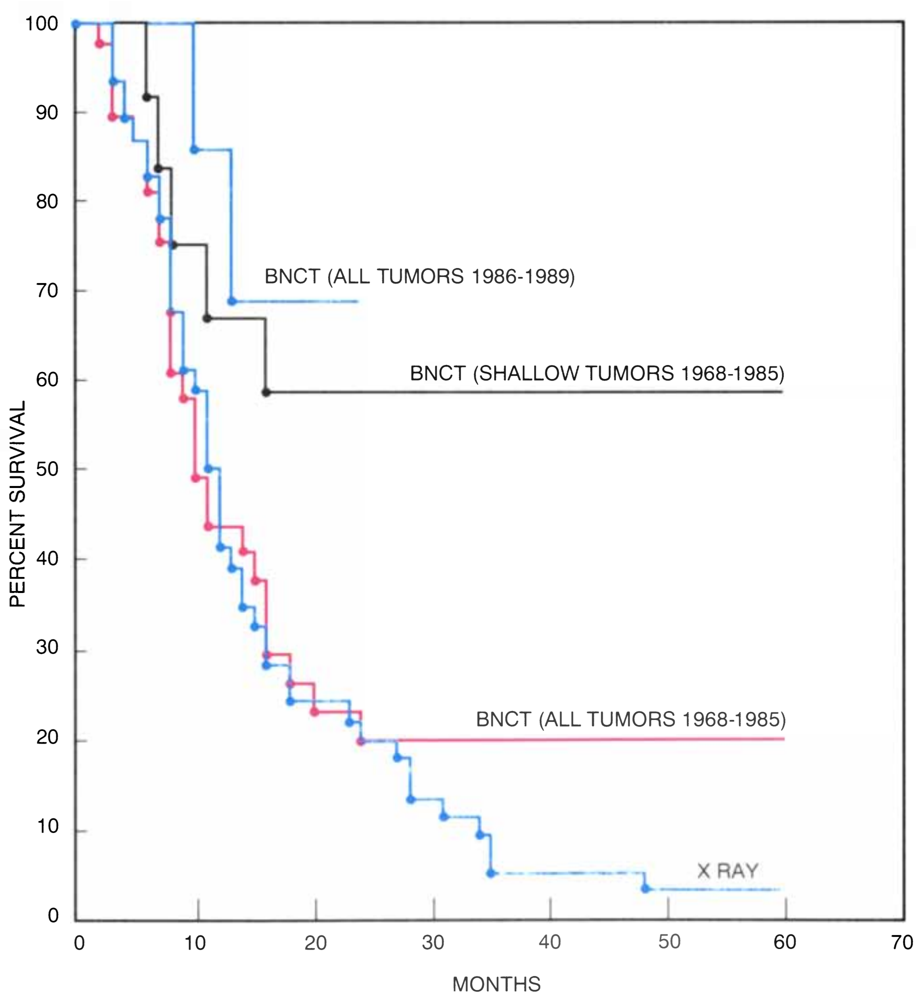
The need to kill enough cancer cells to keep the tumor from regrowing explains the requirement for the intense fluxes of neutrons. BNCT should be able to raise the radiation dose in tumors to levels higher than those achieved with conventional radiation therapy while simultaneously limiting the dose to normal tissues to below their tolerance levels.

The neutrons produced by nuclear reactors are the result of fission. Their initial energy is approximately a million electron volts. They can be slowed to thermal neutron energy by such moderators as heavy water and ordinary water. Nevertheless, thermal neutrons present a few problems. They are rapidly attenuated as they diffuse through tissue. Therefore, it is difficult to get enough of them to the required depth in the body without delivering excessive radiation to surface tissues.

An alternative is to work with beams of epithermal neutrons carrying up to a few thousand electron volts of energy. Such neutrons penetrate better and

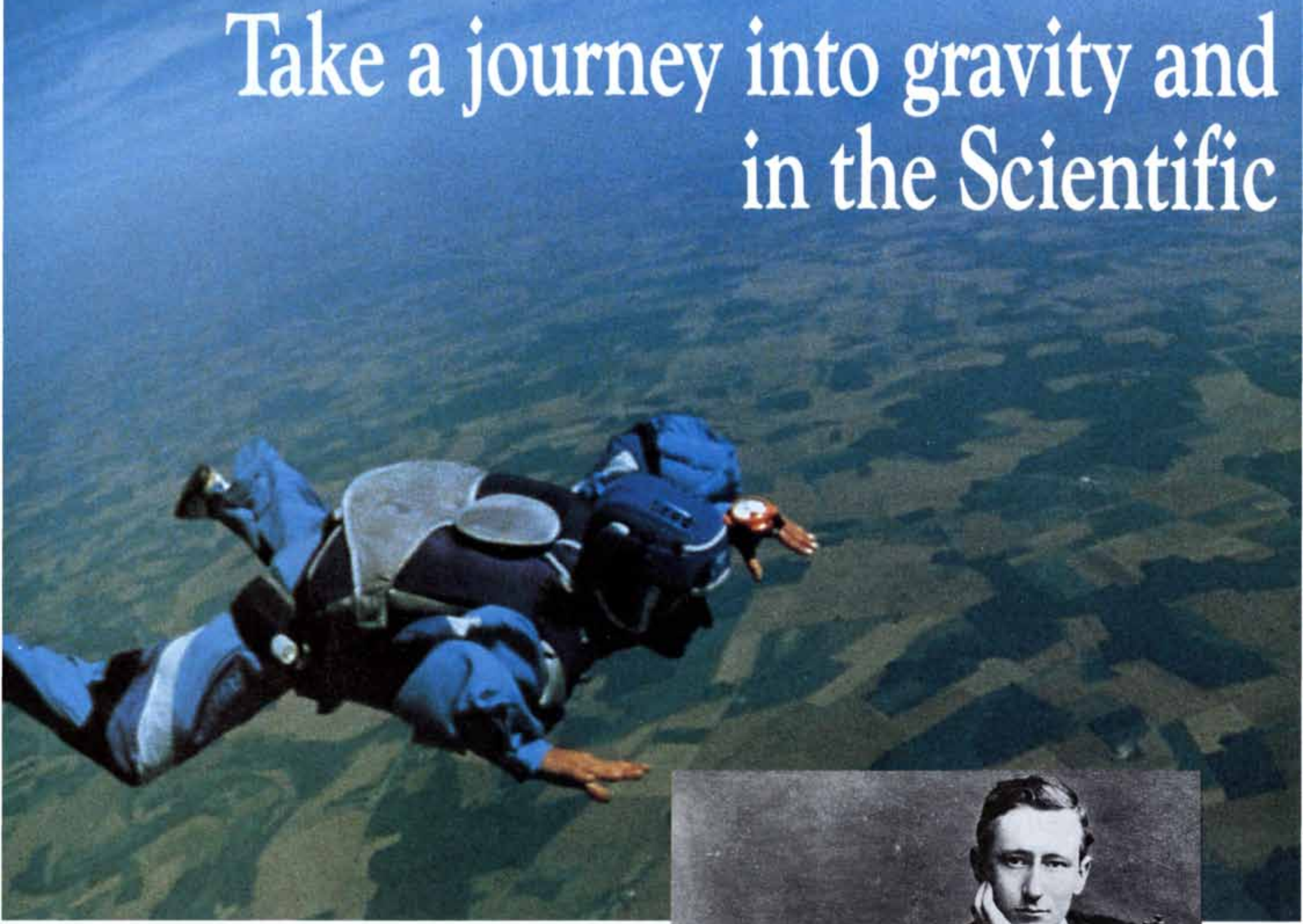


SEQUENCE OF TOMOGRAPHIC SCANS shows a tumor in the right cerebral hemisphere (*top row*). After surgery and BNCT, there is no regrowth (*bottom row*). The procedures were performed by Hiroshi Hatanaka of Teikyo University.



SURVIVAL CURVES summarize Hiroshi Hatanaka's clinical results (1968–89). A combination of surgery and BNCT yields longer survival rates than treatment with X rays.

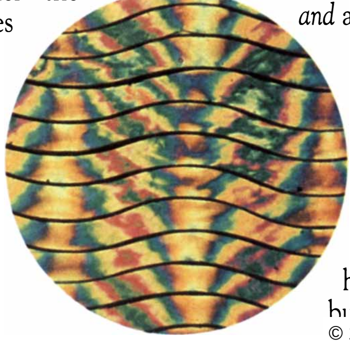
Take a journey into gravity and in the Scientific



Join the Scientific American Library now by previewing *A Journey into Gravity and Spacetime* for 15 days risk-free.

Gravity controls the movement of sky divers, stars, and entire galaxies. It has been studied by the most renowned scientific minds in the world yet it remains a profound mystery of nature.

Now you can see how the latest thinking on this intriguing force is shaping our view of the universe. Join the SCIENTIFIC AMERICAN LIBRARY and receive *A Journey into Gravity and Spacetime* by the preeminent physicist John Archibald Wheeler—the first in a remarkable series of books that will take you on a wondrous journey to the frontiers of scientific discovery.



From The Science of Structures and Materials, *a polarized light view of a stressed graphite fiber-epoxy resin composite. Courtesy Tsu-Wei Chou.*

Written by leading scientists in accessible, compelling prose

SCIENTIFIC AMERICAN LIBRARY books are created in a unique collaboration between our editors and some of the world's most renowned scientists (including many Nobel laureates). The result is a series of books that are readable *and* authoritative.

Each book:

- Covers a topic of major interest with current relevance to ongoing research.
- Presents the firsthand views of a scientist who has made significant contributions to the field
- Delivers clear, compelling prose that even the non-expert will find fascinating.

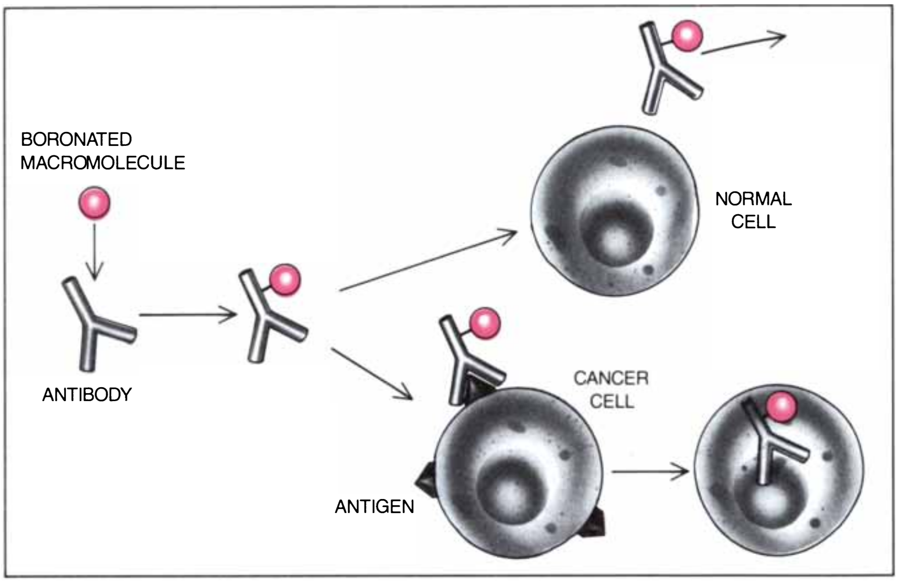
From Signals, *Guglielmo Marconi in 1896. Courtesy Marconi Company Ltd.*

For sheer editorial quality, no other science books can match this fine collection.

Beautiful, richly illustrated volumes that are made to endure

In addition to entertaining and informing, SCIENTIFIC AMERICAN LIBRARY books provide you with an excellent resource for study and research. Each volume features abundant photos, paintings, and illustrations, most in full color. Helpful diagrams, charts, and tables clarify points made in the text.

And SCIENTIFIC AMERICAN LIBRARY books are produced to exacting



ANTIBODIES WORK AS GUIDED MISSILES to carry boron to tumor cells. Other homing devices, such as drugs and naturally occurring compounds, also transport boron.

therefore do not injure normal surface tissues as much as thermal neutrons do. Furthermore, epithermal neutrons lose energy as they pass through tissues and are "thermalized" at depths where tumors tend to grow.

Epithermal neutron beams appropriate for clinical application have been or are being installed in reactors at Brookhaven National Laboratory and M.I.T. In addition, reactors at the University of Missouri at Columbia, the Georgia Institute of Technology and Idaho National Engineering Laboratory are capable of producing effective epithermal beams.

If extensive clinical trials show BNCT to have value, the possibility of developing alternative sources of neutrons would become important because of the high cost and limited numbers of nuclear reactors, together with resistance to the construction of new reactors in major metropolitan areas. Alternative ways of generating neutrons exist. Among them are radioisotopes, proton accelerators and spallation sources. During spallation, neutrons and other light particles from the nuclear reaction are ejected when heavy elements are bombarded by protons.

Among the various radioisotopes that generate neutrons, californium 252 is the most suitable. It produces neutrons by spontaneous fission, with energies similar to those from nuclear reactors. Several grams of the isotope would be needed for beams with enough intensity for cancer therapy. Because californium 252 has a half-life of 2.7 years, it would provide a simple, safe and reliable source of neutrons.

Low-energy proton accelerators could generate epithermal neutrons by irradiating lithium. James W. and Thomas E. Blue (father and son) of the National Aeronautics and Space Administration–Lewis Research Center in Cleveland and of Ohio State University, respectively, have calculated that proton-beam currents approaching two milliamperes would be needed. Such currents are technically feasible but have not been produced so far. Moderation of the beam with heavy water would be needed to obtain suitable numbers of epithermal neutrons.

Spallation in heavy elements such as copper, lead and uranium produces neutrons of various energies. High-energy particles can be moderated to epithermal levels. Measurements made with spallation in copper, however, indicate that the process will take several hours.

We and other workers draw encouragement for continuing our developmental efforts from the work of Hiroshi Hatanaka of Teikyo University. Hatanaka began his clinical tests of BNCT in Japan 22 years ago. He now has treated more than 100 patients afflicted with various types of brain tumors, including the intractable form, glioblastoma. The first step is to remove surgically as much of the tumor as is possible. Following the operation, the patient receives an arterial injection of a polyhedral borane anion having the chemical formula $Na_2B_{12}H_{11}SH$. After the boron reaches the cancer cells, the tumor site is irradiated with a focused beam of thermal neutrons.

Several limitations have reduced the effectiveness of this treatment. One is that the nuclear reactor used as a source of neutrons is of such low power that irradiation may be as long as three to five hours. Moreover, the effective depth of penetration is less than six centimeters, so that the treatment reaches only superficial tumors.

Nevertheless, Hatanaka's results are noteworthy for several reasons. First, only one of 100 patients showed evidence of radiation injury to normal brain tissue and that patient had received an exceptionally large dose of neutrons. Furthermore, among 38 patients who had glioblastoma multiforme and among 12 with more superficial tumors, the mean survival time was significantly greater than that for patients treated with conventional radiation therapy. Several of the patients have survived for years and appear to be cured. Usually the median life expectancy of patients with such tumors is less than one year.

Tests of BNCT on gliomas transplanted into the brains of rats support Hatanaka's clinical data. Perhaps of major significance is the fact that Darrel D. Joel and his colleagues at Brookhaven National Laboratory have investigated a two-part (dimeric) form of the polyhedral borane anion ($Na_2B_{12}H_{11}SH$), which appears to have greater affinity for tumors than the monomer does. They found the dimeric form effective in the treatment of a rat brain tumor. Several groups in the U.S. and Europe have begun to take the first steps toward what may develop into a controlled clinical trial of BNCT for glioblastoma multiforme.

Several recent developments have greatly enhanced the potential of BNCT. New boron compounds are now available that show increased affinity for tumors as well as marked effectiveness in treating animal tumors, and epithermal neutron beams have been developed that permit the treatment of deep-seated tumors.

In the treatment of melanoma by BNCT, Yutaka Mishima and his associates at Kobe University School of Medicine have taken the lead. Working with Duroc pigs, which develop melanomas spontaneously, the investigators injected a boron compound, ^{10}B-BPA (^{10}B-boronophenylalanine), around a skin-level melanoma. They followed the injection with a single exposure to thermal neutrons. Within two months the melanomas began to shrink, and eventually they disappeared.

These results led to a clinical trial now in progress. As of the spring of 1990, eight patients who had skin-level melanomas and who for various reasons were not candidates for surgery have been treated. Mishima has reported that cure of the primary melanomas has been achieved in two patients and the others are currently being closely followed for signs of recurrence.

The challenge is to extend BNCT therapy to melanomas below the skin. Jeffrey A. Coderre and his colleagues at Brookhaven National Laboratory have treated mice and rats bearing transplantable tumors by means of BNCT using ^{10}B-BPA. At the lowest radiation dose, the growth of tumors slowed, while at the highest dose, they either stopped growing or completely regressed.

Other tumors that respond poorly to current therapies eventually may be candidates for BNCT. Indeed, if all the hurdles are surmounted, BNCT could become available at specialized centers with nuclear reactors or at hospitals that have developed alternative neutron sources. For tumors such as the deadly glioblastomas, which cannot be effectively treated by any currently available therapy, BNCT someday might be the treatment of choice. It could also serve as an adjuvant therapy—used together with surgery and conventional chemotherapy or radiation therapy.

Whether BNCT ultimately has a place in cancer treatment depends on research now in progress at a number of centers throughout the world. Whatever the results of these studies may be, it seems likely that binary systems will make an important contribution to the treatment of cancer in the 21st century.

FURTHER READING

BORON-NEUTRON CAPTURE THERAPY FOR TUMORS. Edited by Hiroshi Hatanaka. Nishimura Co., Ltd., 1986.

THIRD INTERNATIONAL SYMPOSIUM ON NEUTRON CAPTURE THERAPY. Edited by Detlef Gabel. *Strahlentherapie und Onkologie,* Vol. 165, No. 2/3, pages 65–257; February/March, 1989.

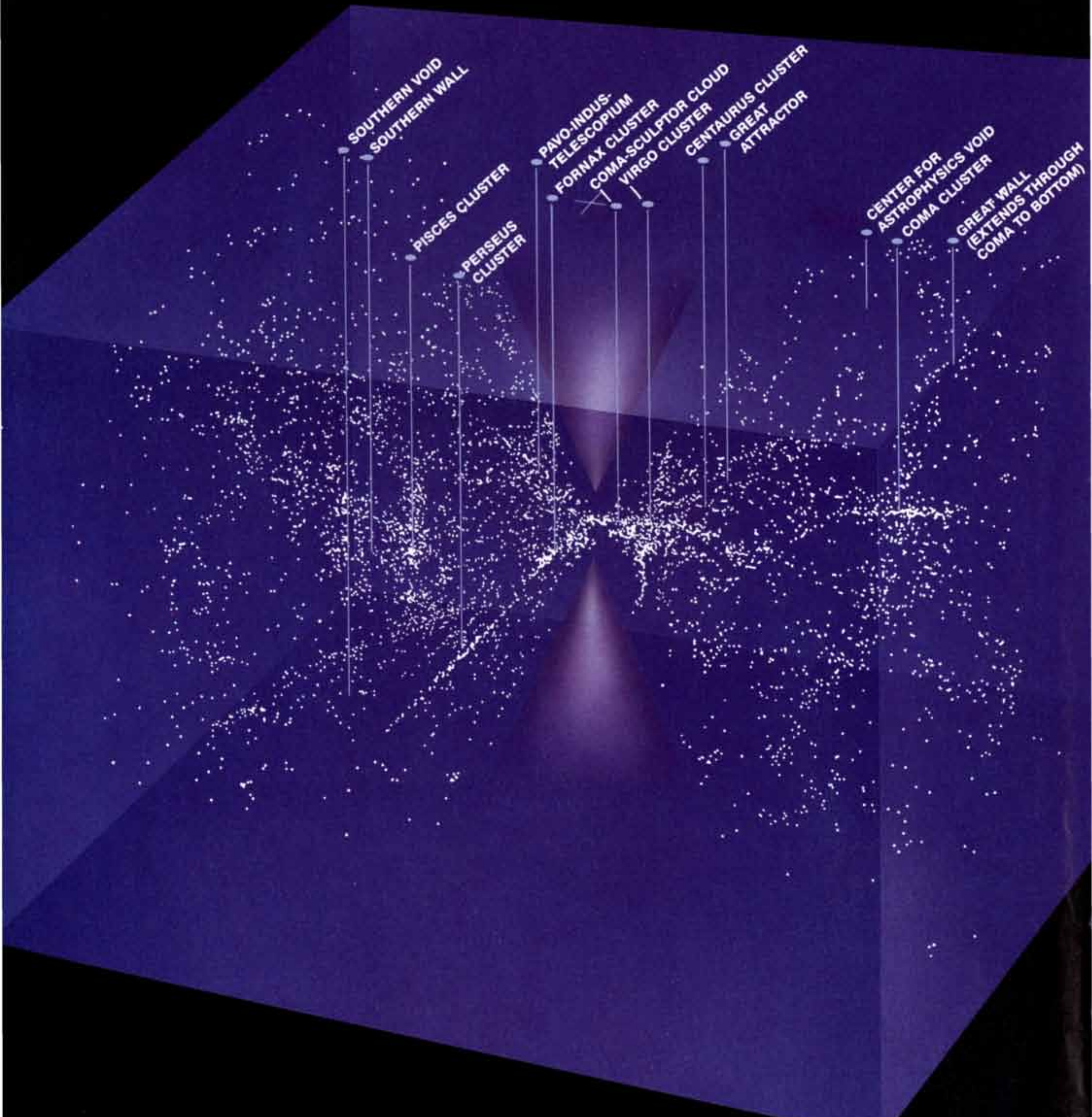
CLINICAL ASPECTS OF NEUTRON CAPTURE THERAPY. Edited by Ralph G. Fairchild, V. P. Bond and A. D. Woodhead. Plenum Press, 1989.

THE SECOND JAPAN-AUSTRALIA INTERNATIONAL WORKSHOP ON THERMAL NEUTRON CAPTURE THERAPY FOR MALIGNANT MELANOMA, Kobe, September 23–26, 1987. Edited by Yutaka Mishima. *Pigment Cell Research,* Vol. 2, No. 4, pages 223–386; 1989.

BORON NEUTRON CAPTURE THERAPY OF CANCER. Rolf F. Barth, Albert H. Soloway and Ralph G. Fairchild in *Cancer Research,* Vol. 50, No. 4, pages 1061–1070; February 15, 1990.

UNIVERSAL TRUTHS

by John Horgan, *staff writer*



The most widely accepted explanation of the cosmos, the cold dark matter model, collides with some cold, hard facts.

THREE-DIMENSIONAL MAP of 6,500 galaxies reveals clusters of galaxies, voids and other large-scale features in the local universe. Approximate depths of notable features have been projected onto the top plane of the map. The Milky Way, which is at the center of the map (below the cross), blocks observations of the shaded conical regions. The map, made for Scientific American *by Changbom Park of Princeton University, is based on data from the Southern Sky redshift survey, the Perseus-Pisces survey and Harvard-Smithsonian Center for Astrophysics surveys. The map shows a region 650 light-years long and wide and 450 light-years deep. A map including the most distant objects ever observed would be about 50 times wider than this one.*

Someone threw sand on the overhead projector. Actually, it is the universe, according to Carlos S. Frenk, the University of Durham, U.K., astronomer who just put the transparency down. To be precise, it is a computer simulation showing how gravity caused matter to condense in the aftermath of the big bang, the cataclysmic explosion that created the universe billions of years ago.

Each grain on Frenk's transparency represents an entire galaxy, itself a mini-universe containing billions of stars. Frenk, a barrel-chested man with a Mexican accent, contends that his simulation bears a close resemblance to real maps of galaxies. "Our model," he says, jutting his chin out defiantly, "can reproduce any of the observed large-scale structures."

If Frenk seems a bit nervous, it may be because his audience includes more than 30 of the world's leading astronomers and physicists. They have assembled at an ascetic lodge in Sweden's rocky, subarctic north for a Nobel symposium on "the birth and early evolution of our universe." For six days in mid-June they will do little but talk cosmology—over dinners of barbecued reindeer, during tramps up a nearby ridge (dubbed the "local maximum" by one physicist) or at the daily scheduled sessions.

Frenk knows that some of these students of the cosmos are growing increasingly skeptical of the theory on which his simulation is based. Called the cold dark matter model, it incorporates the unsettling notion that at least 90 percent of the universe's bulk consists of some invisible type of matter. All the stars and galaxies we see through our telescopes, in other words, are just froth on this dark ocean. The model is the most widely accepted version of how the universe evolved after the big bang; one physicist here calls it the "canon."

But recent observations are squeezing the model from two sides. First, ever more sensitive probes of the so-called cosmic microwave background, a cool (2.7 degrees above absolute zero) bath of microwaves that is thought to be the faint afterglow of the big bang, have yet to reveal any regional variations in intensity. That has forced modelers to assume the early universe was exceptionally smooth, or homogeneous, with matter spread uniformly

SCENES FROM A NOBEL SYMPOSIUM

One awestruck local newspaper called it a meeting of "the greatest geniuses in the world." From June 11 through 16, more than 30 scientists gathered in a no-frills resort in northern Sweden for a Nobel symposium on "the birth and early evolution of our universe." Even geniuses cannot live on cerebration alone. One evening they trooped off to a farm to sample reindeer, moose and bear meat and to watch townspeople perform folk dances and songs. Another evening helicopters whisked everyone up to a mountain lake for lingonberry-punch cocktails. Smaller groups, formed without regard for theoretical compatibility, scaled nearby peaks.

Swedish fiddlers, Stephen Hawking and assistant

Jim

A group of geniuses heads into the woods

John Preskill, Frank Wilczek

through space. At the same time, maps of the universe have revealed ever larger thickets of galaxies surrounded by ever larger voids. If the universe was so smooth early on, how did it come to be so clumpy? Cold dark matter advocates are struggling mightily to answer this question.

One critic of the model is P. James E. Peebles of Princeton University, a dominant figure in cosmology since the mid-1960s. Tall and rail-thin, he perches like a hawk in the front row, eyeing Frenk's simulations. As soon as Frenk heads back toward his seat, Peebles jumps up and flops down his own transparency. It shows a dense band of galaxies surrounded by only a few strays. This is not a simulation, Peebles informs the audience, this is reality: the metropolis of galaxies in which the Milky Way, our galaxy, resides. Peebles circles the big empty spaces above and below the central band. "Where are these voids in your model, Carlos?" he asks, peering through his wire rims at Frenk. "What about these?" Frenk says, striding to the screen and pointing to the stray galaxies. "I prefer to concentrate on this," Peebles says, jabbing the largest void. "You always taught us to be quantitative!" Frenk exclaims. "You should also be reasonable," Peebles retorts, as if chastizing a tardy student.

Cosmology mano a mano, just what physicists at Sweden's Chalmers University of Technology hoped would happen when they organized the symposium for the Nobel Foundation. This is a colorful, contentious bunch. Lectures rarely proceed uninterrupted. As Lennox L. Cowie of the University of Hawaii tries to interpret his latest deep-sky images of galaxies, the audience pelts him with questions. When you say galaxy formation, you mean star formation, right? Why didn't you include helium in that table? Could you quit blocking the screen? Cowie, a big, owl-faced Scot, retreats into self-deprecation. "Any intelligent ideas that I present here are probably stolen from Jim Peebles," he says, his rich burr quavering. "Any totally idiotic ideas are mine."

About the only thing that no one seems to doubt is the underlying premise of mainstream cosmology: the big bang theory. The theory rests on three pillars of evidence. First, there is the discovery by the American astronomer Edwin P. Hubble some 60 years ago that light from galaxies is shifted toward the lower-frequency, or red, end of the spectrum in proportion to their distance from us. Just as the lowered pitch of a police siren indicates that it is moving away from us, so does the "redshift" of the galaxies indicate that they are receding, and the fact that the recessional velocity is proportional to distance confirms that the universe is expanding. (The relation between recessional velocity and distance holds for rubber bands, too. Stretch one and see.) By extrapolating backward from the current rate of expansion, astronomers have inferred that the big bang occurred 13 billion years ago, give or take as many as seven billion years. The imprecision reflects uncertainty over the distances of galaxies, the amount of mass in the universe (which affects the rate of expansion) and other factors.

The second pillar of evidence is the cosmic microwave background, which was discovered in 1965 by Arno A. Penzias and Robert W. Wilson of Bell Telephone Laboratories. Finally, there are the observed abundances of elements in the universe: hydrogen accounts for roughly three quarters of the total mass and helium one quarter, with heavier elements occurring in traces. Various physicists have calculated that a "hot" big bang would forge elements in just such proportions through conventional nuclear fusion, also known as nucleosynthesis.

"The big bang is in fantastic shape," declares David N. Schramm of the University of Chicago, who has played a major role in refining nucleosynthesis calculations. One is inclined to agree with him. Schramm, a.k.a. "Schrammbo," a pilot, mountain climber and former champion in Greco-Roman wrestling, is said by colleagues to be America's second most massive cosmologist (after Geoffrey R. Burbidge of the University of California at Santa Barbara). Schramm is also an ebullient booster of his field. Others—noting the woes of cold dark matter and all other theories of galaxy formation—may

eebles (*left*) and Carlos Frenk

Rocky Kolb (*left*) and Alan Guth

Dave Schramm (*left*) and Swedish hosts

nd Gerard 't Hooft

Andrei Linde

Neil Turok

speak gloomily about the crisis in cosmology, but Schramm calls it "a golden age." "We have the basic framework," he says. "We just need to fill in the gaps."

Some big questions lurk in those gaps, however. Why does the cosmic microwave background appear so smooth? How much matter is there in the universe and what is it made of? How and when did galaxies form? Will the universe expand forever, or will it eventually collapse back on itself?

Theorists have approached these questions in various ways. Some have speculated that galaxies and larger structures were formed by defects in the fabric of space called cosmic strings, or by a similar phenomenon known as textures, or by prodigious explosions in the early universe. Others think the key to many cosmic paradoxes may be a glimmer of energy secreted in empty space: the so-called cosmological constant. Practitioners of a discipline called quantum cosmology seek ultimate answers in the first instant of creation, when all the forces of nature were unified.

Frenk, for one, still thinks cold dark matter represents the most promising approach. When pressed, he concedes that the model has been stretched to the breaking point by recent observations and will—at the very least—probably need revising. But he insists that alternative theories tend to be even more inconsistent with the data or too embryonic to test. The old adage comes to mind: the front-runner is the one with all the arrows in his back.

Right or wrong, the cold dark matter model is an astonishingly bold, imaginative amalgam of ideas from particle physics and general relativity. Perhaps its most crucial component is a concept called inflation. Some astronomers, in fact, refer to the model as the cold-dark-matter-plus-inflation model.

Guth's GUT Feeling

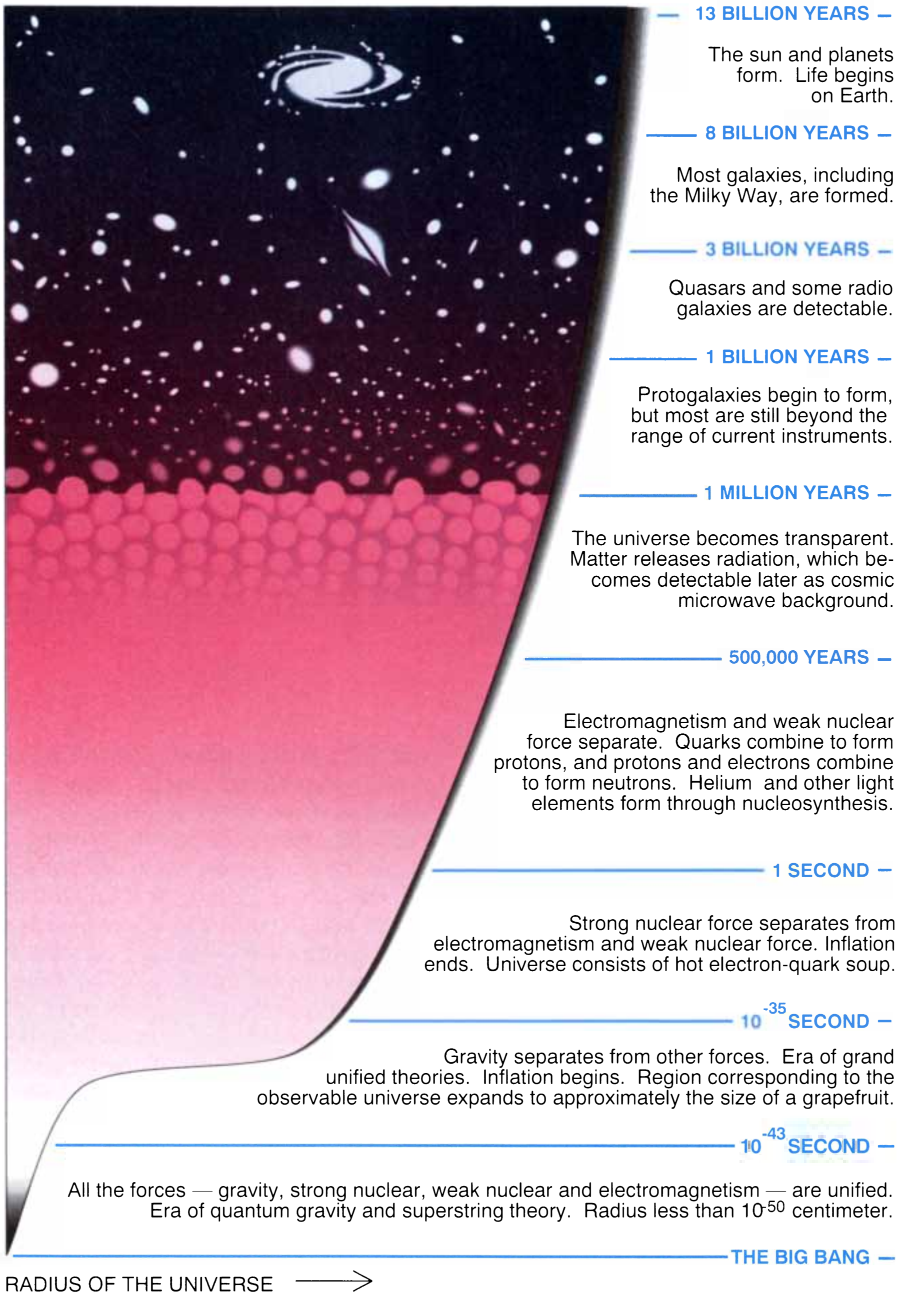
Inflation was first proposed by Alan H. Guth a decade ago. With his round, snub-nosed face and early Beatles haircut, Guth looks more like an undergraduate computer hacker than a tenured professor at the Massachusetts Institute of Technology. At the Nobel symposium, he also tends to doze during other people's talks—conspicuously so, since he sits front row center. One physicist announces proudly early on in his lecture that he has met "the weak Guth criterion: keeping Alan awake for at least five minutes." Guth laughs louder than anyone.

But Guth is a sharp-minded concocter—and marketer—of cosmic ideas. He stumbled on inflation while exploring certain grand unified theories of particle physics. The theories, also called GUTs, hold that the strong nuclear force, the weak nuclear force and electromagnetism—that is, all the forces of nature except gravity—merge into a single force at extremely high energies, or temperatures. These temperatures, which lie far beyond the reach of any earthly experiment, would have prevailed a scant 10^{-35} second after the big bang, during the so-called GUT era.

The GUT era ends when the strong nuclear force separates from the other forces and the universe goes through a "phase transition" akin to the transformation of water into ice. Guth conjectured that just as water can become supercooled, remaining in liquid form below the freezing point, so might certain regions of the universe have become supercooled toward the end of the GUT era. The supercooled regions, he proposed, entered a peculiar state called a false vacuum, in which gravity becomes a repulsive rather than attractive force. As a result, the regions underwent a brief but enormous expansion, or inflation. One region, which started out much smaller than a proton and ended up grapefruit-size an instant later, became our universe.

According to this scenario, the entire observable universe is just a tiny bubble in a vastly greater cosmos, and most of its matter was created virtually out of nothing. Guth pointed out that his theory, in addition to boggling the mind, could resolve some nagging cosmological puzzles. One, sometimes called the homogeneity problem, arises from calculations based on the preinflation big bang model. The calculations suggested that the primordial fireball expanded too fast for all its parts to exchange radiation and so to reach thermal equilibrium. Why, then, some cos-

THE BIG BANG: A CHRONOLOGY



mologists asked, does the big bang's afterglow—the microwave background—have the same temperature no matter where it is measured? Inflation, Guth argued, provides an answer: the region from which the observable universe emerged was so tiny that it had already reached equilibrium.

Inflation also dispatched the so-called flatness problem. The problem stems from Einstein's theory of relativity, which states that matter, through the influence of gravity, causes space to bend. This effect has fateful consequences in an expanding universe. If the amount of matter per unit volume in the universe—that is, its mass density—is above a certain "critical" level, space curves in on itself in such a way that parallel lines converge. Even more significantly, gravity eventually halts the expansion of the universe and causes it to collapse in a "big crunch" (a term Schramm proudly takes credit for). Such a universe is said to be closed. If the mass density falls below the critical level, parallel lines diverge and the universe expands forever; this is known as an open universe. If the density is precisely at the critical level, the universe keeps expanding but at an ever decreasing rate. This is called a flat universe, because space has no curvature: parallel lines behave just as Euclid said they do and as they seem to in our world.

Various cosmologists had argued prior to inflation that the apparent flatness of the universe—and even its very existence—represents an extraordinary coincidence. If the universe had had a bit more matter than it does, it would have collapsed long ago; a bit less matter and it would have flown apart too swiftly for stars and galaxies to form. Lawrence M. Krauss of Yale University has compared the odds of the universe's having just the right density of matter to the odds of someone's guessing *exactly* how many atoms are in the sun. Why were we so lucky?

Again, inflation provided an answer. Just as blowing up a beach ball to 1,000 times its normal size would make its surface appear flatter to a nearby observer, Guth surmised, so would inflation flatten out the region of the universe that we can observe. In such a flat universe, the density of matter would be set at the critical level.

Conundrums such as the homogeneity and flatness problems have led some cosmologists to invoke the anthropic principle: the universe is the way it is because if it were not, we would not be here to observe it. But most scientists find the anthropic principle tautological and unscientific. "When you invoke the anthropic principle," Schramm says, "you're saying you don't know how to solve that problem." Little wonder, then, that many theorists embraced inflation as the most important development since the original big bang theory. Not only did it solve the horizon and flatness problems, it also made specific predictions about the amount of matter in the universe and its ultimate fate.

Inflation even provided an explanation for the formation of galaxies. According to quantum physics, any energy field constantly fluctuates in intensity at the subatomic level, just as waves dance on the surface of a lake. The peaks created by these quantum fluctuations in the very early universe would have become large enough, after inflation, to serve as the seeds of stars and galaxies. Gravity would do the rest of the work. "A lot of people think inflation just smells too good not to be right in some form," Schramm says.

Although inflation resolved some very big questions, it raised another: Where is all the matter? Guth's theory predicts that the density of matter in the universe should be precisely at the critical level, but all the measurements made so far have come up short.

Surveys of visible, luminous matter yield a mass density equal to only 1 percent of the critical level. Studies of the motions of stars within individual galaxies have pushed estimates of the total density higher. The observations indicate that the galaxies must contain about 10 times more mass than they appear to or they would fly apart. Analyses of the motions of galaxies orbiting within clusters of galaxies have led to the same conclusion.

These findings, first made decades ago, initiated the search for the "missing matter." But the revised estimates of the masses of galaxies and clusters still pushed the total mass density of the universe up to only 10 percent or so of the critical level. Moreover, a key assumption of the big bang nucleosynthesis calculations, which so successfully account for the proportions of light elements in the universe, is that baryons—a class of particles that includes neutrons and protons and accounts for most of the ordinary matter of which stars, planets and people are made—*cannot exceed* 10 percent of the critical density. More baryons would have resulted in more helium.

The implication of this assertion is profound. Astronomers originally suspected that dark matter consists entirely of ordinary baryonic matter packaged in difficult-to-detect forms: comets, planets, dim stars known as brown dwarfs, black holes and even entire galaxies that are too faint to see. But if both inflation and the nucleosynthesis calculations are correct, then at least 90 percent of the matter in the universe must consist of so-called nonbaryonic particles. Moreover, the particles must interact very weakly with ordinary matter or they would not be so dark, or hard to detect.

The leading dark matter contender early on was a "hot," or fast-moving, nonbaryonic particle: the neutrino. Neutrinos, although they are produced in abundance in various nuclear reactions, are quite weakly interacting; they can zip through the earth unhindered. Best of all, they are known to exist: physicists first detected one of the three known types of neutrino in 1956.

Neutrino-dominated, or "hot dark matter," models have two problems, however. First, years of experiments have failed to prove that neutrinos have any mass, let alone enough to brake the expansion of the universe. Preliminary data from a sensitive new neutrino detector called the Soviet-American Gallium Experiment recently raised the hopes of hot dark matter enthusiasts that this problem will be eliminated; the data suggest that the sun is emitting far fewer neutrinos than the standard model of particle physics predicts, and a way to resolve the discrepancy is to require that one type of neutrino have mass. But a second problem remains unsolved: according to computer simulations, the fast-moving neutrinos would have taken too long to settle down into galaxies.

The shortcomings of hot dark matter models led theorists to focus on slow-moving, or cold, dark matter, which presumably could clump together more readily. Peebles wrote the first paper on cold dark matter plus inflation in 1982 but later became a critic, he says, because the model had "too many problems." Frenk says Peebles enjoys proposing theories and then criticizing them after they become popular. "That's a myth," Peebles responds. "But like all myths," he adds, "there may be a little truth in it."

The Gang of Four

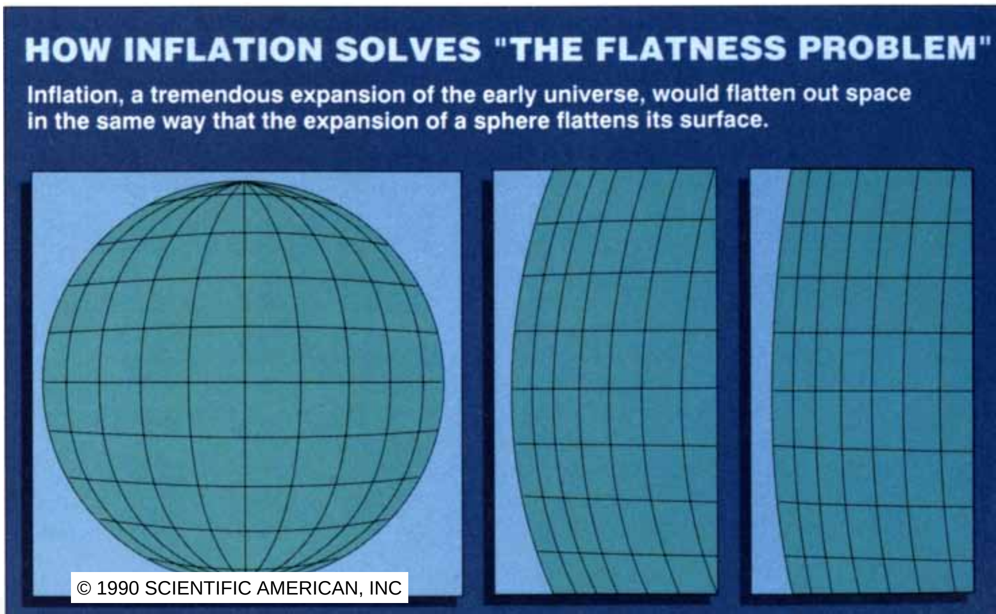
In the mid-1980s Frenk, Marc Davis of the University of California at Berkeley, Simon D. M. White of the University of Arizona and George Efstathiou of the University of Oxford began performing computer simulations based on the cold dark matter model that seemed to reproduce current observations of the distribution of galaxies remarkably well. The so-called gang of four became cold dark matter's chief advocates, and with their help the theory became the leading explanation of galaxy formation.

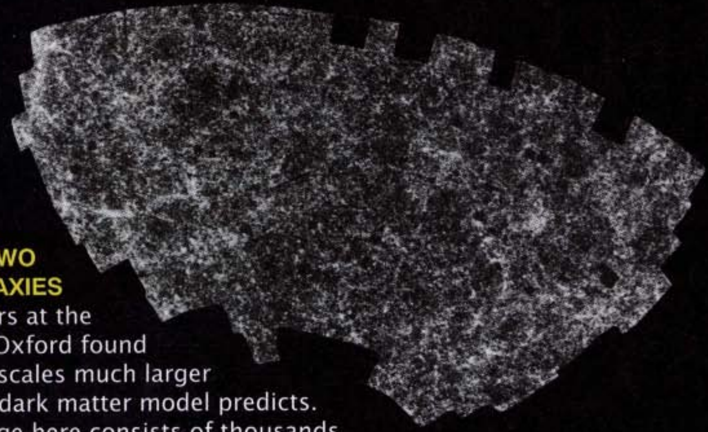
The fact that no stable cold dark particles had ever been detected was seen as a challenge rather than a fatal flaw. After all, particle physicists, in their efforts to devise a grand unified theory, had invented a bevy of hypothetical particles to mediate the interactions between forces. One of the particles, the cosmologists speculated, could still exist as a "relic" from the GUT era in abundances great enough to account for the missing mass.

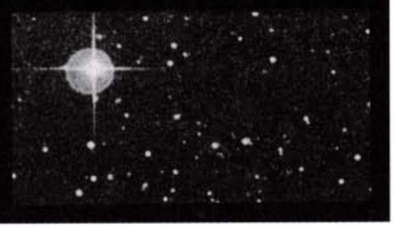
In Sweden, Michael S. Turner of the University of Chicago, renowned for his artistic transparencies, nominates "100-kilogram cosmologist relics" as candidates; his transparency displays oblongs labeled "Dave" and "Rocky," the nicknames of his Chicago colleagues Schramm and Edward W. Kolb. Two candidates considered somewhat more plausible are axions and light supersymmetric particles, or LSPs. Axions are relatively light particles invented to resolve problems in certain GUTs. LSPS are the offspring of a theory called supersymmetry, which holds that all the particles identified so far have weakly interacting twins.

Particle physicists have eagerly joined forces with cosmologists in trying to identify a cold dark matter candidate. John Ellis of CERN, the European laboratory for particle physics, says finding such a particle would represent a satisfying extension of the Copernican revolution. "Not only are we not at the center of the universe," remarks Ellis, whose shaggy beard and staring eyes suggest a disregard for earthly concerns, "we may not even be made of the same stuff as most of the universe." Yet experiments at CERN and elsewhere have already ruled out some candidates (such as a cold, very massive neutrino) and have indicated that others—such as LSPs—may inhabit energy levels beyond the reach of today's accelerators, according to Krauss.

The cold dark matter model has other problems. Ironically, one major source of trouble is the cosmic microwave background, which has heretofore given so much comfort to cosmologists. Its uniformity in all directions, or isotropy, has always been a strong point—evidence of the radiation's primeval origin. But the cold dark matter model predicts that the fluctuations responsible for galaxies and larger struc-







SURVEY OF TWO MILLION GALAXIES by astronomers at the University of Oxford found clustering on scales much larger than the cold dark matter model predicts. The large image here consists of thousands of pixels, each of which corresponds to a photograph of a different patch of sky. Pixels are black, gray or white depending on whether the plate contains no galaxies, from one to 20 galaxies or more than 20 galaxies, respectively.

tures should have left an imprint—in the form of local variations in temperature—on the microwave radiation.

That imprint has yet to be found. Measurements by a satellite called the *Cosmic Background Explorer* (COBE), which was launched last year, have probed the radiation down to a resolution of one part in 10,000 without finding any local variations in intensity. Those data are already on the verge of violating the cold dark matter model, and George F. Smoot of the University of California at Berkeley, a member of the COBE team, hints that new data will show the radiation to be smooth at even finer resolutions. Measurements of the microwave background by an observatory at the South Pole, which scans a smaller region of the sky than COBE does but can achieve higher resolutions, have also failed to find any inhomogeneities.

At the same time, a growing number of observations have shown that the distribution of galaxies is anything but smooth. In the mid-1980s a group led by Margaret J. Geller and John P. Huchra of the Harvard-Smithsonian Center for Astrophysics began mapping the positions of galaxies by analyzing their redshifts—which Hubble had shown are proportional to distance—and found that they tended to cluster along gigantic filaments and sheets. Another Harvard-Smithsonian team found evidence of a huge empty region in the direction of the Bootes constellation, one of many such voids now identified. A group known as the seven samurai (astronomers favor these monikers over lawyerly lists of names) reported that the Milky Way and many neighboring galaxies seem to be falling toward some huge, unidentified mass: the Great Attractor.

Two reports in the past year have exacerbated the situation. Geller and Huchra claimed to have discovered the largest structure yet identified: the Great Wall, a band of galaxies at least 500 million light-years across. Thomas J. Broadhurst of the University of London and his colleagues probed two narrow "pencil beams" of space and found dense globs of galaxies alternating with voids some 400 million light-years across out to a distance of several billion light-years. Geller, among others, contends that neither cold dark matter nor any other model can explain the new observations: there simply has not been enough time for gravity to create structures as large as the Great Wall out of the small density fluctuations that are consistent with the microwave background.

Neither Geller's group nor Broadhurst's is represented at the meeting in Sweden, and their reports take a beating. When Schramm reviews the pencil-beam results, he is constantly interrupted by objections. "I'm just presenting the data!" he exclaims. "I'm not interpreting it!" Frenk calls the Great Wall and pencil-beam data "rubbish." "It's not a statistical analysis," he says. "It's just anecdotal." Even Peebles, no fan of cold dark matter, warns that "one should beware of Great Walls. The eye is a great recognizer of patterns."

Participants of the conference seem more impressed with a survey done by astronomers at Oxford. The group did not plot the distances of individual galaxies based on their redshifts, but it compensated for this shortcoming by analyzing many more galaxies. Whereas the previous surveys mapped only a few thousand galaxies in relatively narrow regions, the Oxford group scanned two million galaxies across a broad stretch of sky.

"We found evidence of clustering on scales at least twice as large as those predicted by the cold dark matter model," Efstathiou, one of the participants in the survey, tells the group. The cold dark matter model must be either revised, he adds, or scrapped. The audience seems stunned: the slim, elegantly dressed Efstathiou exudes a cool rationality; he is also one of the cold dark matter model's original gang of four. "It's like the Pope becoming Jewish," one astronomer remarks.

Another set of observations—those showing objects that are very distant and hence very old—is also causing concern. The cold dark matter model predicts that most galaxies take at least several billion years to form, so few should be found at distances greater than 10 billion light-years (assuming that the universe is 13 billion years old). But hundreds of quasars, extremely bright points of light and radio waves, have been found beyond that limit. Many astronomers are convinced that these odd objects are galaxies with hyperactive cores.

Because quasars are still poorly understood, greater weight has been given to objects that are clearly galaxies. Astronomers have now identified a score of galaxies more than 10 billion light-years away; a half dozen of these are more than 11 billion light-years away. Most are radio galaxies: large elliptical galaxies with gigantic, radio wave–emitting "jets." Cold dark matter cannot be ruled out yet on the basis of these observations, says Kenneth C. Chambers of Leiden Observatory, Netherlands, who is studying distant galaxies, "but if we find many more of these, the model is in big trouble."

Things That Go Bang

Even before the recent wave of observations, theorists had begun offering alternative mechanisms for creating galaxies and organizing them into larger structures. So far none has usurped the role of front-runner from the cold dark matter model. One hypothesis, developed by Alexander Vilenkin of Tufts University and others, is cosmic strings. Cosmic strings have been likened to the crystalline defects that form between regions of a pond that have frozen at different times. In the same way, so might defects have formed in the topology of space as the early universe cooled. Such defects, theorists calculated, could have been massive enough to form galaxies and even clusters of galaxies. But recent work on strings suggests that they may have been too small and unstable to create galaxies. "The picture is not as nice as it was," says Thomas W. B. Kibble of the University of London, a pioneer in the study of topological defects.

A similar fate has befallen proposals by Jeremiah P. Ostriker of Princeton and others involving explosions in the early universe. Exploding stars, or supernovas, could have spurred galaxy formation, according to Ostriker and his colleagues, and explosions of cosmic strings could have pushed the galaxies into clusters. "Jerry likes things that go bang," Frenk remarks. But many theorists, including Ostriker, think that explosions large enough to create huge clusters and voids should have left a detectable imprint on the cosmic microwave background.

There is one alternative theory of structure formation whose popularity is still ascending. Called textures, it has been developed over the past two years by a young Princeton professor and former cosmic string enthusiast, Neil G. Turok, and others. Textures resemble cosmic strings, except that they are knotlike, localized defects—and they are much easier to model on a computer. "Anyone who wants my string codes can have them," Turok tells his colleagues with a rueful smile. Kibble finds textures "rather attractive," but he warns that cosmic strings and explosions also looked promising early on. "The harder you look," he says, "the more problems you find."

If the cold dark matter model has fared better than most other models, that is because its advocates have made it more flexible. They assume, for example, that galaxies present a biased, or inaccurate, picture of the distribution of dark matter, and so they rearrange the matter at will. The problem with this idea, called biasing, is that it adds "free parameters" (variables that can be arbitrarily adjusted) to the model. Free parameters make it easier to match observations, Peebles says, but they also make the model "less predictive and satisfying."

Theorists have also proposed modifications to inflation that could give rise to larger-scale structure. Alexei A. Starobinsky of the L. D. Landau Institute of Theoretical Physics in Moscow has suggested that inflation may occur in several stages, or "steps," and Paul J. Steinhardt of the University of Pennsylvania has advanced a version, called extended inflation, in which the strength of gravity varies.

Perhaps the most inventive inflationary theorist is Andrei D. Linde, a Soviet physicist who went to CERN two years ago and to Stanford University this fall. Linde is something of a showman. At an outdoor cocktail party in Sweden, he does a backflip, karate chops a rock in half and performs magic tricks. Balancing a match on his seemingly motionless finger, he makes it quiver and hop as though jerked by an invisible string. When his colleagues demand to know the secret, he replies in a deep Russian growl: "Is quantum fluctuations."

Linde is also known for using sleight of hand to keep inflation viable. He has been closely associated with the hypothesis since 1982, when he adroitly solved theoretical problems arising in Guth's original formulation. Since then, Linde has proposed countless new forms of inflation—incorporating axions, cosmic strings, neutrinos, textures and other hypothetical concepts. Unlike most other cosmologists, moreover, he is not afraid of invoking the anthropic principle. "There are many strange possibilities," he says.

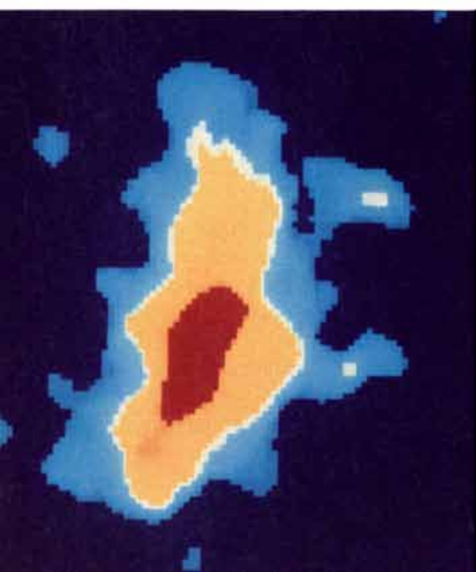
Yet some cosmologists have become increasingly frustrated by the mutability of inflation. During a talk on extended inflation, Peebles looks more and more uncomfortable and finally blurts out: "But you'll get a ring of black holes around your membrane!" Later he broadens his criticism. "Is the 'robustness' of inflation a good thing?" he asks. "Here it's been used to mean that it can fit any data presented." Frank Wilczek, a particle physicist at the Institute for Advanced Study in Princeton, comments: "We need more specific, testable models of inflation. We can't let Andrei go on like this."

Einstein's Greatest Blunder

Those betting on the outcome of all this theoretical gaming must keep in mind a wild card known as the cosmological constant. The term, which was invented by Einstein early in this century, refers to the energy that inhabits a perfect vacuum. Einstein had speculated that vacuum energy, which would create a kind of outward "pressure" on space, might be needed to counteract the inward pull of gravity in a static, or nonexpanding, universe. Einstein abandoned the idea after the discovery that the universe is expanding—he is even said to have considered it his "greatest blunder"—but it keeps rearing its head in cosmology.

Cowie, for example, says a cosmological constant could help to explain deep-space images he has recently made of galaxies. These images suggest that galaxies were several times as numerous about five billion years ago as they are in the present epoch. What happened to all the galaxies?

Cowie notes that many of them could have merged or become too dim to see, but he prefers another explanation: the cosmological constant. A slight vacuum energy would distort our view of

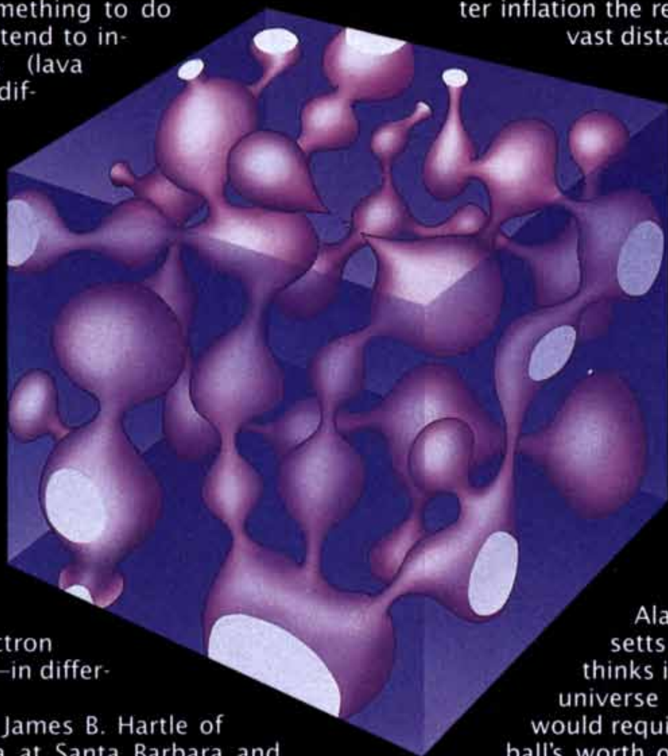


IMAGES OF DISTANT GALAXIES CHALLENGE THEORISTS

The most distant galaxy identified so far was discovered by Kenneth C. Chambers and George K. Miley of the Leiden Observatory (*left*). It is almost 12 billion light-years away. Theorists have a hard time explaining how galaxies formed so soon after the big bang. An image made by Lennox L. Cowie of the University of Hawaii and others showing a field of galaxies about five billion light-years away poses a different puzzle (*right*). The galaxies seem to be much more numerous than in nearer regions. What ex

Cosmologists are always putting us in our place. Not content to tell us that our sun is just one of billions of stars in the Milky Way, which itself is one of billions of galaxies in the universe, some cosmologists would also have us believe that our entire universe is only one in an infinitude of cosmos.

There are a number of multiverse theories. Although they all have something to do with quantum physics and tend to inspire similar illustrations (lava lamps come to mind), they differ in significant ways.



Many-worlds theory. Hugh Everett III of Princeton University proposed this theory more than 30 years ago to resolve questions concerning the role of the observer in quantum physics. According to quantum physics, a particle such as an electron seems to follow a number of paths simultaneously, yet when a physicist observes the electron he finds it following only one path. Bothered by the arbitrary way the alternate paths are eliminated, Everett proposed that the electron actually follows every path—in different universes.

Many-histories theory. James B. Hartle of the University of California at Santa Barbara and Murray Gell-Mann of the California Institute of Technology have extended Everett's concept to the entire universe. Immediately after the big bang, Hartle explains, the universe was so small that it could be considered as a subatomic particle following different paths. Hartle prefers the term "histories" to "worlds," and, unlike Everett, he thinks of the alternate paths as "potentialities" rather than realities.

Chaotic inflation. Soviet physicist Andrei D. Linde speculates that when the universe was about 10^{-35} second old, it was a chaotic froth; different regions had different physical properties and underwent inflation, a brief but tremendous growth spurt, at different times. After inflation the regions are separated by such vast distances that they are largely beyond one another's influence. In effect, they are separate cosmos.

Wormholes. Stephen W. Hawking of the University of Cambridge, Sidney R. Coleman of Harvard University and others suggest that just as electrons can suddenly "tunnel" from one spot to another, so can space-time itself. The tunneling of space-time creates wormholes, which can lead either to other points in the same universe or to cul-de-sacs called baby universes or to other universes as large as our own.

Do-it-yourself universe. Alan H. Guth of the Massachusetts Institute of Technology thinks it may be possible to create a universe in the laboratory. The task would require little more than a bowling ball's worth of matter, he says. The hard part would be compressing the matter to black hole–like densities and then somehow getting it to expand—just as the universe did when the big bang took place. "I like to think of that as an engineering problem," Guth remarks, "possibly solvable by some future civilization."

space, causing us to underestimate its volume on very large scales. This effect, Cowie says, could explain why the galaxies in his deep-space images seem so closely packed together.

In a similar way, the constant could eliminate an embarrassing discrepancy between the age of globular clusters (dense knots of stars hovering near the center of the Milky Way) and the age of the universe. Spectral analysis of the stars in the globular clusters implies that they are more than 15 billion years old, whereas estimates of the distances of galaxies and their rate of recession have led many astronomers to believe that the universe is only 13 billion years old. A slight vacuum energy would lead astronomers to underestimate the total volume and so the age of the universe.

Finally, a cosmological constant could explain why the universe appears so flat even though the apparent density of mass falls far short of the critical level. Since energy and mass are in a sense equivalent, vacuum energy could serve as the missing mass. "If you were bewildered by Len Cowie's talk," Peebles remarks to his colleagues in Sweden, "you were paying close attention."

The cosmological constant is not terribly popular among theorists, however. The most common complaint is that it is "ugly": it introduces yet another free parameter into models, making them both less predictive and less elegant. Moreover, no one has suggested why nature would prefer a positive cosmological constant, let alone one with a particular value.

In fact, calculations by Sidney R. Coleman, a physicist at Harvard University, have suggested that the cosmological constant should be exactly zero. His explanation is a bit complicated. Just as subatomic particles can instantaneously jump from one spot to another through the quantum effect known as tunneling, so can space and time. The tunneling of space-time, according to Coleman, creates "wormholes" that link our universe to other universes in such a way that the vacuum energy vanishes.

Quantum Cosmology

Coleman's pronouncement on the cosmological constant is one of the more concrete results emerging from quantum cosmology. This exotic branch of physics addresses the first 10^{-43} second of creation, when the universe was so small, dense and hot that gravity and quantum effects—which ordinarily belong to separate realms—would have been deeply entangled. Even grand unified theories stop short of this era, but some physicists hope to illuminate it by developing a quantum gravity theory—also called the theory of everything—uniting all the forces of nature.

Discussions of quantum cosmology have a science-fictionish ring to them, and they are viewed with both skepticism and confusion by more empirically minded scientists. "All the astronomers can fall asleep now," Coleman says at the beginning of his talk on wormholes. "But Alan has to stay awake." Yet practitioners such as John P. Preskill of the California Institute of Technology think their efforts could lead to the breakthrough that cosmology increasingly seems to need. "Eventually these ideas may help us understand large-scale structure," he says.

One of the most adventurous explorers of quantum cosmology is Stephen W. Hawking of the University of Cambridge. Stricken by a degenerative nerve disease more than two decades ago, Hawking must breathe through a tube in his throat, and he can move only two fingers of his left hand. That is enough. Slumped in his wheelchair, he presses a button cradled in his palm and constructs sentences on a computer screen propped up in front of him. A speech synthesizer utters the words for him—"Integrating wormholes leads to alpha parameters that are a quantum field on superspace"—in an incongruously deep, authoritative voice.

Hawking, whose work served as a launching pad for Coleman's assault on the cosmological constant, thinks wormholes may eventually lead to a quantum gravity theory. Others are doubtful. Gerard 't Hooft of the University of Utrecht, Netherlands, thinks wormholes violate principles of cause and effect that are fundamental to physics. "I'd much rather see a theory forbidding wormholes," he says.

Good-bye, Go-Go Junk-Bond Days

Another approach to quantum gravity involves superstrings. Not to be confused with cosmic strings, superstrings are hypothetical particles that have length but no width and that wiggle in as many as 26 dimensions. Proponents say these wiggles can account for the behavior of all observed particles. But other physicists contend that the highly abstruse, mathematical theory must prove itself at lower energies—in the realm attainable by particle accelerators—before it is accepted as a viable description of the universe's creation. "String theory is still promising," Wilczek of the Institute for Advanced Study says dryly, "and promising, and promising." That is true as well of all quantum cosmology, he adds. "These are healthy attempts, but they may all be wrong."

Turner, the University of Chicago physicist, applauds these theoretical efforts but predicts that the future of cosmology will be dominated by observations. The "go-go junk-bond days of cosmology," in which theoretical speculation is unchecked by data, are over, he says. Yet he voices hope that theories will serve to guide experiments. "If all you have are observations, that's botany," he says. "If all you have is theory, that's philosophy."

Turner points out that in spite of the impression conveyed by all the reports on large-scale structure lately, the vast bulk of the observable universe remains uncharted. That situation should be redressed somewhat by a joint Chicago-Princeton project to map the positions of a million galaxies and several hundred thousand quasars by the end of the century. A few years from now the *Hubble Space Telescope*, if it is fixed, could reveal the structure of distant galaxies and quasars in detail, and a radio telescope now being built in India could look even further back, detecting the embryos of galaxies.

Other experiments could demystify dark matter. A group led by J. Anthony Tyson of AT&T Bell Laboratories is searching deep space for objects whose images have been distorted—bent into arcs or rings or split into multiple images—by gravitational lenses made of dark matter. Pierre Sikivie of the University of Florida and others want to convert a huge magnet at Lawrence Livermore National Laboratory—built for a fusion experiment that was canceled—into an axion detector. Laboratories in the U.S. and Europe are designing instruments for detecting light supersymmetric particles.

In the near term, Turner says, the most crucial test of all the models will come from the cosmic microwave background. If no fluctuations are found soon by COBE or the South Pole observatory, he says, "that would shake the foundations of all cosmology."

What would happen then? Could the big bang itself be overturned? Although no one at the Nobel symposium engages in big bang bashing, scientists elsewhere are less restrained. Hannes Alfven, a Swedish Nobel laureate, has attracted attention to a model that dispenses with the big bang and posits that electromagnetism and not gravity is the dominant shaper of matter in the universe. Halton C. Arp, an astronomer at Germany's Max Planck Institute for Astrophysics, has argued that the redshifts of galaxies may not derive from the expansion of the universe.

Schramm, the big bang's biggest defender, insists the theory will survive these attacks—even if no fluctuations are found in the cosmic microwave background. Although that eventuality would doom most mainstream explanations of galaxy formation, including cold dark matter, Schramm says it could favor "more exotic and bizarre" models. One such model, he notes, holds that the universe acquired the seeds of large-scale structure during a phase transition that occurred very late, after the microwave background was released.

This kind of talk concerns Turok, the young textures advocate. He fears that as observations reveal the universe in ever greater detail, cosmologists may resort to ever more complicated, jury-rigged models—like the one that Ptolemy devised to show how the sun and planets revolve around the earth. In that event, Turok says, he may abandon cosmology for another field—condensed-matter physics, perhaps. "Maybe the problems cosmology has set for itself will turn out to be just too difficult to solve scientifically," he says. "After all, we've got a lot of gall to suppose that the universe can be described by some simple theory."

Peebles, when told of Turok's concerns, smiles. He recalls that when he first looked into cosmology in the late 1950s he was "repelled" by its lack of substance. "Since then the field has gotten richer and richer," he says, "and I see no reason why that shouldn't continue." Peebles views the ongoing surge in observations as a source of joy, not despair. The observations may eliminate many theories, but eventually, given the abundance of talent and creativity in cosmology, they will spawn new theories with far more explanatory power. In the mean time cosmologists—and the rest of us—may have to forgo attempts at understanding the universe and simply marvel at its infinite complexity and strangeness.

FURTHER READING

THE INFLATIONARY UNIVERSE. Alan H. Guth and Paul J. Steinhardt in *Scientific American*, Vol. 250, No. 5, pages 116–128; May, 1984.

PARTICLE ACCELERATORS TEST COSMOLOGICAL THEORY. David N. Schramm and Gary Steigman in *Scientific American*, Vol. 258, No. 6, pages 66–72; June, 1988.

THE COSMIC BACKGROUND EXPLORER. Samuel Gulkis, Philip M. Lubin, Stephan S. Meyer and Robert F. Silverberg in *Scientific American*, Vol. 262, No. 1, pages 132–139; January, 1990.

THE FIFTH ESSENCE: THE SEARCH FOR DARK MATTER IN THE UNIVERSE. Lawrence M. Krauss. Basic Books, 1990.

ORIGINS: THE LIVES AND WORLDS OF MODERN COSMOLOGISTS. Alan Lightman and Roberta Brawer. Harvard University Press, 1990.

Drug burst, new old TV, compacting rad waste

Gigabit Connection

Scientists want a network, but technical troubles loom

Julian G. Rosenman wants a more accurate way to aim beams of radiation at tumors. He would like a three-dimensional image of the dose distribution throughout the body that can be modeled within seconds, letting him raise radiation levels enough to increase remissons for certain types of cancer.

The problem is that Rosenman, an associate professor of radiation oncology at the University of North Carolina School of Medicine in Chapel Hill, needs to send information from a Cray supercomputer that can calculate the dose levels to a specialized graphics computer that can immediately produce a three-dimensional model. Unfortunately, this must happen at speeds well beyond the capacity of existing computer data networks.

Rosenman is one of many researchers who are looking expectantly toward efforts by the federal government and industry to build high-speed fiber-optic networks that could combine the computing power of distant computers. That would create, in effect, a coast-to-coast computing laboratory.

BUILDING A GIGABIT NETWORK

*Research programs coordinated by the Corporation for National Research Initiatives**

AURORA	**Goal:** Designing switches and communications protocols; developing applications for video conferencing, networked virtual reality (under consideration) and distributed data-base applications **Proposed participants:** Bell Atlantic, Bell Communications Research (Bellcore), IBM, Massachusetts Institute of Technology, MCI Communications, Nynex and University of Pennsylvania
BLANCA	**Goal:** Developing switches and communications protocols and network applications for atmospheric storm modeling, radio astronomy imaging and multimedia digital libraries **Proposed participants:** AT&T, Ameritech, Astronautics, Bell Atlantic, Cray Research, Lawrence Berkeley Laboratory, National Center for Supercomputing Applications, Norlight, Pacific Telesis, University of California at Berkeley, University of Illinois and University of Wisconsin
CASA	**Goal:** Creating hardware interfaces to connect computers to the network and programming techniques for physically separated computers; developing applications for geophysical, global climate and chemical-reaction modeling **Proposed participants:** California Institute of Technology, Jet Propulsion Laboratory, Los Alamos National Laboratory, MCI Communications, Pacific Telesis, San Diego Supercomputer Center and U.S. West
NECTAR	**Goal:** Revamping computer operating systems, hardware interfaces and other systems software needed to tie computers into a high-speed network **Proposed participants:** Bell Atlantic (Bell of Pennsylvania), Bellcore, Carnegie-Mellon University and Pittsburgh Supercomputing Center
VISTANET	**Goal:** Designing switches; imaging for radiation therapy **Proposed participants:** Bell South, GTE, MCNC and University of North Carolina

*Funding from a $15.8-million award from the National Science Foundation and the Defense Advanced Research Projects Agency combined with an estimated $100 million or more from industry.

Source: Corporation for National Research Initiatives

Legislative proposals to create such a network have been snaking their way through Congress, mirroring a plan by the White House's Office of Science and Technology Policy. The plan would involve spending up to $2 billion over five years for supercomputers and a network to tie them together. A fifth of the total appropriation would go toward a National Research and Education Network (NREN), an optical fiber superhighway that could transfer more than a billion bits of data each second—an encyclopedia's worth of text.

Meanwhile the National Science Foundation and the Defense Advanced Research Projects Agency (DARPA) are spending $15.8 million, coupled with what is estimated to be more than $100 million in contributions from private industry, to pursue advances in network technologies.

"One of the assumptions is that a gigabit network is fundamentally important to the research community, but we need to get people to use it to prove that is true," says Robert E. Kahn, president of the Corporation for National Research Initiatives in Reston, Va. Kahn's organization is coordinating the efforts of academic, industry and government researchers to establish five test programs that will develop a technological base for such a network.

A major challenge is to come up with a more powerful version of switches used to route digitized data through long-distance computer networks. That technology breaks up a computer message into a series of "packets" that are routed by the switches through the network and then reassembled before they are received by another computer. A separate, dedicated circuit is not needed for each message, because packets flow across the network mixed together in one continuous stream.

Packet switches generally operate too slowly to send video images and voice conversations across a network. Data streaming into a present-day packet switch are stored in its computerized memory. The switch's software then determines the best path for the packet to move through the network before sending it on its way. "If you do all those steps in a high-speed network, you get flooded," says David J. Farber, professor of computer and information science at the University of Pennsylvania and one of the founders of the NREN concept.

For that reason, the National Science

Foundation Network, the main artery for a web of existing packet networks used by researchers and universities, is used primarily for the exchange of electronic mail and computer files. To get more speed, a research effort known as Aurora is currently evaluating two high-performance packet switches developed separately by IBM and Bell Communications Research (Bellcore), the research arm of the local telephone companies.

The IBM switch prototype, called Paris, tries to eliminate processing delays in the switch by placing all the routing information in the packet itself. The device takes in a packet and channels it to one of eight ports, each of which is connected to a line running at a gigabit per second. Bellcore tries to achieve higher speeds by designing customized switching hardware. Its switch, known as Sunshine, is built around a single integrated-circuit chip, a "switching fabric," whose 32 input lines can connect to 32 output lines. Each line operates at speeds of 150 million bits per second. Up to four lines can be combined and switched together.

Switch design is only one of many engineering hurdles. The type of hardware needed to connect a computer to the network has yet to be determined. Software must be redesigned to eliminate having both the sending and receiving computers acknowledge that a packet was properly transmitted. This "I'm OK, you're OK" exchange can bog down the network and may prove unnecessary because optical transmission is so efficient. Programmers must also figure out how to keep the flood of incoming data from overwhelming the operating-system software that controls computer programs.

Even with these problems resolved, it takes 10 milliseconds for a packet to get from the California Institute of Technology to the Los Alamos National Laboratory at the speed of light, enough time to perform 10 million arithmetic operations on a supercomputer.

Casa, another of the five projects, aims to develop new algorithms to hide these delays. Borrowing a technique used for the internal design of a computer, Cal Tech will "pipeline" a string of linear equations through a fiber link to a supercomputer at the Los Alamos National Laboratory. By the time the computer has finished solving one equation—say, a chemical-reaction model—the next one will be waiting. "It's like an automobile assembly line," says Paul Messina, director of the Cal Tech Concurrent Supercomputing Facility. "It takes a day or so to get going, but once it's up and running, it can produce one car a second."

The ultimate goal of a gigabit network, however, is still nothing short of a fundamentally new way of doing scientific research. The Blanca project, for example, will try to prove how researchers spread across the country can jointly work on atmospheric and radio astronomy models.

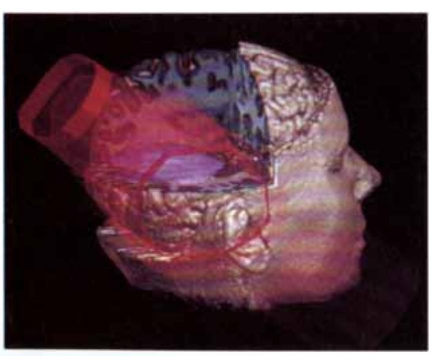
In the works is a three-dimensional model of a thunderstorm that could be computed at the University of Illinois' National Center for Supercomputing Applications while being manipulated on screens at the University of California at Berkeley and the University of Wisconsin at Madison. "It is going to be like a bunch of people all huddled around a computer screen," says Charles E. Catlett, the center's manager of networking and systems development. "A guy in Berkeley will be able to release a weather balloon into the clouds while a colleague in Madison follows its path."

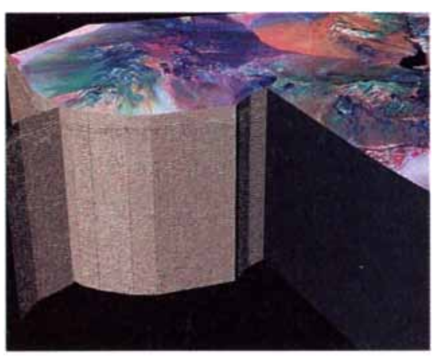
In addition to a cross-university electronic laboratory, some researchers want to apportion processing tasks among different machines, letting each one do what it does best. At Carnegie-Mellon University, computer scientists and chemical engineers are collaborating as part of the Nectar project to create a mathematical simulation of a chemical-processing plant.

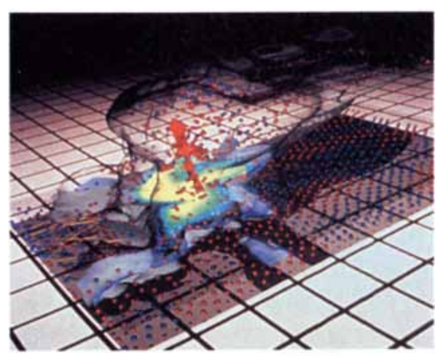
A Cray supercomputer cannot handle the millions of equations rapidly enough to model a plant's operation faster than the process itself, thereby allowing engineers to make changes that would improve yields and cut processing time while the operation is in progress. To overcome that problem, the supercomputer's more flexible processing abilities would be used to calculate chemical concentrations and other data that would then be sent over the network to serve as variables for systems of linear equations. The equations would be computed on specialized parallel processors, whose performance is comparable to the Cray's for highly structured work.

Passing data from one computer to another may be easier than figuring out how to pay for a network. Legislative and business foes of industrial policy have dubbed the overall congressional initiative a "Cray welfare bill" and insist on a purely commercial network in which universities use their research funding to buy the lowest-cost service.

But advocates of a stronger government role argue that such a network is critical to continued U.S. technological prowess. "If you want just electronic mail, then giving people money to buy services and letting the commercial marketplace handle this may achieve your objective," says Frederick W. Weingarten, who directed a computing and networking study for the congressional Office of Technology Assessment. "But if you want a very innovative service where the market is risky, and which requires R&D and significant innovation, more government intervention may be needed." —*Gary Stix*







HIGH-SPEED COMPUTER NETWORKS will combine the power of supercomputers and specialized processors to produce three-dimensional images that can be used to direct radiation beams for cancer therapy **(left),** ***model geophysical data from several large data bases*** **(center),** ***or let researchers spread across the country manipulate models of thunderstorms*** **(right).**

On Again, Off Again

Custom drug-delivery systems emulate the body's own patterns

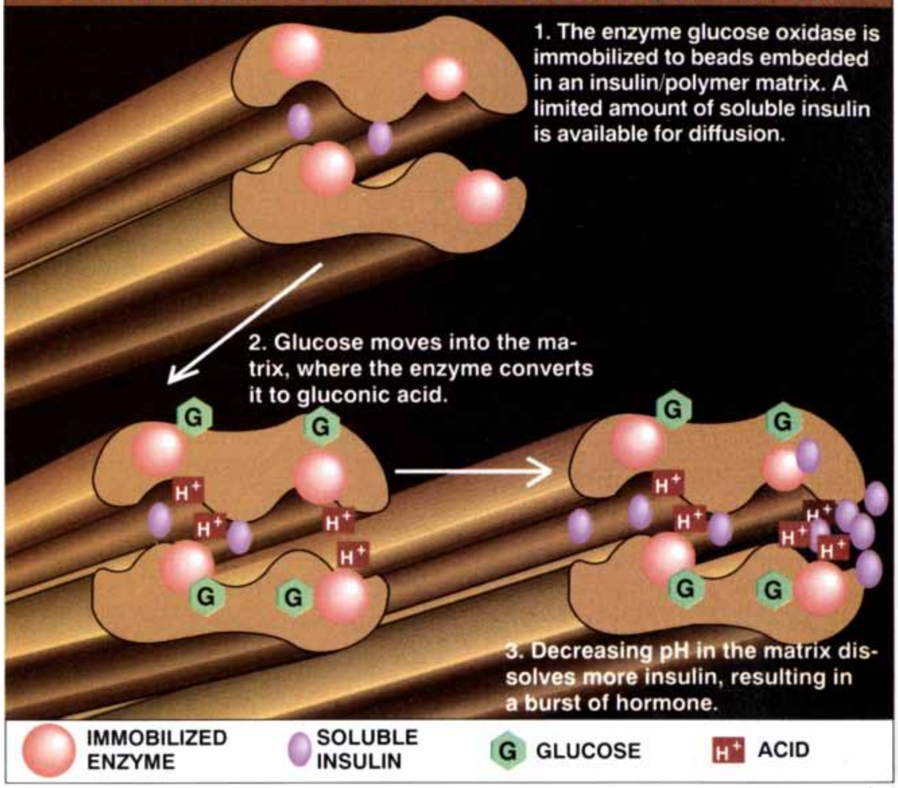
"If you want to grow up to be big and strong...go to sleep." This could be the most justifiable bedtime admonition ever. For it is primarily at night, during sleep, that a young person's brain releases growth hormone in short bursts every 90 minutes or so. Each several-second burst is terminated by an opposing hormone in a perpetual feedback loop that works like a thermostat. Other pituitary hormones, such as those controlling fertility, are also released in pulses. Insulin, although not under pituitary control, follows a similar on-off pattern, revving up only at mealtimes.

Yet current methods of therapeutically replacing these hormones in people who lack them bear little resemblance to the way the healthy body does it. Injections, even when taken several times a day as diabetics do, simply dump a lot of drug at once, with the expectation that enough medication will be circulating when the body needs it. Portable pumps that administer drug through a catheter operate at strictly set rather than responsive intervals or require the patient to calculate when and how much to release. The few long-acting drug formulations that exist do spare the patient frequent injections but do not interact in a natural manner.

Finding a way to mimic the body's pulsatile patterns is high on the agenda of drug companies and researchers. They think that emulating natural rhythms will enhance the effectiveness of hormones, as well as other proteins and even classical chemical compounds. Blood pressure medications, for instance, are good candidates. They are less important when a person is resting or asleep than during a stressful work day. Pulsed doses might also mitigate the tolerance problems some patients develop when taking cardiac and anti-inflammatory drugs for chronic conditions.

The pulsatile-delivery strategies exciting the most enthusiasm now are polymers that release drugs in response to signals within the body. Robert Langer, professor of biochemical engineering at the Massachusetts Institute of Technology, pioneered the controlled release of proteins from polymers in the 1970s. More recently his laboratory has developed a variety of polymers, called polyanhydrides, that erode in the body like a bar of soap. The material has been exploited by licensees such as Nova

A PULSATILE POLYMERIC DRUG-DELIVERY SYSTEM



Pharmaceuticals in Baltimore. The firm embeds wafers of the polymer mixed with an anticancer drug in the brain after local tumor surgery.

Those wafers, however, provide a slow continuous release of the drug. Getting polymers to release drug in pulses that respond to the body is a lot trickier. One solution is to make them react to a metabolic trigger, so that an elevated concentration of a certain metabolite in the blood releases a burst of drug. Langer co-developed such a system for delivering insulin to diabetics in collaboration with Larry Brown, now principal investigator at Enzytech, a young drug-delivery company in Cambridge, Mass.

Enzytech's approach is to use glucose as a trigger. It works like this: after a meal, when a diabetic digests food, the body's glucose level rises. The sugar diffuses into an implanted polymer slab impregnated with insulin. Also embedded in the slab are polymer beads containing an enzyme that converts glucose into gluconic acid. The resulting change in acidity makes the insulin more soluble, so it diffuses out of the slab and into the bloodstream.

The insulin-glucose interaction works for another polymeric approach, based on a phenomenon called competitive binding. Sung Wan Kim, director of drug-delivery research at the Center for Controlled Chemical Delivery, University of Utah, attaches a chemical tail to insulin that causes it to bind to a protein called concanavalin A. Glucose has an even greater affinity for binding to the same protein. Kim puts the insulin-concanavalin complex into a polymer membrane pouch and implants this in the abdominal cavity. As glucose seeps in, the insulin is dislodged.

"If we have the right metabolite, we could do this for other diseases," Kim says. "We don't always have to use laboratory materials. We could use what's in the body." Low-density lipoproteins, for example, could nudge drugs for hyperlipidemia into the bloodstream, he suggests. Other triggers might work for epilepsy or schizophrenia medications.

Many drugs have no known biological trigger or preferred binding material but could still be fitted with internally regulated delivery systems. Temperature-sensitive polymers that degrade in a feverish person might do the job. A five-degree Celsius change, from 25 to 30 degrees C, can turn a "concrete wall of polymer into spaghetti," Kim explains, allowing quick release. Unfortunately, a five-degree change amounts to a lethal fever, so the method is far from ready to go.

A process as basic as osmotic pressure can make oral pulsatile delivery possible for certain drugs, says Jane E. Shaw, president of Alza Corporation in Palo Alto, Calif. Her firm has already used osmosis—water molecules' habit of moving into areas of high ion concentration—to create continuous-release forms of drugs such as appetite suppressants. In Alza's Oros system, a drug is coated with a semipermeable membrane, and a hole is drilled through this coating with a laser. Water enters the tablet through the membrane and dissolves the drug, which is forced out through the hole. Shaw asserts that layering drug in the tablet's core between sections of inert material, perhaps with an expanding polymer to give an extra push, could produce bursts of drug.

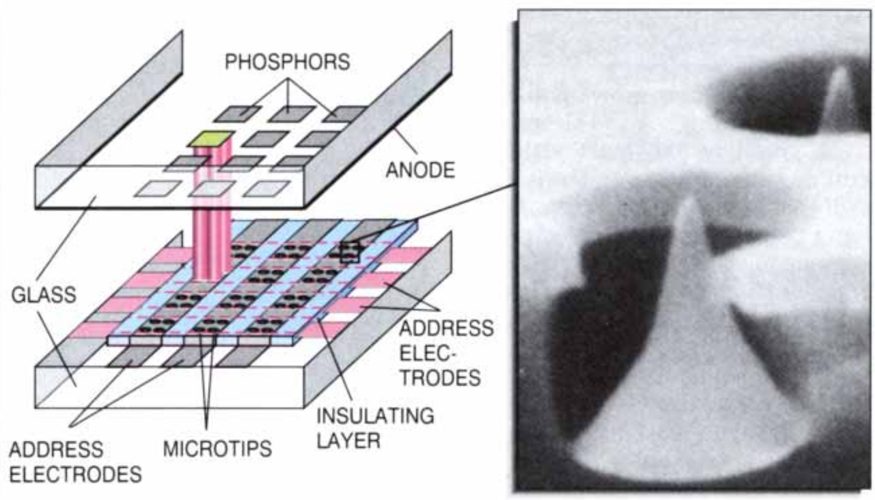
"Optimizing delivery is definitely the way of the future. It would be of value for any disease state," Shaw declares. Yet would-be developers of pulsatile-delivery systems face particular challenges. Chief among them is ensuring stability of the drugs, so that implants are long-lived. Easier for chemicals, stability is a difficult thing to ask of proteins, which are large, fragile molecules. The most stable form of insulin, for instance, lasts about three months, although medical researchers expect they will eventually be able to make it last a year and more.

Workers are looking at other, external triggers to degrade polymers, including magnetism, ultrasound and light. These forces could activate not only implantable systems but also injectable drugs encapsulated in polymer microspheres. Pulsatile, implantable pumps may get a boost from new biological sensors. "But sensors and motors take up volume," points out Ronald A. Siegel, a biochemical engineer at the University of California at San Francisco. He is working on an implantable insulin pump that uses the same acid reaction as Enzytech's polymer system, not electric current, to force insulin into the bloodstream.

"Implantable, injectable... these all sound mighty invasive," says Gary Cleary, chairman of Cygnus Research, Redwood City, Calif., a company devoted to delivering drugs through the skin. Pulsatile transdermal delivery could be achieved, he suggests, by rhythmic application of a weak electric field.

Proving the advantage and safety of new product formulations to the U.S. Food and Drug Administration has never been easy. The task stands to become even more challenging, given the recent installment of an expert group on pharmacokinetics at the agency's Center for Drugs. The panel is expected to increase pressure on companies to evaluate how the timing of administration and fluctuating concentrations of drug in the blood influence a medication's performance. For the developers of pulsatile drug-delivery systems, that's a healthy decision. —*Deborah Erickson*

VACUUM MICROELECTRONICS DISPLAY



Cold Cathodes

Vacuum microelectronics enter the flat-display race

Television designers have long sought a replacement for the bulky cathode-ray tube. Their goal is to create thin, lightweight screens that would preserve the intrinsic brightness and color produced by the tube, a technology nearly 100 years old. Flat-display panels using liquid crystals, electroluminescence and gas discharge have all come forth as prospective contenders. Now another technology has thrown its pixels into the ring. It is none other than the old picture tube itself. This time around, however, it is a flat panel built of millions of micron-size cathodes.

The idea of miniaturizing tubes and other vacuum devices so that they would consume less energy and occupy less space is more than 20 years old. But the integrated circuit got there first. The research effort nonetheless continued, largely to create military electronics that would be immune to bursts of electromagnetic radiation. Researchers have now borrowed the technology used to print circuit lines on silicon semiconductors to make arrays of millions of tiny cathodes called field emitters.

The first display has been demonstrated by a team at the Laboratory of Electronics, Technology and Instrumentation (LETI), part of the French atomic energy commission. "It's a major, major advance," says Capp Spindt, director of the vacuum microelectronics program at SRI International in Menlo Park, Calif., and the developer of much of the basic technology adapted by the French laboratory.

If commercialized, a flat-panel vacuum microelectronic display could be fitted into portable computer, aircraft and automobile instrumentation—and it would be ideal for large high-definition television screens.

Still, the French researchers have a long way to go before their technology can compete with a $40 cathode-ray tube. Their 4.3-by-3.5-inch prototype display uses vacuum microelectronics to illuminate a grid of only 256-by-256 picture elements, or pixels, on a glass substrate.

The possibility of winning funding for further development work was behind a decision by the French team to describe its effort at the Institute of Electrical and Electronics Engineers' Third International Vacuum Microelectronics conference in July. Their group had lost its research support from the French-based Thomson Consumer Electronics. Thomson decided that the project was too embryonic to be worthwhile, says Jacques C. Duchene, LETI's project director for flat-panel development. "They said, 'We can't wait 10 years.'"

Still, the LETI group caused a stir at the IEEE meeting with a slide of a black-and-white video image of Mary Poppins that was controlled by 70 million tiny cathode-ray guns, about 1,000 for each pixel. Each cathode is a 1.2-micron-high cone of molybdenum that is implanted in a grid of electrodes. The tiny structures were formed using the same photolithographic and electron-beam deposition technologies as those that produce integrated circuits.

The miniaturized geometry produces extremely intense fields at the point of each microtip—more than 10 million volts per centimeter—when a charge of 50 to 80 volts is passed through the electrodes. This field helps drive electrons across a 200-micron-wide vacuum to activate a phosphor deposited on the display's glass faceplate. These cold cathodes bear scant resemblance to the cathode-ray guns found in bulky home television sets, which emit electrons only when heated to 1,000 degrees Celsius.

The LETI group has already made progress against some of the technical problems of building vacuum microelectronics displays. Early in its efforts, the team found that current flowed abundantly from just a few of the microtips that form each pixel, whereas other tips generated practically no current at all. "The pixels were not uniformly bright," Duchene says. Adding a two-micron layer of silicon below the microtips supplied enough resistance to smooth out the disparities in current so that brightness varied by less than 10 percent within each pixel.

Duchene is optimistic that a commercial product could be manufactured within three years. But his team must still scale up to larger displays, produce ones with color pixels and prove that the panels can be made in large quantities.

Because the displays require relatively few lithographic steps, Duchene believes that they may prove cheaper to produce than some liquid-crystal displays, a major competing technology. But the various steps added to improve display brightness might undercut the goal of achieving a straightforward manufacturing process, says Webster E. Howard, who manages flat-panel display technologies at IBM's Thomas J. Watson Research Center in Eastview, N.Y. "When you start solving problems by adding complexity, you have to be careful that when you're finished you still have the low-cost approach you're pursuing," he cautions.

Other vacuum microelectronics displays may be on the drawing boards. The Soviets have done work in the field. Hughes Aircraft is pondering the use of miniature displays that could one day brighten airplane cockpits or automobile dashboards. SRI International, the original developer, has an indirect ownership stake in Coloray Display Corporation, a Fremont, Calif., start-up company. The Japanese, too, have their own well-cloaked efforts to develop a display. "If you can think of any of the major Japanese electronics companies, they are probably working on it," says SRI's Spindt.

Now that the French have shown that these displays can work, others are likely to step up their efforts. So a technology with its roots in the 19th century may soon be playing a role in the electronics of the 21st. *—G.S.*

Infrastructure Woes

Going high-tech to diagnose decaying roads

Each summer the state of Pennsylvania's Department of Transportation teaches 200 college students the difference between an alligator crack, a block crack and a pothole. They then spend the rest of their vacation riding and walking along the 43,000 miles of the state's highways, rating each half-mile stretch of road for damage from the constant pounding of trucks, cars and the elements.

But some officials think there has to be a better way to inspect highways than having bored and tired students plod the road shoulders while 18-wheelers whiz past. They are experimenting with a range of technologies, from machine vision for spotting potholes to "smart" pavements that can tell a computer when they become weakened. At various stages of development are ground-penetrating radar, infrared thermography and laser ultrasonics, among others.

The need is acute. The Federal Highway Administration estimates that keeping major roadways and bridges in good working order would have required $35 to $40 billion in 1988, about twice what was spent. The National Research Council's Strategic Highway Research Program (SHRP) and its Transportation Research Board and the U.S. Army Corps of Engineers are among the institutions now road testing new technologies for diagnosing highway maladies.

Some of them have reached the early commercial stage. Infrasense in Cambridge, Mass., is one of a handful of companies that use ground-penetrating radar. It can detect deteriorating concrete and subsurface voids while also measuring pavement thickness and moisture content. This replaces time-consuming drill sampling. Signal-processing techniques are later used to interpret the data.

But highway officials are not rushing to contract such equipment and services. Admittedly, a few of the systems under development may be simple to master with a half hour of training. But most require outside consultants to wrestle with gigabytes of data. "The equipment needs to be made so you don't need a doctorate in radar technology to use it," says John P. Broomfield, a SHRP program manager.

A case in point are machine-vision systems: vans equipped with boom-mounted video or 35-millimeter cameras that move along at highway speeds looking for the cracks and crevices that lead to road decay. One system requires that a technician review the filmed output on a television screen, noting the places where road damage has occurred. Although a truck can film 100 miles or more of road in a day, the technician can inspect only 15 miles of filmed road, as much as a field inspector using nonautomated methods.

Ultimately, pothole scanning may become easier as the materials used in building roads and bridges are endowed with electromagnetic "tags." The Westinghouse Science and Technology Center in Pittsburgh is experimenting with adding inexpensive ferromagnetic particles to concrete. In new roads, they would be evenly distributed and emit a uniform signature in response to a small current or magnetic field. But when the pavement begins to degrade, the signature will vary.

For now, however, public-works crews may stick with home-grown technologies. The best way to find an ailing bridge deck, many old-timers insist, is simply to drag a chain along the concrete surface. A hollow, drumlike thud signals where a crack has occurred from corrosion of reinforcing steel within the concrete.

In addition, purchase orders for basics like road salt usually take precedence over those for lasers, so major repair programs tend to get pushed back. "Something gets done when it gets to the point that a pothole screws up the alignment on the governor's car," says SHRP's Broomfield. *—G.S.*

MACHINE VISION uses a boom-mounted camera to look for pavement damage.

Less of a Problem?

Compacted nuclear waste is still a hot issue

The hydraulic press bears down on a 55-gallon drum with 23,000 pounds per square inch of pressure in a matter of seconds. The metal and its contents of surgical gloves, clothing and other articles from a northeastern research hospital are reduced to an eight-inch-high disk dubbed a "hockey puck."

Scientific Ecology Group (SEG), an Oak Ridge, Tenn., company, is part of a small subindustry of the waste-management business that compresses, burns and shreds low-level radioactive waste. These companies have cut the volume of the least dangerous nuclear wastes to 1.6 million cubic feet in 1989, a reduction of about 55 percent from 1980 levels. "Nobody expected it to decrease by that much," says Gretchen H. McCabe, who directed a study on low-level wastes for the congressional Office of Technology Assessment and is now a researcher with Battelle Seattle Research Center.

Here at SEG, a unit of Westinghouse, some 1.2 million cubic feet of wastes were shrunk to a fourth of their original volume last year. As the largest volume reducer, SEG acts as a nationwide garbage hauler, picking up solid wastes in its fleet of tractor-trailer trucks. Once processed, the waste is sent on its way to one of the three commercial burial sites that currently accept low-level radioactive materials.

Volume reduction is not an unmixed blessing. Radioactivity does not go away but is only concentrated in remaining residues, sometimes requiring additional shielding on burial. Nevertheless, lesser volumes should ease the problem of getting rid of the low-level radioactive wastes generated by nuclear power utilities, universities, hospitals, research laboratories and industry.

Commercial wastes can include anything from a lab coat to animal carcasses to clam shells extracted from the tepid waters inside a nuclear reactor condensor. With an average concentration of .1 curie per cubic foot of radioactivity, low-level wastes pose a biological hazard for up to several hundred years. By contrast, the spent fuel from nuclear reactors averages 200,000 curies per cubic foot and is required to be isolated from the environment for 10,000 years.

The efficiency of waste-reduction technologies has become part of a debate that extends from town meetings to the halls of Congress. Under present law, states have until 1992 to meet a final federal deadline for spelling out how they plan to bury their wastes. The nine compacts that represent groupings of 43 states, as well as the seven unaffiliated states, must each build a repository for their wastes—or make some arrangement to share the disposal burden. If their paperwork is not in order, they may be cut off from access to existing waste repositories.

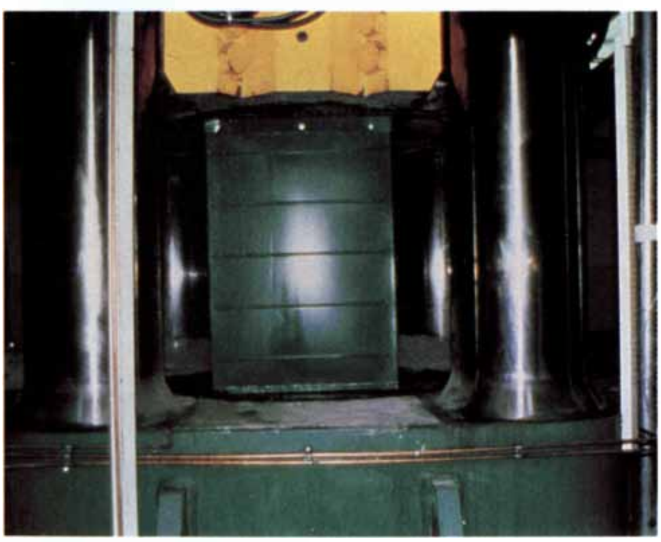
Declining waste volumes have become a point of contention among states who oppose the federal law and who ask why new landfills are needed, each bearing a price tag that starts at $35 million and which can go much higher. "I don't think it was the intention of Congress to end up with a dozen sites," says James F. Cleary, commissioner of the Michigan Low-Level Radioactive Waste Authority.

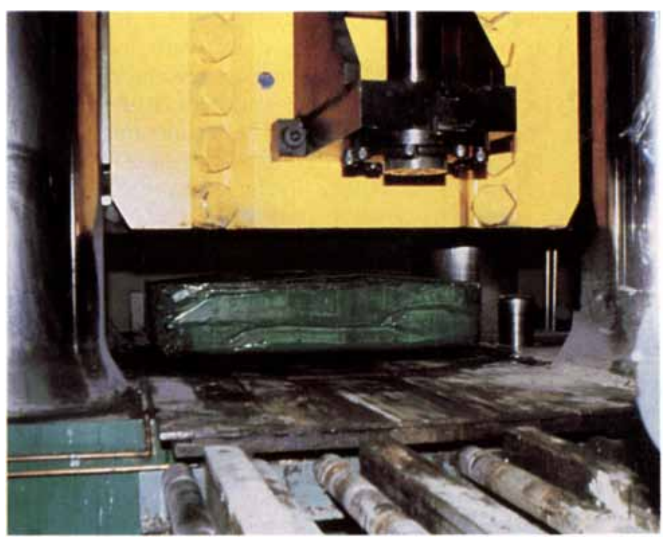
So far only a single site is in place: one of the three existing repositories in Richland, Wash., has agreed to serve as the compact for six other states. Although California may be close to getting a license for a location in the Mojave desert, many other states are at an impasse.

As pins go in the map to mark out prospective sites, protests often begin. In New York the governor and members of a siting commission were either burned or hung in effigy. In Nebraska the head of a local citizens' committee who favors a dump site had his home peppered with bullets.

Michigan and New York have gone to court to get the federal law overturned; a citizens' group from Nebraska has done the same. It is not just the states that think the current plan should be scrapped. A New York–based environmental group, the Radioactive Waste Campaign, wants utilities, the biggest generators, to store their own wastes, in addition to taking in the relatively small amounts produced by hospitals and universities.

The three states that have been accepting other states' wastes for at least two decades see no reason to let the others off the hook. "The purpose of the law was to make states share the burden of waste disposal equitably," says Elaine M. Carlin, executive director of the Northwest Interstate Compact, which is served by the Richland repository. But with the issue so polarized, the half-life for finding a 500-year resting place for the nation's low-level wastes is likely to match that of a long-lived radioisotope. —*G. S.*





***VOLUME REDUCTION** of low-level nuclear wastes is carried out at an Oak Ridge, Tenn., company by using a hydraulic press to crush a four-foot-high (38-cubic-foot) steel box* **(left)** *to about a third of its original height and volume* **(right).**

THE ANALYTICAL ECONOMIST

David or Goliath?

One perennial question in the debate over U.S. industrial policy—or its lack—is the issue of big versus small. Are only small start-up companies nimble enough to take advantage of the fast-moving opportunities offered by advances in technology, or do only large companies have the financial resources and long-term outlook required to sustain ever more expensive research and development?

Such questions take on practical significance because any kind of industrial policy—even relatively noninterventionist strategies such as different tax treatment for capital gains or for debt—will affect small and large firms differently. Billions are at stake. Unfortunately, though, the answers are elusive. And some economists are opting out of the argument: both small and large firms are important, they say, particularly if they are joined together in combines such as Japan's *keiretsu* to fend off foreign competition.

Certainly the U.S. government has not set a clear direction in the debate. It funded the Very High Speed Integrated Circuits program for big companies such as IBM, Rockwell, Honeywell and TRW. Big companies have been the backbone of such research and development consortia as MCC (which now has former DARPA director Craig I. Fields as CEO), and Sematech, which receives $100 million a year in government money to help improve semiconductor fabrication methods. For small companies, there have been countless research and development contracts and even one equity investment.

Semiconductor industry leaders such as Andrew S. Grove, president of Intel, have called for even more massive government funding to keep big U.S. companies competitive in chip manufacturing. No single company outside the industrial combines of the Far East, they say, can afford the billions of dollars required to get into the market for dynamic random-access memories or even the $250-million price tag of a new leading-edge fabrication facility. Grove characterizes many entrepreneurs, meanwhile, as little better than thieves who steal the intellectual property of their former employers. The venture capitalists who fund them, he says, are the financial equivalent of ambulance chasers.

Yet small-company boosters, such as investment banker Tom Volpe, call the venture-capitalist firms that fund start-ups "ultraefficient R&D operations." They bring new technology to market long before large companies can move on a project, he says.

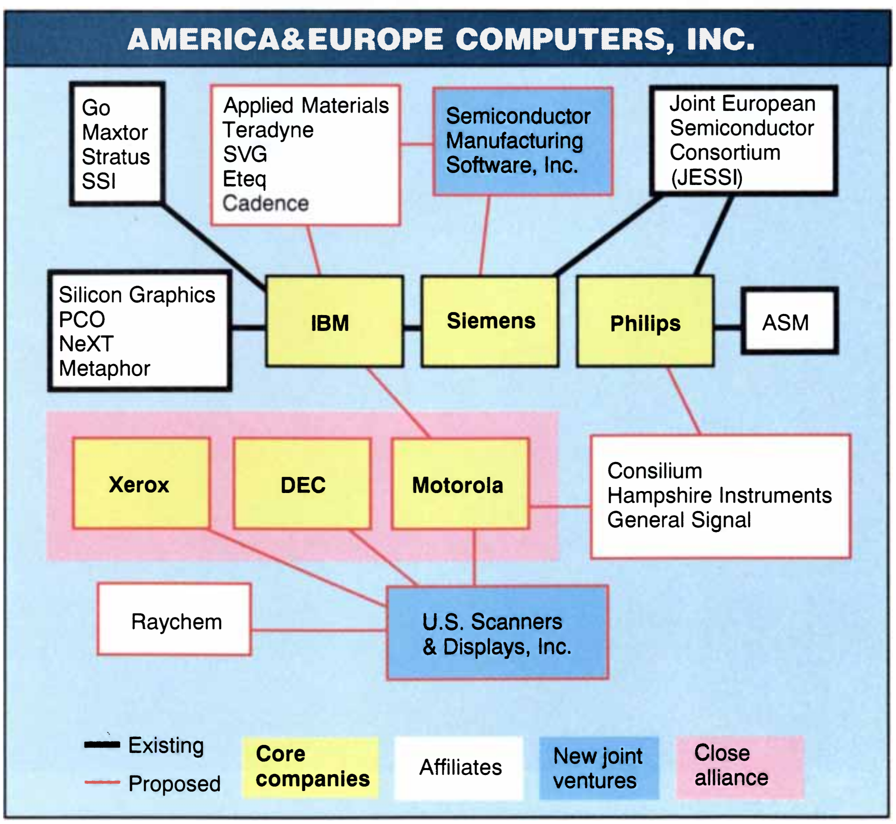
And T. J. Rodgers of upstart Cypress Semiconductors calls talk by Grove "scare tactics to fortify large companies' positions for the government dole." He says that the stultifying management practices of large companies give managers incentives to spend rather than to save, so that the giants inevitably move slowly and at great cost. (Cypress has spent only $73 million to build facilities that make more than $220 million worth of chips a year. At that rate $1 billion would buy almost enough capacity to match industry leader NEC.)

Both sides make strong statements, but there is "more passion than analysis" in the debate, opines economist Kenneth Flamm of Brookings Institution. "There's been very little empirical work," Flamm says. "I've never seen data that I would find compelling," Volpe says. Indeed, the only thing most observers seem to agree on is that there are hardly any data at all to support any of the positions so strongly defended.

Volpe notes that the numbers that might convince him—overall venture-capital returns that include failed companies—are closely guarded. Charles H. Ferguson of M.I.T. puts it more strongly: "The question is impossible to answer by a controlled experiment or by conventional analysis." The only answers, he says, come from the "guesses of smart people with a lot of exposure to various pieces of the puzzle."

Flamm's guess is that the corporate world forms a kind of ecology in which start-ups play a "socially useful role." Large companies may slow innovations to avoid cutting into existing product lines—IBM's minicomputer and office-product sales fell $600 million the year after IBM introduced the PC—whereas small companies have no such cause for compunction.

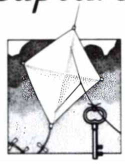
Ferguson, who once championed large companies against start-ups, now says he thinks that the role for giants is as leaders in Japanese-style trading blocs, or *keiretsu:* channeling innovation and helping to guarantee markets and distribution for start-up-engendered technology. He also says he's reconsidering that position because big U.S. firms may not be competent to exercise long-term guidance. At MCC, meanwhile, Fields claims he has always found the big-versus-small debates irrelevant. He is making plans to build a *keiretsu* around MCC—spinning off start-ups and joint ventures with the big shareholder companies—regardless of what the experts say. —*Paul Wallich*



Source: Charles Ferguson, Massachusetts Institute of Technology

THE AMATEUR SCIENTIST

A remote-control camera that catches the wind and captures the landscape



by Forrest M. Mims III

I once envied the local hawks and buzzards for the spectacular view they have of the countryside around my rural home and laboratory. Now I get a glimpse of what those birds see by towing a camera aloft by balloons or a small blimp. This technique has long been used by archaeologists to take aerial photographs of excavations. I recently spoke with a professor of agronomy who plans to photograph experimental crops with a balloon-lofted camera. Such systems also aid surveyors, farmers and meteorologists.

Oblique views of the countryside near the author's rural home and office

Balloon-suspended camera systems are inexpensive and can be stored in a comparatively small space. They can quietly float at low altitudes over places totally inaccessible to conventional aircraft; they can also fly over historic monuments, neighborhoods and other areas forbidden to aircraft for reasons of safety and noise. I have rigged balloon-borne cameras to photograph my family's home, the field where we pasture sheep and the creek that borders our property. The resulting images were better than those that I obtained while flying in an airplane because no windows, wings or struts obstructed the view.

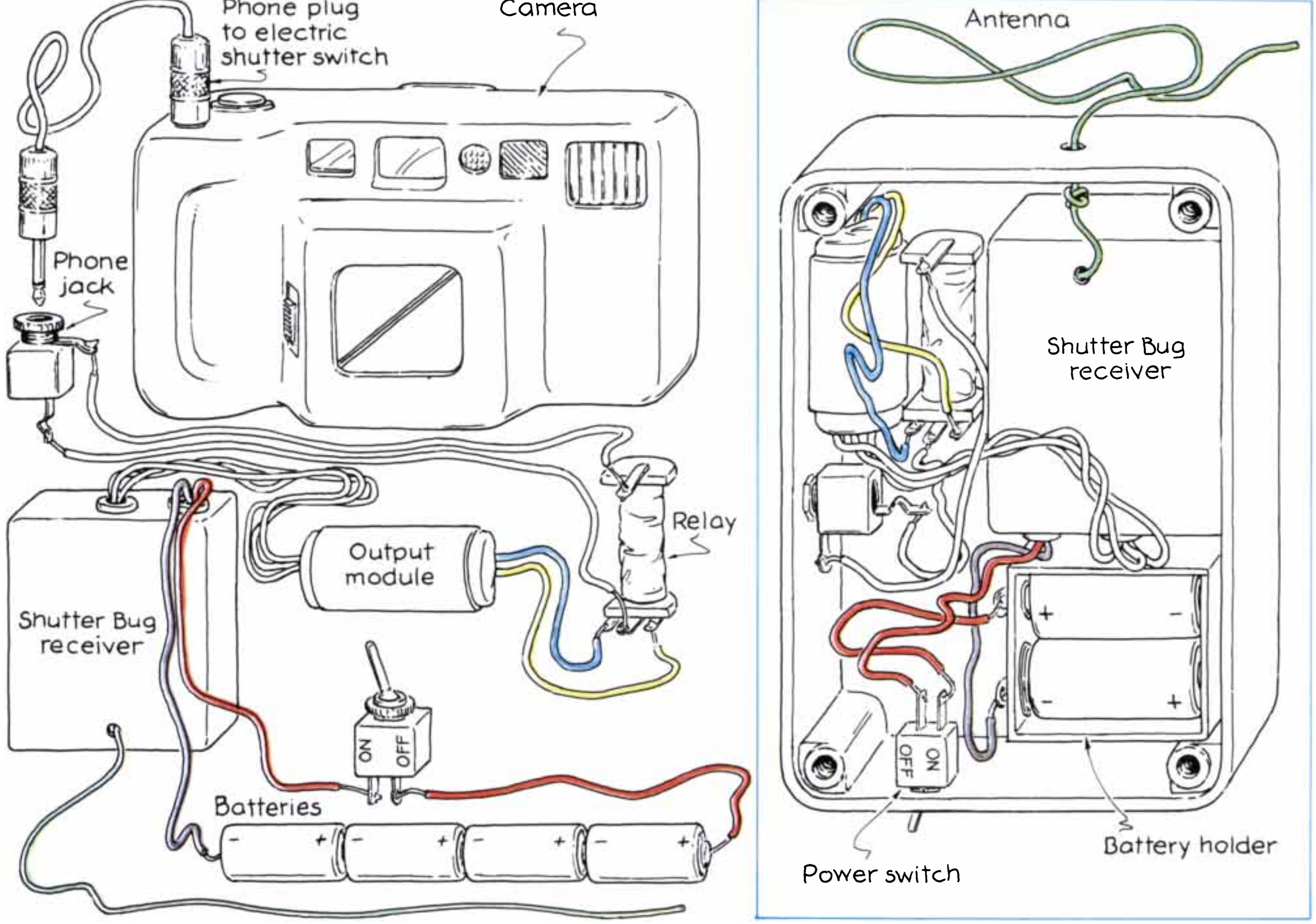
The ideal cameras for aerial photography are the point-and-shoot 35-millimeter kind with automatic focus and film advance. Some can automatically trigger their shutters at preset intervals. Because these cameras do not require a remote-control system, you can simply attach one of them to a balloon and launch it. The camera automatically takes a series of aerial photographs.

Because this method produces rather haphazard results, I much prefer a radio-controlled aerial camera, which is more flexible and reliable. Most cameras can be triggered by a radio-actuated servomechanism like those made for radio-controlled cars, but the mechanism is heavy and consumes much power. A better design incorporates a point-and-shoot camera that can be directly actuated by an electronic signal. Although the details that follow apply to my camera, a Ricoh FF-7, they should be applicable to many electronically actuated cameras manufactured by other companies.

Many kinds of radio-control systems will remotely trigger a camera. Controllers salvaged from toy cars work, although they have limited range and are highly susceptible to false triggering from citizens-band radios and other sources of radio waves. More costly systems designed for large model cars and boats have considerably more range and are much less prone to false signals.

Be sure never to select a radio-control system designed for model airplanes. In the U.S., the Federal Communications Commission assigns specific frequencies to radio-control systems for model airplanes because interfering radio signals can cause a model airplane to crash. You would not want to accidentally redirect your neighbor's model airplane through a window or into a wall.

For two years I have remotely trig-



How to connect the camera to the remote-control receiver

gered cameras with a single-channel radio-control system made by Ace Radio Control, Inc. (Box 511, Higginsville, MO 64037). Ace recently introduced a system called the Shutter Bug, which has greater immunity to interfering signals than do earlier models. The system, consisting of a transmitter, a receiver and an output module, costs about $85 (catalogue number 40E83).

The diagram above shows how to connect the receiver of the Shutter Bug to my camera. The receiver, which is housed in a small plastic enclosure, has several wires emerging from it. The single wire is the antenna. The twisted red and black wires with a small toggle switch are the power-supply connections. Three twisted wires connect the receiver to its output module. Other parts can be obtained from an electronics supplier, such as Radio Shack.

The wires from the output module should be soldered to the coil terminals of a miniature five-volt relay with a coil resistance of 250 ohms or more. (The relay is an electromechanical switch that can actuate an electronically controlled camera.) You should next connect one end of a pair of wires to the relay's moving and normally open terminals. Solder the other end to a 1/8-inch phone jack. Make a cable to connect this output jack to the camera's remote-control jack. Solder a 1/8-inch phone plug to each end of the cable.

The receiver requires a six-volt power supply, which can be made from four 1.5-volt cells connected in series. Obtain a four-cell AAA battery holder. Solder the receiver's red lead to the positive terminal of the battery and the black lead to the negative terminal. Double-check the connections before reinstalling the receiver's batteries.

Then turn on the power switches of the transmitter and receiver. Be sure the camera is turned on. When you press the transmitter's fire button, the relay's contacts will close, the camera shutter will trip and the film will be advanced. If the camera does not respond, keep the transmitter button pressed down for a second or so. If nothing happens, check the wiring and make sure the batteries are delivering five to six volts.

Install the receiver system and battery pack in a plastic enclosure [*see illustration above*]. Secure the receiver with double-sided tape or hook-and-loop fastening tape, and stuff foam plastic between the components. Shake the assembled system to make sure it still works.

You can attach the receiver system to the back of the camera with rubber bands. Even better, mount the camera and receiver on a platform made from a thin sheet of aluminum or model-airplane plywood [*see illustration on next page*]. A pattern of holes drilled in the platform saves weight and provides attachment points for the receiver, camera and rigging.

The receiver can be secured with a machine bolt, and the camera can be mounted directly on the platform with a screw (from a camera store) that fits its tripod receptacle. You can reorient the camera by rigging the platform in various ways or by affixing the camera to a ball-and-socket mount (found in most camera stores). As a safety measure, tie the camera to the platform with a short line.

To guarantee that the platform hangs vertically, I sometimes attach the receiver to a strip of aluminum extending from the platform. The strip offsets the weight of the camera. To prevent the camera from rotating and swinging, I tie the nylon tether that prevents the camera system from floating away directly to the platform instead of to the balloons.

You will next need to obtain the balloons themselves, the helium and a reel to wind the tether around. Surplus weather balloons can be used, but they are fragile. Whenever I do use them, I tie the camera system to a parachute whose canopy is tied to the balloon, or I deploy two or more balloons. These precautions twice saved my camera when a balloon suddenly burst high in the sky.

A more robust and very inexpensive alternative to weather balloons and costly rubber balloons is what I call the BPTB lifting device, better known as the black plastic trash bag. Trash bags are available in various colors, so you will need to change the designation accordingly (a yellow bag is a YPTB, red is RPTB and so forth). Black is best because it absorbs the heat from the sun, thereby expanding the volume of the bag and enhancing the bag's lifting capacity. Find the largest-capacity bags possible.

Helium can be purchased from welding shops and party stores. Helium and cylinder rental prices vary widely. In my area, enough helium to fill a cylinder with a capacity of 6.5 cubic meters (230 cubic feet) sells for around $40 at a welding supply center. There is an additional charge for renting the cylinder ($3.75 a month) and a balloon-filler nozzle fitted with a pressure gauge ($7.50 a month). If you plan to launch many balloons or trash bags, you can save money by buying a nozzle and leasing a cylinder.

Gas cylinders are heavy; when transporting them, be sure they are well secured so they cannot fall. Although helium is an odorless, nontoxic and inert gas, it becomes an asphyxiant when it displaces the oxygen required to sustain life. You should therefore fill balloons outdoors or in a well-ventilated room.

At sea level, helium can lift approximately 1.1 kilograms per cubic meter, or 1.1 ounces per cubic foot. The simplest way to determine the lifting capacity of a single BPTB where you live is to fill one with helium, attach a small container and add water to the container until its weight exceeds the lifting capacity of the BPTB. Then weigh the container. This method reveals that a 50-gallon BPTB will lift around 70 grams (2.5 ounces).

Unless you have a large balloon or a small blimp, you will need a cluster of balloons to lift your camera. The total weight of my camera and a Shutter Bug receiver installed in a plastic enclosure is 460 grams. The platform and tether add another 100 to 200 grams. A cluster of nine 50-gallon BPTBs will lift the system with an excess margin of 10 to 15 percent. Avoid the temptation to fly your camera from a cluster that provides just enough lift to float. Instead, add one or two extra BPTBs in order to keep the camera airborne should one of the BPTBs develop a leak. Doing so will also help keep the system from blowing down in a light breeze.

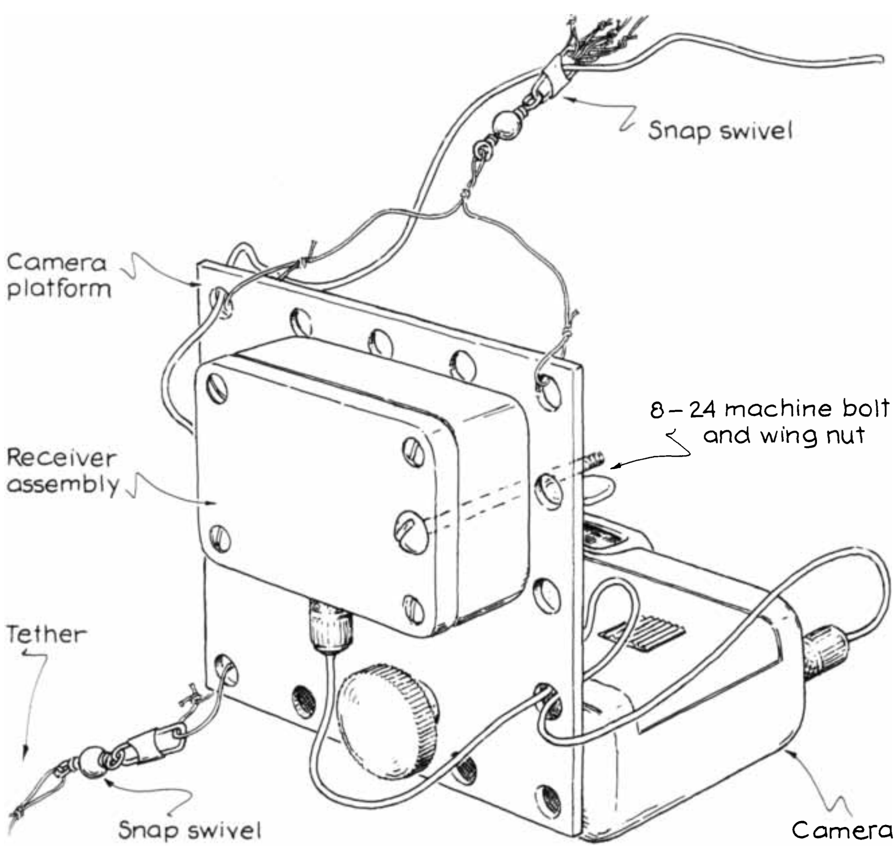
For your first flight, gather all the required equipment in a suitable spot. Load the camera with film and test the radio-control system. Then launch a small helium balloon tied to a 10-meter tether. If the balloon is blown to the ground, it is too windy to fly the camera platform.

Before you inflate a trash bag with helium, squeeze all the air out. Next gather the bag's open end over the helium-tank nozzle and hold it in place while the gas flows inside. When the bag is full, twist the gathered opening, bend it over against itself and secure it in place with tape. Tie a two-meter length of nylon tether above the taped region. Make a loop out of the other end and temporarily connect the loop to a snap swivel, such as those used in fishing tackle [*see illustration at left*]. The snap swivel should be attached to a block of wood, which you can later use as a ground anchor.

When all the BPTBs are inflated, attach the loops at the ends of their lines to another snap swivel tied to the camera platform. Be certain that the tether attached to the camera platform is temporarily but securely anchored to the ground.

Trash bags and spherical balloons have the aerodynamic properties of tumbleweeds. Even a light breeze will blow down a tethered balloon if its lifting power is only slightly greater than its payload. You can compensate for this imbalance by launching upwind from the point you wish to photograph. Attach the tether reel to a ground anchor downwind from the target. Hold on to the balloon and walk upwind, unwinding the tether as you go. Then release the balloon and race back to the tether reel. (Please watch the ground during this phase of flight.)

Watch the camera carefully as it approaches the target. When it is reasonably still, press the transmitter fire button. If the camera is swinging, press



The camera platform and its rigging

the button when the camera is at the apex of each swing.

The balloon may be blown down when it reaches the end of its tether, so be sure there are no roads, power lines or other hazards below its flight path. Avoid pulling on the tether when the balloon is over trees or other obstacles, because you may pull the balloon down onto the obstacle you are trying to avoid. In addition, always fly when the air is calm and never when an electrical storm is nearby. Use only helium to inflate balloons. Hydrogen and other lighter-than-air gases are highly flammable and can explode. Never use a wire tether. And never attempt to trigger a camera with a pair of wires leading to the ground. Should the wire or wires touch a power line, you or anyone coming in contact with them may receive a fatal shock. Avoid flying tethered balloons near airports, tall buildings, power lines and busy roads.

If you wish to upgrade from trash bags, the blimps used for lifting advertisements and meteorological instruments are well suited for lifting cameras. Because they are aerodynamically stable, they can be flown in a breeze, and their large capacity permits them to lift several kilograms to 200 meters or more. Blimps are expensive—small ones cost from around $400 to $1,000.

I fly my cameras and various instruments from a blimp supplied by Atmospheric Instrumentation Research, Inc. This blimp, now manufactured by Aerial Innovations (8111 Raleigh St., Westminster, CO 80030), has a length of 4.2 meters and a maximum diameter of .9 meter. It will lift two cameras and a radio receiver system.

Much can be learned from aerial photographs. For example, you can measure the dimensions of objects in a vertical photograph. Dividing the height (h) of the camera by its focal length (f) gives a scale factor known as the photo scale reciprocal (PSR). If the image dimension of an object is d and the actual dimension is D, then $PSR = D/d$. If d and the PSR are both known, then $D = d \times PSR$.

Assume that a camera with a focal length of 35 millimeters provides a photograph of a fence from an altitude of 100 meters. This gives a PSR of 100/.035, or 2,857. If the length of the fence as measured on the photograph's negative (d) is 20 millimeters, the length of the fence (D) is then .02 × 2,857, or 57.1 meters.

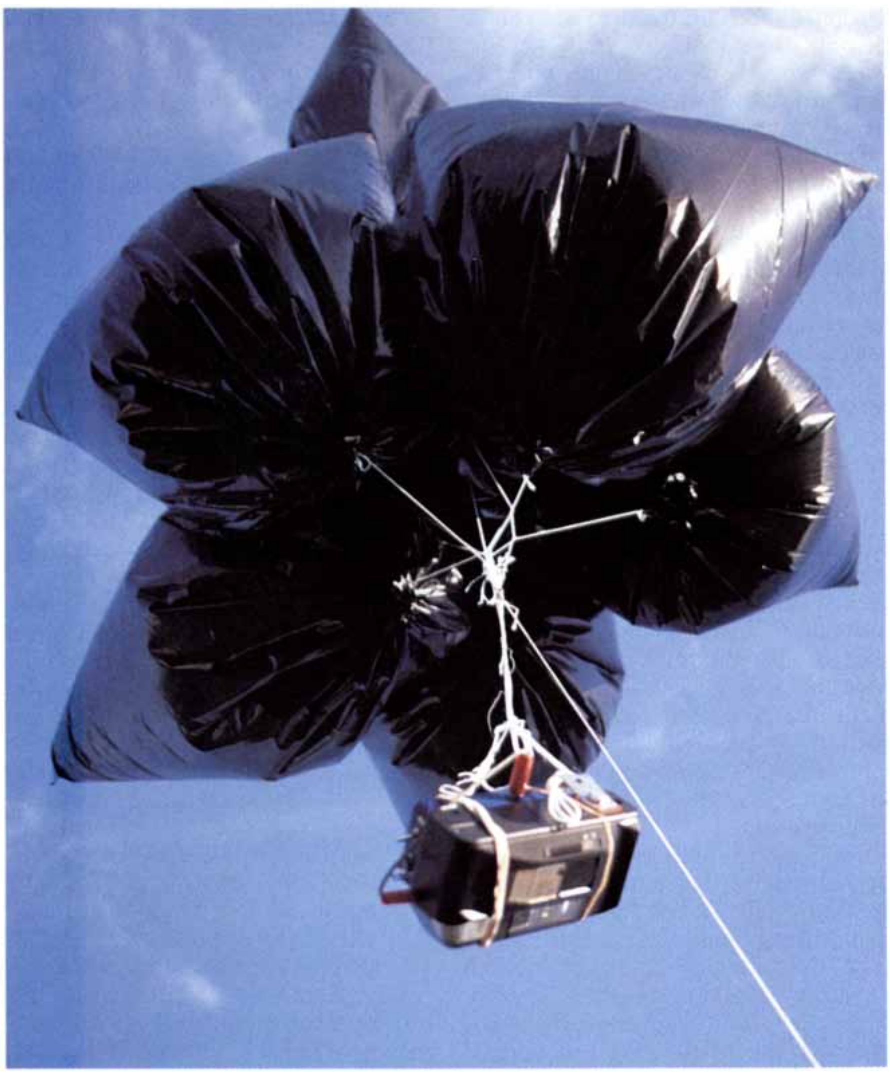
You may want to know the altitude of the balloon. It can be calculated if you know the actual dimension of an object in a photograph. The formula is $h = (f \times D)/d$. Assume, for example, that a driveway you know to be 20 meters long has a length of 10 millimeters on a negative produced by a camera with a focal length of 35 millimeters. The altitude is $(.035 \times 20)/.01$, or 70 meters.

These formulas will help you determine how high to fly a camera to photograph a given area. A 35-millimeter slide or negative measures about 23 millimeters wide and 35 long. A 35-millimeter camera with a lens whose focal length is the same size will therefore cover a space that measures 197 by 300 meters when the camera is flown at an altitude of 300 meters.

Those readers who fly balloon-borne cameras will soon accumulate a collection of adventure stories, all nicely illustrated by portfolios of aerial photographs. May your breezes be light and your landings soft.

FURTHER READING

AIRBORNE CAMERA: THE WORLD FROM AIR AND OUTER SPACE. Beaumont Newhall. Hastings House, 1969.



***A camera payload suspended from a blimp* (left) *and a cluster of helium-inflated trash bags* (right)**

BOOKS

The president who would be scientist, escaping flatland, art and science, sensing the world

by Philip Morrison

Thomas Jefferson: Statesman of Science, by Silvio A. Bedini. Macmillan Publishing Company, 1990 ($29.95).

Two days before he would be inaugurated as the second vice president of his country, an office he foresaw as "honorable and easy," Thomas Jefferson, the owner-architect of Monticello, author and revolutionary, ex-governor, ex-ambassador, keen natural philosopher and collector, arrived in Philadelphia, the nation's capital. There he was welcomed by cannon fire and a flag that proclaimed "Jefferson the Friend of the People."

On that occasion he had brought with him a scientific paper to be published in the *Transactions of the American Philosophical Society*. His manuscript was a detailed study of a fossil that he named *Megalonyx*, for he considered it to be a clawed predator "of the family of the lion... but as preeminent over the lion in size as the Mammoth is over the elephant."

Jefferson was most fortunate; while browsing in Philadelphia's bookstores, he came across an article by the great French naturalist Georges Cuvier, in which a similar fossil species, which had been found near Buenos Aires, was described. Cuvier concluded (on the basis of an almost complete skeleton) that the species, which he called *Megatherium*, was a giant relative of the leaf-eating tree sloth. Jefferson quickly revised his paper, adding a footnote citing Cuvier's paper, and allowed that his own specimen was probably not a carnivore either.

The ground-sloth paper supports—perhaps more than any of his other works do—the notion that Vice President Jefferson had a real talent for science. The "tranquil pursuits of science" were his "supreme delight," he once wrote. He was an eager and gifted amateur throughout the realm of science and technology, a weaver of international scientific acquaintances and correspondents, an avid patron of useful novelties, a wide reader, a good calculator and, above all, a superb promoter, the quintessential statesman of science.

His formal scientific education, although brief, was effective. Jefferson's father, a farmer and surveyor, was a powerfully built yeoman among the Piedmont gentry who lived along the upper James River in Virginia. He taught his lanky, red-haired, sunburnt son the art of woodcraft and the rudiments of surveying, but he died when Thomas was only 14. After his father's death, the boy was sent to the best country tutors, learned Greek and Latin brilliantly and at 17 found himself a student at the College of William and Mary, the only academic institution in the South at that time.

The college was located in Williamsburg, the first settlement Jefferson had seen that was home to more than 20 houses; there, dispersed among the flat tobacco fields, were 200 buildings occupied by 1,000 people. Such telling perceptions of 18th-century life are what enrich this readable and comprehensive chronicle. The college boasted six professors and a student body of 115, most of whom were sent there to study by wealthy (but not scholarly) Virginian families.

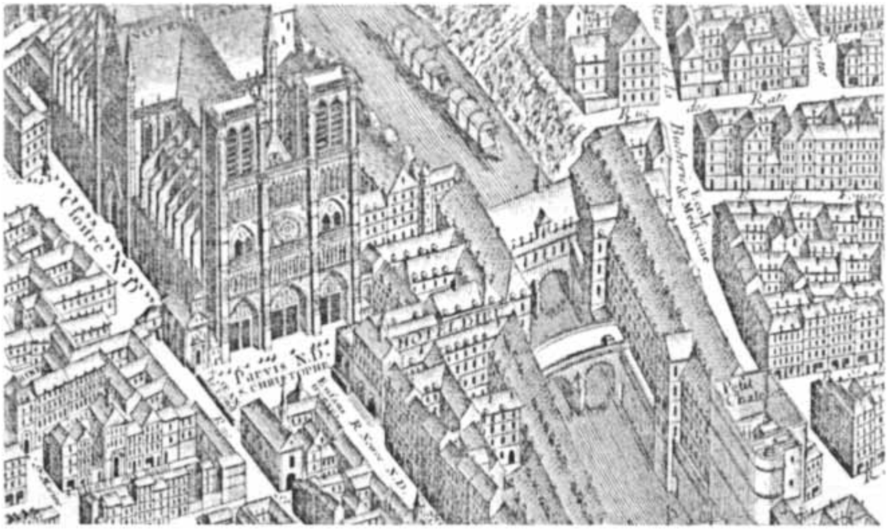
A young professor of mathematics from Scotland, William Small, had an unusually enlightened view of the world and befriended Jefferson, who never forgot him; "from his conversation I got my first views of the expanse of science & of the system of things in which we are placed." Small introduced Jefferson to an elite circle of intellectuals, which included the governor of Virginia. Their meetings had the flavor of the English Enlightenment, focusing as they did on the use of such prized instruments as the microscope, telescope, orrery, globe and barometer. In later life Jefferson would buy such instruments for himself; indeed, the inventory he put together is testimony to his wide-ranging interests and his lively curiosity, which extended far beyond the needs of his surveying profession.

Jefferson also apparently had an excellent memory. The reader learns that trigonometry and the principles of navigation were still at his fingertips when he turned to them 40 years after college. In his youth he had studied Anglo-Saxon history and learned to read French and Italian. Although he heard Italian spoken for the first time by the gardeners who built his Monticello vineyards some 10 years later, he was able to converse enjoyably with them.

After spending a dozen years practicing law and colonial politics, the Virginian, a prose stylist of repute at 33, found himself drafting our country's Declaration of Independence. He wrote it at a laptop mahogany desk that he designed to his own specifications and commissioned from a distant kinsman, a cabinetmaker with whom he had lodged in Philadelphia. He used that writing board for some 50 years and along with it accumulated a whole catalogue of clever aids: copying presses, dumbwaiters, macaroni molds—the list goes on. Several inventions, including a moldboard plow and an ingenious ciphering wheel, which was a century ahead of its time, owe their existence to Jefferson.

His technical eye was keen. Visiting England during the years when he was ambassador to Paris, he described his encounter with a steam engine, noting the eight pairs of millstones turned by one engine. From France he reported the manufacture of a product with interchangeable parts: musket locks made by the gunsmith Honoré Blanc. He strongly urged the government in Washington to promote similar methods of production, but without success. Replaceable parts were not mass-manufactured in the U.S. for another 15 or 20 years. Whether Eli Whitney ever manufactured such parts on a mass scale "remains in serious doubt," Bedini says, although Whitney more than once claimed credit for doing so.

The rich narrative spins on. Bedini describes the readying and return of the Lewis and Clark expedition and of the three or four other less celebrated surveying ventures that President Jefferson aimed westward. He also describes the later years of Jefferson's life when he saw his own University of Virginia flower. Jefferson had hoped to found a national university, but that was beyond his political reach, so he settled instead for a state-run institution. He lovingly designed the beautiful Charlottesville campus, including the rooms that would both exemplify the orders of architecture and house the young men well.



DETAILED and complex information can actually simplify an illustration, if it is properly arranged, as it has been in this three-dimensional map of Paris, drawn by Michel Etienne Turgot and Louis Bretez in 1739.

One night some students of the first resident class, "animated ... with wine" and donning masks, derided "European professors" (most of the nine faculty were newcomers from Britain) and even stoned two of them. The next day the Board of Visitors gathered all 40 students together. On the platform in grave monitory concern were three former presidents: Thomas Jefferson, James Madison and James Monroe. Jefferson could not make his comments clear through his tears. Much moved, most of the offending students came forward in submission.

In these pages we pay a documented scientific tour to a yeasty nation of 10 million; born of the Enlightenment, although still mired in slaveholding, what a hopeful land!

ENVISIONING INFORMATION, by Edward R. Tufte. Graphics Press, 1990 (P.O. Box 430, Cheshire, CT 06410) ($48, postpaid).

Once again Tufte, a professor of statistics and graphic design at Yale University, offers us in samizdat a colorful, beautifully illustrated and well-argued volume on "cognitive art," which is the visual representation of experience and reason. There is an ingenuous stance that the reasoning mind, which works by alphabet and formula, is somehow different from the mind of the artist, who models by visual image. Fortunately, such a view appears to be a nearly exhausted fad. Tufte's book copiously confirms the mind's unity, as it "arrays exemplary designs ... for all types of information ... at the intersection of image, word, number, art."

The galleries of this paper museum are rather sunnier than those of Tufte's first book, *The Visual Display of Quantitative Information,* where more of the graphics were informative specimens of graphic pathology. In the current book the annotated displays center on coherent and helpful strategies for visual presentations that both work and delight, or work because they delight ... or delight because they work. It is the fascination of what we see, both as rendered design and as bearers of interesting information that makes this savvy book as much a pleasure for the armchair amateur as it is a guide for the professional.

However cogent the text, it is the hundreds of apropos images that provide its real strength. I cannot tithe them in a list, but a sample or two may lead the reader of this review to seek more. For instance, the chapter on micro/macro readings opens with Michel Etienne Turgot and Louis Bretez's wonderful *Plan de Paris,* which shows each building of that royal city; it is paralleled today by a three-dimensional map of midtown Manhattan, designed by Constantine Anderson. Even bus shelters and telephone booths are marked among the forest of tall buildings aligned on the deep-textured page. The map represents a bold tactic: "*to clarify, add detail.*" The extraordinary artistic and popular triumph, Maya Ying Lin's Vietnam Veterans Memorial, is displayed and analyzed as another gifted exemplar of the power of detail.

Tufte demonstrates ever so clearly how the grids that enclose tabular numbers on a railroad schedule can kill clarity by providing signals that have no meaning. In contrast, two illustrations, each a series of postage-stamp images, are enchanting. One is a page of 30 close-ups showing how the calligrapher's brush executes a single *kana* character. The second aligns 10 renderings of a single pictograph that was found on a much inscribed rock along the Taunton River in Massachusetts. The forms were drawn long ago by unknown Algonquin artists and, until the invention of the camera, were redrawn and interpreted in pictoral form by many different hands. The amazing sequence of icons—from the first effort in 1680 to the last one in 1854—demonstrates how hard it is to image what you do not understand. Astronomers, be not proud; are there canals on Mars?

In the chapter on color and information, color is described as "subtle and exacting." A set of color photographs showing a grove of birch trees at different times during the year beautifully illustrates the use of colors found in nature, especially those that are so elegantly subdued. Maps and emphatic graphics often fail in this; a grim example is shown of a bright-blue U.S. map edged in white whose colors overwhelm its informational content. In contrast, a splendid map of the world's oceans does nothing cartographically unusual, but its blue values for ocean depths and apt contoured brown shades for land surfaces create a textbook example of what can be achieved with fine craftsmanship.

Some of the ideas presented in this book are brand-new, yet others are surprisingly old. The book closes with a page from Galileo's 1613 paper in which he describes the first telescopic observations of Saturn. The poorly resolved planet and its enigmatic handles (which would one day be seen to be rings) are drawn within the Italian text itself, in sketches only a line and a half high. In his epilogue, Tufte writes three lines of free verse: "The stunning images, never seen before, were just a sentence element. / Saturn, a drawing, a word, a noun. / The wonderful becomes familiar and the familiar wonderful."

CONNECTIONS: THE GEOMETRIC BRIDGE BETWEEN ART AND SCIENCE, by Jay Kappraff. McGraw-Hill Book Company, 1990 (cloth, $39.50; paperbound, $19.50).

It was Jorge Luis Borges, that endlessly inventive fabulist-librarian of Buenos Aires, who imagined the universe as a library, where bookstacks extended in all directions, unchanging and yet varied. No two volumes, however uniform in their serviceable bindings, held the same contents, for between their covers they contained all the possible strings of letters, a set finite but immense. In this imaginary library, readers would wander lifelong among oceans of random text, searching for meaning.

The author of this engaging compilation, a designer-mathematician from the New Jersey Institute of Technology, reminds us of another sort of imaginary universe, which also has the air of a magical dream. This one, however, was rigorously constructed a decade or so ago by two English mathematicians: Roger Penrose and J. H. Conway.

First, Penrose showed how to cover an endless floor with simple space-filling tiles that are neither triangles, rectangles nor hexagons but two different rhombs, each one related to the regular pentagon. As one proceeds to tile, it is necessary to choose a tile that fits; sometimes the choice is forced; sometimes it is free. The result is a floor that "always seems to be striving" to replicate itself but never does.

It was Conway who demonstrated the repetitive nature of such a universe. Like the imaginary readers of Borges' library, one can carefully study a Penrose-tiled universe; across an area of diameter *d*, many variant patterns emerge. But now suppose one is magically transported to another point on a Penrose-tiled floor, where different choices were made when the tiles were laid out. There the exact replica of the first region will be found somewhere within 2*d* of the new starting location!

Kappraff's narrative opens most disarmingly for the nonmathematical reader. He outlines three classical canons of architecture: partial repetition, harmony among the parts and variety (the last is not entirely predictable). His account of the geometry of the two- and three-dimensional spaces in which we live and design proceeds past the classics, past the half-mythical Pythagoras, the real Plato and the Renaissance architect Leon Battista Alberti, to the work of such modern figures as the French architect Le Corbusier and the composer Bartok. Architecture is, after all, a form of frozen music; this book brings that comparison to life at a simple technical level.

Kappraff builds on the three classical numbers by which two-dimensional objects are described: the golden mean, the diagonal of the square and the sacred cut, all of which are ways to divide a rectangle by one or by two horizontal and vertical lines, to yield a few self-similar rectangles. From this base of information, which is so clearly explained and well drawn, the reader is led through the wonderful self-similar proof of the great theorem of Pythagoras. Next comes a brief discussion of fractal dimensions, followed bravely by the freewheeling geometry of dots and lines, which in turn lead to graph theory, froths, maps and surface partitions. Three-dimensional space is introduced by rolling and curling a plane, until the reader learns how an unambiguous map of a doughnut's surface may require seven colors, not just the four that suffice for an ordinary plane map. Penrose tilings come next, and at last the full crystallographic symmetries of the plane are described.

Three-dimensional space symmetry is awarded perhaps a third of the whole volume. This is no place to learn much about scale in the world, although the subject is given a page or two. The only account of dynamic matter, namely, the geometry of stems and leaves, is given a fresh treatment in the Fibonacci way. The focus instead is on static form.

This book, on the mathematics of natural and artful form, is a lively new entrant to the small shelf of those fine works. Its topic is already somewhat familiar to those readers who have followed my witty colleagues, the mathematical columnists of *Scientific American*, for more than three decades. The antique sources of some of this matter flavor the text with a hint of occult appeal, mainly through assertions made by some credulous old scholars. But the persuasive demonstrations are marvels enough.

A NATURAL HISTORY OF THE SENSES, by Diane Ackerman. Random House, 1990 ($19.95).

Poet and essayist Diane Ackerman has chosen the right title for her book, the style of which evokes the last fin de siècle decade, a time of mannered reaction to long-feigned loftiness. Here and there her prose is written in musk and old purple, and even a little ordure, not to an excessive degree but one that is proportionate to our candid epoch.

The book is a celebration of our "sense-luscious" world. Its literary model is not the Paris nights of Joris-Karl Huysmans, but instead it captures the darkness of Helen Keller, who so vividly knew her own curtailed world. By smelling people, she could decipher "the work they are engaged in.... The odors...wood, iron, paint, and drugs cling to the garments [as] a person passes quickly [through] the kitchen, the garden, or the sickroom."

The book's main chapters predictably muster the five senses: smell, touch, taste, hearing and vision. Each one is subdivided into a lively string of small essays, 100 or so in all. The best ones express informed reflection and well-traveled experience over the sensory range; some encapsulate the underlying science; others outline a rich lore derived from topical sources. A good deal of armchair ethnography is cited; some of those descriptions, as in the accounts of cannibalism, probably perpetuate tall tales more than they illustrate real life.

The book's style is best illustrated by letting the perceptive Ackerman speak for herself. "The most scent-drenched poem of all time, 'The Song of Solomon,'...weaves a luscious love story around perfumes and unguents.... [W]ater was rare, people perfumed themselves often and well, and this betrothed couple...converse amorously in poetry," invoking as loving metaphor all the spices of the fragrant garden. Elsewhere the author writes: "Weighing from six to ten pounds, [skin] is the largest organ of the body, and the key organ of sexual attraction" harboring the sense of touch.

Chocolate—Montezuma drank it over snow brought to him by runners from the high slopes of the Mexican volcanoes—is well known to induce craving. It seems that the craving can no longer be blamed on a neurotransmitter, as was once thought, because the same neurotransmitter exists in even greater quantities in salami and cheese, neither of which enthrall as does chocolate.

Ackerman writes four pages in aromatic praise of vanilla, too, describing a luxurious bath with vanilla, followed by the donning of vanilla perfume "with a bitter sting," and the possibility of six dinner dishes touched by the flavor, ending with a glass of warm brandy into which a vanilla pod has been chopped, enabling her to "fall into a heavy orchidlike sleep." Her account of vision is perhaps the most public of the five sensory chapters, as she looks sharply into skies and autumn leaves and dilated pupils and notes that "polar bears are not white, they're clear," like snow itself.

William Blake asked, "How do you know but that every bird...is not an immense world of delight closed to your senses five?" Science claims to offer him a partial reply. After some excursions, the poet Ackerman closes by saying: "The great affair...with life is to live as variously as possible.... It began in mystery, and it will end in mystery, but what a savage and beautiful country lies in between."

ESSAY

The business of science education

by James L. Johnson

The evidence is clear: American students are technologically illiterate, possessing few if any of the mathematics and science skills needed to function successfully in the world today or as it will be in future decades. Comparisons of test scores show that U.S. students lag behind their counterparts in much of the Western world in science and mathematics. Worse, their scores in these critical disciplines show no improvement.

This trend is confirmed by U.S. business. It is increasingly difficult for high-technology firms to find the scientists and engineers who will make tomorrow's discoveries. The shortage of competent people is a long-term problem that threatens the competitiveness of the U.S. in today's global economy.

American business recognizes this problem and is prepared to do its part to solve it. The question is: Where do we begin? Improving higher education, although essential, is not the only answer. Help must begin with support of general education—even at the elementary school level—to ensure that students develop strong basic skills well before they graduate from high school.

In 1989, the Business Roundtable, a national organization of corporate leaders, challenged each of its members to select one city in which to make a strong commitment to improve education. Virtually all 200 members of the Business Roundtable have accepted the challenge. GTE is supporting the introduction of a School-Based Management system in cooperation with Hillsborough County Public Schools in Florida. IBM sponsors mentor and other programs in Austin, Tex. The Monsanto Company gives special attention to science education in St. Louis. Borg-Warner supports "Project: Achievement" in Chicago.

Many educational programs sponsored by business attempt to change the traditional methods of teaching students and operating schools. The School-Based Management system, for example, trains teachers to solve problems as a group and empowers them to implement their solutions. Principals and faculties receive increased authority in the areas of budget, staffing and curriculum. Business-supported programs also place a premium on establishing goals and measuring progress.

Business must also reach out to those already in the work force. Many companies train new employees to help them develop the skills required for the work they will perform. Employees also participate in educational activities to increase competency and qualify for promotions. One estimate indicates that U.S. companies spend $25 billion a year to upgrade employee skills.

These efforts are most successful when businesspeople are involved in the educational process. Their involvement affects not only the quality of education available to their own children but also the quality of life in their communities. Employees of America's businesses serve on school boards, participate in PTAs, speak to students about their jobs and even teach classes at local universities. Many corporations encourage this involvement through employee-volunteer programs, leaves of absence and executive-loan programs.

Companies can also directly support education through their products and services. We can apply technology to enhance educational programs. The nation's telecommunications companies, for example, are developing advanced fiber-optic communications networks that will one day be able to bring educational video programming to any classroom connected to the telephone network. Using satellite networks, we already can achieve similar results with rural schools. This so-called distance learning can bring advanced language, mathematics and science teaching to any classroom where personal instruction is not available.

Finally, business must encourage the development of innovative educational programs. For technology companies, this means targeting mathematics and science as the disciplines most in need of support. To be successful, such programs must highlight the human aspect of mathematics and science while providing hands-on activity.

Training in mathematics and science typically involves numbers and inanimate objects rather than words and people. Many younger students find this type of instruction cold and impersonal and shy away from these subjects. Innovative educational programs must frame mathematics and science as human endeavors conducted by people for the benefit of humankind. Humanizing these subjects is essential to capturing student interest.

Successful mathematics and science training must also include hands-on activity. The often-romanticized vision of a scientist working in solitude with obscure formulas and theorems is inconsistent with the process of learning. To learn, students must interact with the people and instruments of science: scientists, mathematicians, engineers, computers, lasers, laboratories.

Examples of innovative mathematics and science programs already exist. My company, for instance, cosponsors a program with the United Negro College Fund (UNCF) to encourage minority students to pursue a teaching career in mathematics or science. Teams of faculty members and students from five UNCF institutions are invited to participate in a two-week summer internship at the firm's laboratories. There they interact with scientists, work with the latest scientific equipment and see first-hand how research relates to real-life issues. The first internship was held this year, and I am told the response was very positive.

At GTE our belief in, and commitment to, mathematics and science education now includes the national underwriting of *Scientific American Frontiers,* a new hour-long science television series premiering this month on Public Broadcasting Service (PBS).

We believe this series does an excellent job of showing that science is a human enterprise, that it is not just high-tech equipment and cold facts but men and women who are excited about what they do. At the same time the programs provide young people with entertaining examples and projects that include classroom activities. The *Scientific American Frontiers* package includes the distribution of more than 250,000 sets of free classroom teaching aids that directly supplement the PBS broadcasts. An estimated 150,000 teachers and 16 million students will use *Scientific American Frontiers* classroom materials in 1990–1991.

I encourage *Scientific American* readers to watch *Frontiers* and send the magazine your comments and suggestions. GTE has made a five-year, $11-million commitment to support the series, and we look forward to an even longer partnership in our efforts to help improve the scientific and technical skills of America's youth.

JAMES L. JOHNSON is chairman and chief executive officer of GTE Corp.


INTRODUCING
GM
Let's get it together...buckle up.
1990 GM Corp. All rights reserved. Park Avenue is a registered trademark of GM Corp.
J.D. Power and Associates 1990 Initial Quality Survey. Based on owner-reported problems during the first 90 days of ownership.

THE GOLD STANDARD. NOW IN BLACK.

Unique dual-locking mechanism assures a tight, shock-resistant seal.

Now that data processing has outgrown the controlled environment of the "Computer Room," it's become a tough world for the floppy disk.

Today diskettes are at work almost everywhere. Even in such inhospitable environments as bumpy airplanes, crowded commuter trains, and dusty factories.

All putting your invaluable data at risk.

Now Maxell introduces the MF2-HD Super RD.

With advanced technology whose time has come because it assures outstanding performance even in extreme operating conditions.

The key is our revolutionary new Dual Interlocking Flex Shutter. Ten Maxell patents are pending on this unique mechanism, designed to protect the internal environment and to lock tightly even when accidentally jarred or dropped.

But MF2-HD Super RD superiority doesn't stop there.

New high-energy magnetic particles improve the signal-to-noise ratio by 10%, thus enhancing drive compatibility. The "superuniform" particle dispersion achieved in our Neo-Cross Linkage Magnetic Layer delivers extremely stable output. And an improved HB lubricant and HT binder helps keep both disk and drive head free from foreign particles.

And the distinctive black shell, precision-molded of a super-flexible composite material, offers excellent drive compatibility and improved head-to-disk contact.

The Gold Standard. Now in black, and ready for your most demanding applications.

maxell®
THE GOLD STANDARD

Maxell Corporation of America, 22-08 Route 208, Fair Lawn, NJ 07410