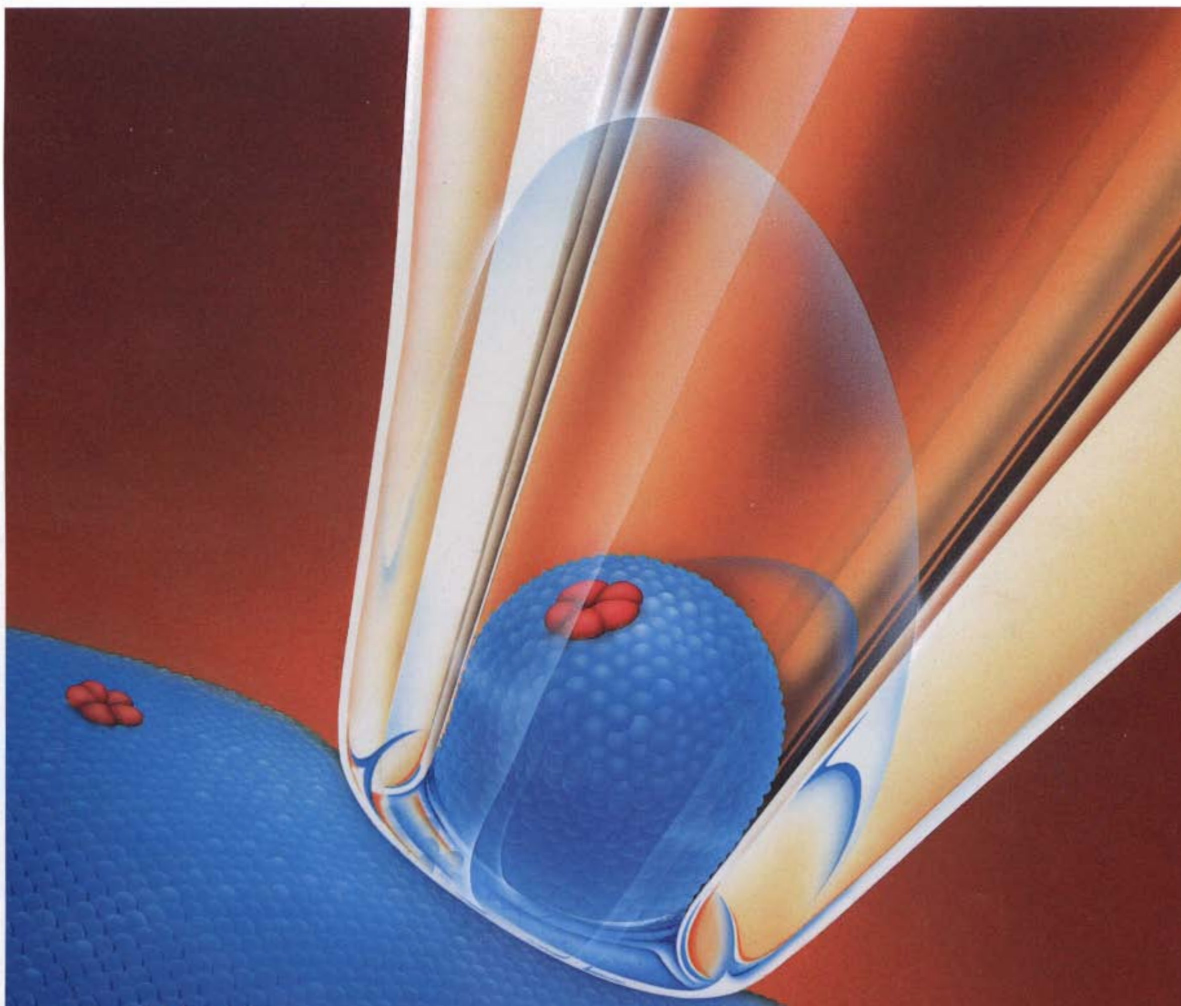
SCIENTIFIC AMERICAN

MARCH 1992
$3.95

Can light beams catch a gravity wave?

Explaining the clumpy structure of the universe.

Spider silks: nature's high-tech fibers.



Tiny pipette *isolates a pore-forming protein that allows signals to pass through cell membranes.*

WE CAN'T PROVE T ARE BETTER DRIVERS, MAYBE THER

Ask the owners of one of the current crop of new luxury automobiles why he or she bought the car he or she bought, and you'll no doubt get a variety of answers.

Some will probably cite the image and prestige of the automaker. Others may wax poetic about the car's quiet operation and fine craftsmanship. Still others may offer up the concept that a car is just a car is just a car, so why not buy the one that costs less?

"You sit up high in a BMW, there's glass all around. It's easy to see, you've got lots of visibility to take in what's going on around you. I like that." Don Orman, Owner 1969 BMW 2000A

While these are all sound reasons for buying a home appliance, BMW believes that a car is not just a car, and a car with a premium price should do more than simply improve the driver's image.

It should improve nothing less than the driver's ability.

Our owners seem to agree.

"When I took the car for a test drive, I knew the difference. I could feel it in the steering wheel...we were one." Dr. Jim Dacey, Owner 1980 BMW 320i

"The BMW brakes, accelerates and handles so well, it will correct your mistakes much quicker than a lot of other vehicles." Victor Bruno, owner BMW 850i Coupe.

And when you consider that from the beginning BMW has looked upon the driver as an integral, fully functioning part of the car itself, the human part of the equation which completes the car's mechanical system, it becomes clear rather quickly why a BMW owner confronted with the same question of why

"The car gives you a fair understanding of where it's at, at all times. It constantly gives you feedback through the steering wheel and other controls. It's very predictable." Jeff Parks, Owner 1988 BMW M6

this car versus that car would say, "When I get into the car, it's like putting on a leather glove. I become united with the car and it becomes an extension of myself." Howard Rogers, owner 1991 BMW M5.

Improving the driver's ability means that every BMW design begins not with a clean sheet of paper, but instead, with the driver.

Research conducted at the University of Freiburg, for

"The car feels strong. It handles well. It doesn't slip and slide. It's easy to drive. It's enjoyable to drive." Monica Dragul, Owner 1987 BMW 325ES

example, gives dramatic importance to the kind of steering response that should be designed into your car.

This research concludes that typical emergency corrections are made not with a conscious thought but by automatic, subconscious

HAT BMW OWNERS BUT IF THEY BELIEVE IT, E'S A REASON.

reflexes. And therefore that the single most meaningful source of information for the driver concerning the behav-

▲ "In a BMW, you drive the car. It does not drive you." Victor Bruno, Owner 1991 BMW 735i/850i

ior of his automobile, and the condition of the road surface is not, as generally supposed, his eyes, but rather, his steering wheel.

So rather than deprive the driver of road feel — as do the "dead" steering systems found in many of today's

▶ "No matter how good a driver you are, it's hard to be 100% focused on driving all of the time...having a car that can react when you demand it is critical in the traffic I drive in." Melissa Hilton-Silver, Owner 1992 BMW 325i

passive, auto-piloted luxury sedans — BMW's are designed with a positive offset steering geometry to better connect the driver to the functioning parts of the suspension system through the steering wheel itself.

Thus providing him with all of the information needed to react instantly and confidently whenever the occasion happens to arise.

In other words, "The car is so connected to you that it barely needs any stimulus beyond your own thoughts to respond in the correct manner. That leaves more of your conscious energy available to apply to other areas of safety when you're driving." Jeff Parks, owner 1988 BMW M6.

Inside a BMW, any BMW, you'll find a driving compartment designed with the understanding that the driver is more than merely human luggage.

Driver physiology and the critical interrelationship between steering wheel angle, seat location, visual position, pedal direction and pedal pressure have all been studied for thousands of hours by designers, engineers, and medical doctors to optimize control under all driving conditions, while at the same time, greatly reducing fatigue.

Fulfilling what may very well be the most obvious way to improve a driver's ability.

◀ "These cars have a remarkable ability to make the driver feel that the car is almost human. The car works with you.... It practically talks to you through the suspension." Howard Rogers, Owner 1991 BMW M5

Keeping the driver awake.

To receive literature, or for the location of the authorized BMW dealer nearest you, call 800-334-4BMW.

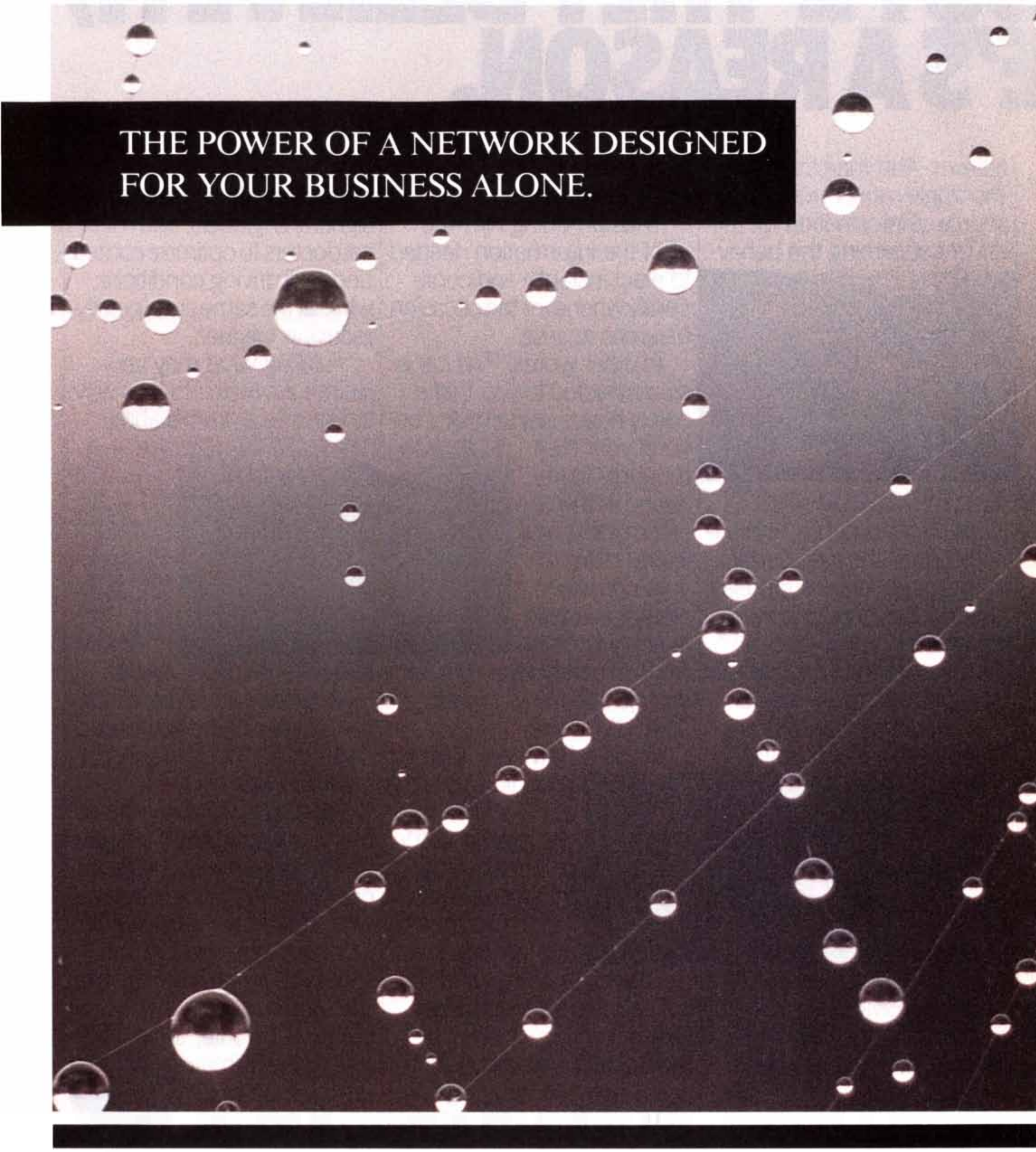
We also invite you in for a thorough test drive. It will give you an opportunity to discover what BMW owners mean when they say things like, "It just feels different." Or, "It spoils you for driving any other car." And of course, our own personal favorite, "It really is the ultimate driving machine."

THE ULTIMATE DRIVING MACHINE.®

For the growth of your business, the prosperity of your business, there is no more powerful tool than the telecommunications network.

And at GTE, there is nothing vague about the power we are putting into your hands when we design that network for you.

For one of the nation's leading retailers, we are now implementing one of the largest satellite-based private



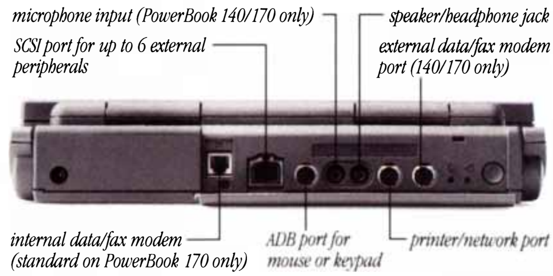
SCIENTIFIC AMERICAN

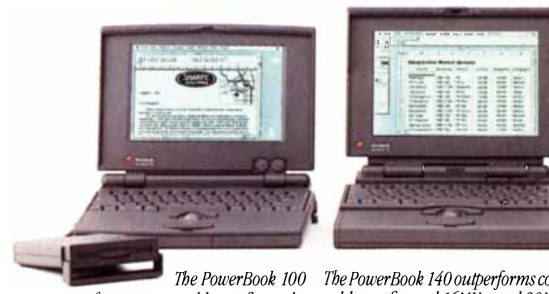
March 1992 Volume 266 Number 3

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017. Copyright © 1992 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. Canadian GST No. R 127387652. Subscription rates: one year $36 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. and Canada 800-333-1199; other 515-247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537.

DEPARTMENTS


Thursday, June 4
Macintosh

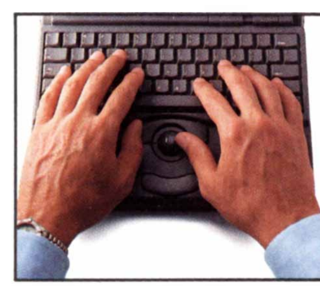




The PowerBook 100 outperforms comparably configured 286 PCs running Windows 3.0.**

The PowerBook 140 outperforms com rably configured 16MHz and 20MH 386sx PCs running Windows 3.0.**

The AppleTalk® Remote Access software (included) gives you access to your Macintosh personal computer via phone line – so you can use all your files or e-mail, or an entire AppleTalk network, from wherever you are.



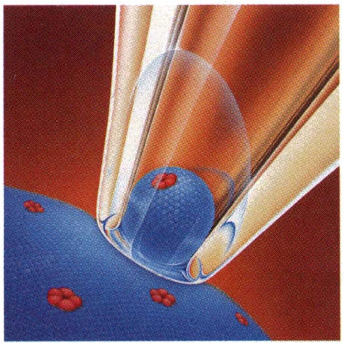
A trackball and palm rests make the PowerBook extraordinarily comfortable for both left- and right-handed people to use.

PowerBook offers a new level of compatibility – its built-in Apple SuperDrive™ disk drive reads from and writes to Macintosh and MS-DOS formatted disks. Add SoftPC and you can run MS-DOS programs, too. (You can even run MS-DOS and Macintosh programs at the same time.)

PowerBook	Weight	Memory	Hard Disk	SuperDrive	Processor	Internal Modem
100	5.1 lb.	2MB	20MB	Optional ext.	16 MHz 68000	Optional
140	6.8 lb.	2/4MB	20/40MB	Internal	16 MHz 68030	Optional
170	6.8 lb.	4MB	40MB	Internal	25 MHz 68030/ 68882 coprocessor	Standard

All PowerBook models expandable to 8MB RAM.

For the name of your nearest authorized Apple reseller, call 800-446-3000, ext. 700. *To run MS-DOS, all you need is a program called SoftPC and an are based on a 1991 independent research study conducted by Ingram Laboratories that tested a variety of personal computers running applications registered trademarks, and PowerBook and SuperDrive are trademarks of Apple Computer, Inc. MS-DOS is a registered trademark and Windows is a tr



THE COVER painting depicts the patch clamp technique, in which a microscopically thin glass tube traps proteins embedded in a living cell's membrane. Researchers long suspected that porelike proteins regulating the flow of ions participated in the generation of nerve impulses and the control of secretion and other processes. Only since the advent of the patch clamp technique has it been possible to analyze the function of individual ion channel molecules (see "The Patch Clamp Technique," by Erwin Neher and Bert Sakmann, page 44).

THE ILLUSTRATIONS

Cover painting by Tomo Narashima

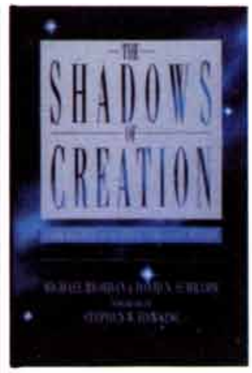
Page	Source
35	Christopher Pillitz/Matrix
37	Chris Niedenthal/Sygma
38	Paula Allen/Matrix (*top*), A. Keler/Matrix (*bottom*)
45	Bert Sakmann
46-50	Dana Burns-Pizer
51	Julio M. Fernandes, Mayo Foundation (*left*), W. Almers, University of Washington (*right*)
52-53	M. J. Geller and J. P. Huchra, Harvard-Smithsonian Center for Astrophysics
54-55	Michael Goodman
56-57	David N. Spergel and Neil G. Turok
58	Marc Skinner
59	David N. Spergel and Neil G. Turok
63	IBM Almaden Research Center
66	Anglo-Australian Telescope Board (*top*)
70	From *Charlotte's Web*, by E. B. White, ©1952, 1980 E. B. White; illustrations © Garth Williams; reprinted by permission of Harper-Collins Children's Books
71	Fritz Vollrath
72-73	Ian Worpole
74-75	Patricia J. Wynne
76	Fritz Vollrath and Donald Edmonds, University of Oxford (*top*), Brent Opell, Virginia Polytechnic Institute (*upper bottom*), Hans Peters, University of Tübingen (*lower bottom*)
78-79	Jerry Silverman
80	Jerry Silverman (*bottom*)
81	Jerry Silverman
82	Jerry Silverman (*top*), Kitt Peak National Observatory (*bottom*)
83	Jerry Silverman
84	Timothy Taylor (*left*), National Museum, Copenhagen (*right*)
85	National Museum, Copenhagen; Kitt Weiss (*photo*)
86	Joe Lemonnier; insets (*clockwise from top, left*): National Museum, Copenhagen; Mikhail Rostovtseff; British Museum (*photo:* Timothy Taylor); National Museum, Sophia; Society of Antiquaries, London (for Sark Island)
87	Erling Benner Larsen, Royal Academy of Fine Arts, Copenhagen
88	National Museum, Copenhagen (*left*), National Museum of India (*right*)
89	National Museum, Copenhagen (*top*), Timothy Taylor (*bottom left*), Scala/Art Resource (*bottom center*), British Museum (*bottom right*)
90-92	George Retseck
93	John Long; courtesy of Peter Michelson, Stanford University
94-95	George Retseck
96	Ken Rogers/Black Star (*top*), Erik Leigh Simmons (*bottom left*), Long Photography (*bottom right*)
97	George Retseck (*left*), Alex Abramovici, California Institute of Technology (*right*)
98	Max Planck Institute for Quantum Physics, Garching
99	Ray Idaszak and Donna Cox, National Center for Supercomputing Applications
113	Barron Storey

SCIENTIFIC AMERICAN®

Established 1845

MATHEMATICA IN ACTION
Stan Wagon
Stan Wagon guides the reader through the spectrum of Mathematica functions—exploring the program's enormous power, its tremendous graphics capabilities, and its ability to create unprecedented insights into mathematical materials—from standard computations to the frontiers of mathematics.
419 pages, illus., **$29.95** paper
(ISBN 2202)
$39.95 cloth (ISBN 2229)
Your price $31.99

Save 20% on clothbound books from W. H. Freeman and Company



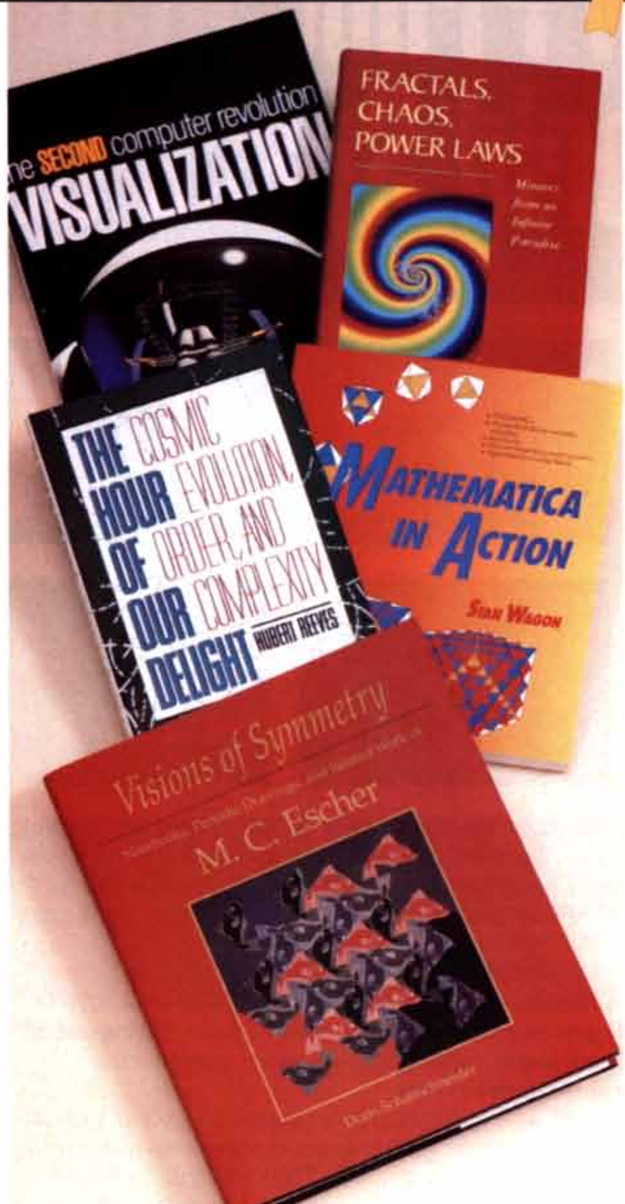
THE SHADOWS OF CREATION
Dark Matter and the Structure of the Universe
Michael Riordan and
David N. Schramm
Foreword by *Stephen W. Hawking*
The authors explore the mystery of dark matter—the more than ninety percent of the universe that is unseen and unknown. Why do cosmologists say this "missing mass" must exist? What is it? What do recent discoveries tell us about the birth and structure of our universe? "Fascinating...."—Isaac Asimov
278 pages, illus., $18.95 cloth
(ISBN 2157) **Your price $15.59**

Science—Explore

FRACTALS, CHAOS, POWER LAWS
Minutes from an Infinite Paradise
Manfred Schroeder
A distinguished physicist traces the development of the concept of self-similarity in mathematics and reveals the fascinating implications for the nature of physical reality. Schroeder explores self-similar symmetries in physics, chemistry, music, and the visual arts, aided by more than 200 illustrations, including 8 pages in full color.
432 pages, illus., $32.95 cloth
(ISBN 2136)
Your price $26.59

VISIONS OF SYMMETRY
Notebooks, Periodic Drawings, and Related Work of M.C. Escher
Doris Schattschneider
This beautiful volume presents M.C. Escher's theory of symmetry and the interplay of art and science. The book features the artist's complete set of symmetry drawings—about 80 of which have never before been published—reproduced in full color. It also includes the two notebooks in which Escher outlined his theory of symmetry.
354 pages, illus., $42.50 cloth
(ISBN 2126)
Your price $34.00

VISUALIZATION
The Second Computer Revolution
Richard Mark Friedhoff
Visualization is the first book to describe in nontechnical language all the technologies underlying the generation and manipulation of computer images. An indispensable resource for learning how phenomena are being depicted today, *Visualization* also reveals how the computer enhances new kinds of thinking in the arts, architecture, science, medicine, engineering, and communications.
216 pages, illus., **$25.95** paper
(ISBN 2231)

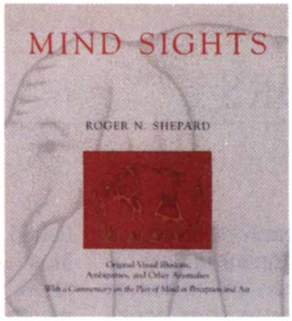


THE HOUR OF OUR DELIGHT
Cosmic Evolution, Order, and Complexity
Hubert Reeves
One of France's most distinguished researchers examines the evolution of our universe from both scientific and philosophical perspectives. This thoughtful analysis, a bestseller in France, won the prestigious Prix Blaise Pascal.
246 pages, illus., $17.95 cloth
(ISBN 2220)
Your price $14.59

SCHIZOPHRENIA GENESIS
The Origins of Madness
Irving I. Gottesman
"For anyone who has ever had an interest in understanding schizophrenia, [this] is the definitive source book. It is a clearly written book that is, nevertheless, so packed with information about all aspects of schizophrenia... that readers will find themselves going back again and again to absorb the full fare that is offered."—L. Erlenmeyer-Kimling, Columbia University
297 pages, illus., **$14.95** paper
(ISBN 2147)

THE VANISHING MIND
A Practical Guide to Alzheimer's Disease and Other Dementias
Leonard L. Heston, M.D., and
June A. White
The authors provide a concise, authoritative introduction to the dementias, including Alzheimer's disease. First published under the title *Dementia*, this greatly expanded version covers the probable causes, characteristics, and treatment of these debilitating diseases to help the afflicted and their families plan for proper care.
208 pages, illus., **$13.95** paper
(ISBN 2192)
$22.95 cloth (ISBN 2131)
Your price $18.59

MIND SIGHTS
Original Visual Illusions, Ambiguities, and Other Anomalies, with a Commentary on the Play of Mind in Perception and Art
Roger N. Shepard
Shepard introduces us to his wonderfully original visual tricks as he discusses the origins of his scientific and artistic work and reflects on the nature of art, perception, and the mind. *Mind Sights* is an elegant introduction to visual perception.
228 pages, illus., **$14.95** paper
(ISBN 2133)



CAN YOU WIN?
The Real Odds for Casino Gambling, Sports Betting, and Lotteries
Michael Orkin
Michael Orkin, statistician and expert on games of chance, shows how to calculate the *real* odds of winning lotteries, casino games, and sports bets in an entertaining and thorough manner. "Required reading for anyone interested in sports betting."—Michael "Roxy" Roxborough, head oddsmaker, Las Vegas
181 pages, illus., **$11.95** paper
(ISBN 2155)

EXPLORING COMPLEXITY
An Introduction
Grégoire Nicolis and *Ilya Prigogine*
This provocative book chronicles unexpected discoveries in nonequilibrium physics that are changing our understanding of complex phenomena. It offers innovative views on nonlinear systems, complexity, and chaos that might further challenge the laws of science.
313 pages, illus., $24.95 cloth
(ISBN 1859)
Your price $19.99

the Frontiers

Our Most Popular Reference Books

MOLECULAR CELL BIOLOGY
Second Edition
James Darnell, Harvey Lodish, and *David Baltimore*
Scientific American Books
1100 pages, $54.95 cloth
(ISBN 1981)
Your Price $43.99

CALCULUS
Harley Flanders
984 pages, $49.95 cloth
(ISBN 1643)
Your price $39.99

ECOLOGY
Third Edition
Robert E. Ricklefs
912 pages, illus., $44.95 cloth
(ISBN 2077)
Your price $35.99

THE FRACTAL GEOMETRY OF NATURE
Benoit Mandelbrot
460 pages, illus., $39.95 cloth
(ISBN 1186)
Your price $31.99

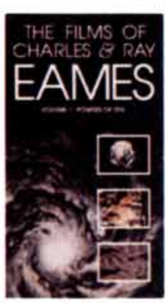
ORNITHOLOGY
Frank B. Gill
680 pages, illus., $49.95 cloth
(ISBN 2065)
Your price $39.99

FUNDAMENTALS OF HUMAN NEUROPSYCHOLOGY
Third Edition
Bryan Kolb and *Ian Q. Whishaw*
912 pages, illus., $41.95 cloth
(ISBN 1973)
Your price $33.59

BIOCHEMISTRY
Third Edition
Lubert Stryer
1089 pages, illus., $54.95 cloth
(ISBN 1843)
Your price $43.99

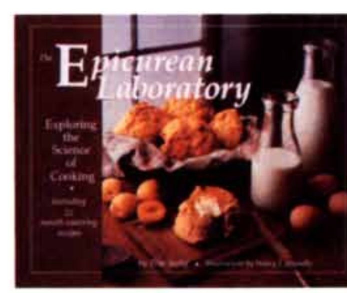
For Your Video Library

FRACTALS
An Animated Discussion with Edward Lorenz and Benoit B. Mandelbrot
H.-O. Peitgen, H. Jurgens, D. Saupe, and *C. Zahlten*
This 63-minute video turns the Mandelbrot set and the Lorenz attractor into visible objects as their discoverers discuss the details of their work. It also features new computer-graphic illustrations of chaos and self-similarity, as well as music composed according to fractal principles.
63-minute VHS video $59.95 (ISBN 2213)
Your price $47.96



POWERS OF TEN
The Films of Charles and Ray Eames
This famous, award-winning film takes us on a trip whose distance spans forty powers of ten, out to the cosmic reaches of the universe (10^{24}m) and back to the heart of the atom (10^{-16}m). The cassette also includes "A Rough Sketch for a Proposed Film Dealing with the Powers of Ten," a remarkable film in its own right. Introduction by Gregory Peck.
21-minute VHS video $39.95 (ISBN 5029)
Your price $35.00

THE EPICUREAN LABORATORY
Exploring the Science of Cooking
Tina L. Seelig
The art of cooking is also a fascinating science, whose principles explain why gelatin works in a salmon mousse; why bread rises; and why the tongue burns from red peppers. Each culinary principle discussed is followed by a mouth-watering, kitchen-tested recipe that demonstrates the chemistry behind the dish. Includes 22 recipes.
163 pages, illus., **$15.95** paper
(ISBN 2162)



Your satisfaction matters. Return any book in good condition within 15 days of receipt and we'll send you a prompt refund.

To order, mail this coupon or a copy to:

W.H. Freeman and Company
4419 West 1980 South
Salt Lake City, UT 84104

All orders must be prepaid in U.S. dollars only.
☐ I enclose my check or money order made payable to W.H. Freeman and Company.
☐ Visa ☐ MasterCard Exp. date ________

Acct. no. ________

Signature ________
(Credit card orders must be signed)

Name ________
(Please print)

Address ________

City/State/Zip ________

Qty.	Author	ISBN	Special Price	Total

Postage & handling* (Add $1.95 for first book, $1.25 for each additional book) $ ________

NY, CA, and UT residents add sales tax $ ________

TOTAL $ ________

*Canadian residents add $2.25 for postage and handling; $1.50 for each additional book. Please add 7% GST. Allow 4 weeks for delivery. All orders should be prepaid in U.S. dollars

4913

LETTERS TO THE EDITORS

Life Science, Old and New

Perhaps I was not the only reader to observe with fascination the juxtaposition of "Antichaos and Adaptation," by Stuart Kauffman, and "Smart Genes," by Tim Beardsley [SCIENTIFIC AMERICAN, August 1991]. Kauffman's article was the first of its kind I have ever seen. It brought into focus for me that for decades, microbiologists have been preoccupied with a bits-and-bytes view of biological processes. I have seen little evidence that they would acknowledge the importance of the architectural approach Kauffman is pursuing.

Modern microbiology is an undeniable tour de force with all its genetic engineering mechanics, but when it comes to actual illumination of the life process, much of its thinking misses the point. Maybe *Scientific American* can serve as an early crucible for bringing the architectural and mechanical views together.

RICHARD MELMON
Objective Software, Inc.
Palo Alto, Calif.

Jarring Jargon

In her October 1991 essay on "Quantum English," Anne Eisenberg fails to mention one of the worst sources of misuse of technical terms: scientists themselves. I have yet to see definitions of such terms as modality, methodology or dynamical that distinguish them in any significant way from mode, method or dynamic, respectively. These terms appear to be no more than the pitiful efforts of some scientists to sound more "scientifical." More pitiful yet is that these terms have become almost universal instead of being thrown on the linguistic trash heap where they belong.

If Eisenberg really wants to illustrate how scientific terms have become whimsical, she should turn to the naming of fruit fly mutations. Quarks and fuzzy logic may be whimiscal, but naming fruit flies "tricky-dick," "hedgehog," "shaven•baby" and "killer of prune" is lunacy. (All these names come from characteristics of the flies.)

I'd like to offer organized science a deal: I will stop using "quantum leap" if you will stop using "forward" to denote the prefatory remarks to your books and papers. Much obliged.

K. A. BORISKIN
Bellingham, Mass.

GARETH PENN
San Rafael, Calif.

Trail of the Wolf

"Howls of Dismay," by John Rennie ["Science and the Citizen," SCIENTIFIC AMERICAN, October 1991], correctly indicates that my studies of fossils argue against a hybrid origin for the red wolf. An even better argument is provided by a large series of skulls collected in southeastern North America between about 1890 and 1930. Had the red wolf actually been a hybrid between the gray wolf and the coyote, this series should have shown a complete blend, without geographic or morphological lines separating the three entities. In fact, such blending is restricted to central Texas, and even there it is limited to red wolves and coyotes. Hybridization with the coyote spread but never totally engulfed the red wolf. I recently examined skulls of animals from the current red-wolf breeding program and found them statistically inseparable from those of the original species.

RONALD M. NOWAK
Office of Scientific Authority
U.S. Fish and Wildlife Service
Washington, D.C.

After Thoughts

In his essay "Animal Thinking" [SCIENTIFIC AMERICAN, November 1991], Donald R. Griffin apparently finds it necessary to set up a behaviorist straw person to make his point. The problem is that the behaviorism he criticizes was beginning to be replaced as early as 1938, with the publication of B. F. Skinner's *The Behavior of Organisms,* and is now of historical interest only.

Any reading of modern behaviorism (no, it is not dead) shows quite clearly that behavior analysts do not reject events just because they are unobservable. They deal with unobservables as in the other natural sciences: they interpret the events according to established scientific laws and principles. Most such analysts simply attempt to show that the behaviors from which mental events are inferred can be explained without speculation about mind, thoughts or consciousness.

HENRY D. SCHLINGER
Department of Psychology
Western New England College
Springfield, Mass.

Having a (Bucky)Ball

For readers like me, an article such as "Fullerenes," by Robert F. Curl and Richard E. Smalley [SCIENTIFIC AMERICAN, October 1991], is the ultimate reward. Few researchers make the effort to write about what has been learned over time. Few bother to assess the impact of scientific branches or twigs for which they have not been personally responsible. Of those who do try, almost none can match the balance of detail and personal discovery delivered by Curl and Smalley.

To both those scientists and *Scientific American:* Thank you! Please stay on the ball(s) and get back to us with an update soon.

Candlelight will never be the same.

PHILIP W. TOWER
Dow Chemical Company
Midland, Mich.

ERRATA

In the caption on page 55 of "The Early Life of Stars," by Steven W. Stahler [July 1991], the tracks of stellar luminosity and temperature are mistakenly called Hayashi tracks. They are actually Iben tracks, named for Icko Iben, Jr., of the University of Illinois, who first computed them. A Hayashi track is one feature of an Iben track.

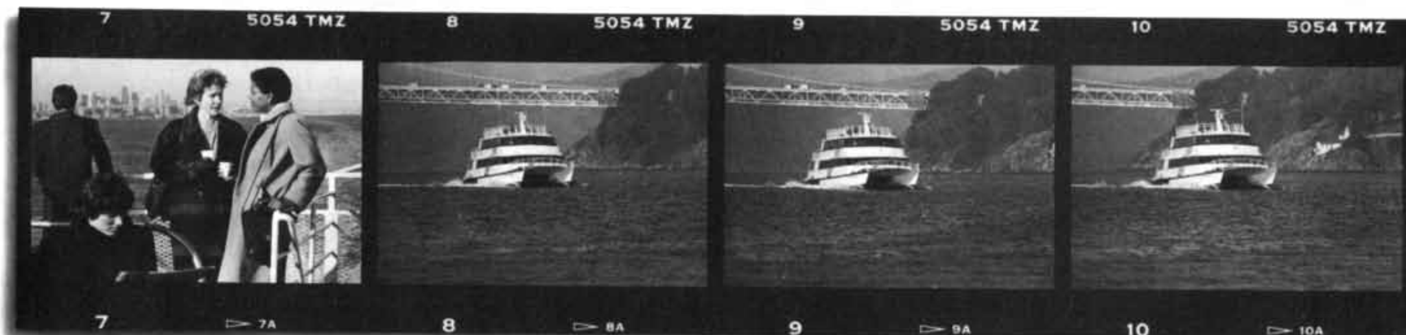
The caption on page 49 of "Firearms, Violence and Public Policy," by Franklin E. Zimring [November 1991], should have read, "Handguns were involved in half of the 23,000 homicides in the U.S. last year."

On page 102 of "Chemical Fuels from the Sun," by Israel Dostrovsky [December 1991], an exajoule should have been defined as one billion billion (10^{18}) joules. In the first equation in the box on page 104, the term "$1/SO_2$" should read "$1/2\ O_2$."

"Steve called from DC."

"And?"

"The presentation went better than we could have hoped."



"Great!"

"He made some last-minute changes, though. Turned up some better data."

"What did he do, hire an expert?"

"Said he did it himself."

"Himself? How?"

The answer is in DIALOG OnDisc.

For complete, precise, current information—the kind that can give your business the edge—thousands of companies count on Dialog®: the world's first and largest electronic library. Now, DIALOG OnDisc® offers an expanding choice of Dialog databases in convenient CD-ROM format. Our easy menu-mode search option helps you find answers quickly. And our exclusive disc-to-online connection makes it easy to get updates directly from Dialog.

Talk to your organization's Information Specialist, or call for a free kit on DIALOG OnDisc offerings for your industry.

1-800-3-DIALOG

Outside U.S., 415-858-3785. Fax 415-858-7069.

DIALOG OnDisc Information Tools: Over 35 databases, including full-text newspapers, covering general business, finance, law, engineering and electronics, health and bioscience. CD-ROM format allows unlimited searching at a fixed cost. Updated discs issued regularly. Powerful, easy-to-use search software. Convenient and easy DIALOG OnDisc-to-Dialog online connection.

DIALOG INFORMATION SERVICES, INC.

A KNIGHT-RIDDER COMPANY

50 AND 100 YEARS AGO

MARCH 1942

"While it is not news to tell of the inroads that war-time production is making on civilian materials of all kinds, it seems proper to record here some things that will influence the future course of this magazine. The paper will incline more toward a natural or yellow shade. This is due to restrictions on civilian uses of chlorine, the chemical that is used to bleach the pulp from which paper is made. Also, as will be noticed, the issue now in your hands is set in a different type from that of previous numbers. This new type, while just as readable as the former font, is a bit more compact; as a result it is possible to get about 10 percent more words on a page."

"Although the living eye automatically focuses objects at varying distances, optical engineers have looked askance at many ingenious proposals to accomplish this purpose in a photographic lens. But a four-element motion picture lens has now been developed by Bausch & Lomb in which one double-concave element is electronically oscillated on its axis. The oscillating element is confined to a movement of three-tenths of a millimeter, but the oscillations are at the rate of 23,200 times per minute, thus continuously altering the focus so that all objects are uniformly in register from four feet to infinity. Although all objects are slightly softer in focus than with conventional lenses, many photographers regard this as an improvement."

"Bischoff, one of the leading anatomists of Europe, thrived some 70 years ago. He carefully measured brain weights, and after many years' accumulation of much data he observed that the average weight of a man's brain was 1350 grams, that of a woman only 1250 grams. This at once, he argued, was infallible proof of the mental superiority of men over women. Throughout his life he defended this hypothesis with the conviction of a zealot. Being the true scientist, he specified in his will that his own brain be added to his impressive collection. The postmortem examination elicited the interesting fact that his own brain weighed only 1245 grams."

SCIENTIFIC AMERICAN

MARCH 1892

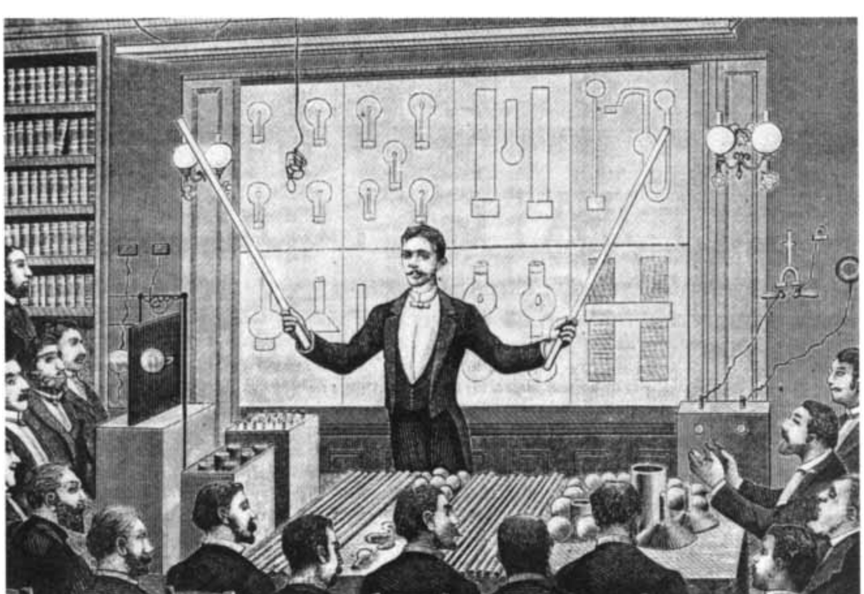
"Compressed air is, perhaps, the chief rival of electrical transmission. It is at present used mainly in mines, where it is still a very successful rival of electricity, but from present appearances it is likely that it will gradually be replaced by the latter method. In Paris there is a large central station for the distribution of compressed air, and it seems to be in successful operation. From a paper by Professor Unwin it appears that the transmission of power by compressed air is practical to a distance of at least 20 miles. It seems that 10,000 horse power can be transmitted 20 miles in a 30-inch main at 132.3 pounds per square inch with a loss of pressure of only 12 per cent."

"If you take a good conductor like copper, and run the temperature down, its resistance almost disappears at very low temperatures. Hundreds of degrees below zero, copper is almost a perfect conductor. If you heat it up, it becomes more and more resisting. Let us take glass—a good insulator—or any insulating material, and run its temperature up, it loses its insulating power, and if we run it up until it gets to red heat, it approaches a conductor; so that all substances are conductors when they are hot enough."

"The philosopher known to fame as Sir William Thomson has joined the ranks of the British aristocracy under the new name of Lord Kelvin. This lord lately took his seat in the House of Lords, being introduced by scientific nobleman Lord Rayleigh."

"C.R.W. asks: What power prevents the bottom course of brick in our large structures from crushing with the tremendous weight above it? *A: A single common red brick, when properly laid in strong cement, is equal to a load of 12 tons, and it will require a column of cement-laid brick nearly 10,000 feet high to crush the bottom course.*"

"Mr. Nikola Tesla terminated his lecture with a few experiments in the illumination of tubes of rarefied gases without wires or electrodes, the tubes being simply placed in the periodical electrostatic field produced between one of the insulated poles of the bobbin on the one hand and an insulated metallic plate placed above the experimenter and communicating with the other pole of the bobbin on the other hand. The accompanying figure shows one of these experiments, in which Mr. Tesla is producing the illumination of two tubes at once."



Mr. Tesla lecturing in Paris

They look alike and drive alike. Camouflaged by their sameness, they vanish in parking lots, lost among mirror images of one another.

Cars may well be on their way toward becoming interchangeable. But if they ever get there, they'll do so without the help of the Saab 900, the car no cookie cutter could ever create.

*The Saab 900 Series: From $19,880 to $36,230.**

*The Saab 9000 Series: From $25,465 to $37,615.**

For more information, call 1-800-582-SAAB.

The 900 is what becomes of a car when form follows function instead of fashion.

Take its odd-looking profile. That vaguely hunchbacked shape houses one of the world's most protective steel safety cages. One reason why, based upon actual highway accident reports, the Saab 900 has repeatedly been ranked among the safest cars in its class.

That eccentric profile also houses the largest carrying capacity in its class. Fold down the rear seat, and there's 53 cu. ft. of cargo space, rivaling some station wagons.

But nowhere is the difference between conventional cars and a 900 more apparent than out on the road.

There, its front-engined, front-wheel-drive system provides the superior traction required of a car engineered for Swedish winters. A nimble suspension and tactile steering system unite car, road and driver in the same lively enterprise; no one was ever anesthetized by a Saab 900.

Or shortchanged on amenities. Besides a spirited fuel-injected engine, standard equipment includes a driver's-side air bag, anti-lock brakes, air-conditioning, power windows, central locking—even electrically heated front seats.

The result is a complete car instead of a compromise, combining rather than choosing between the virtues of safety, utility, performance and value—all in a package that's been described as, well, idiosyncratic.

But then cars, like people, are made infinitely more interesting by their idiosyncrasies. A point your Saab dealer will be happy to prove through a test drive of the 900.

SAAB

**WE DON'T MAKE COMPROMISES.
WE MAKE SAABS.℠**

*MSRP, excluding taxes, license, freight, dealer charges and options. Prices subject to change. Prices do not include the 9000 CD Turbo Griffin Edition.

Reflecting Differences

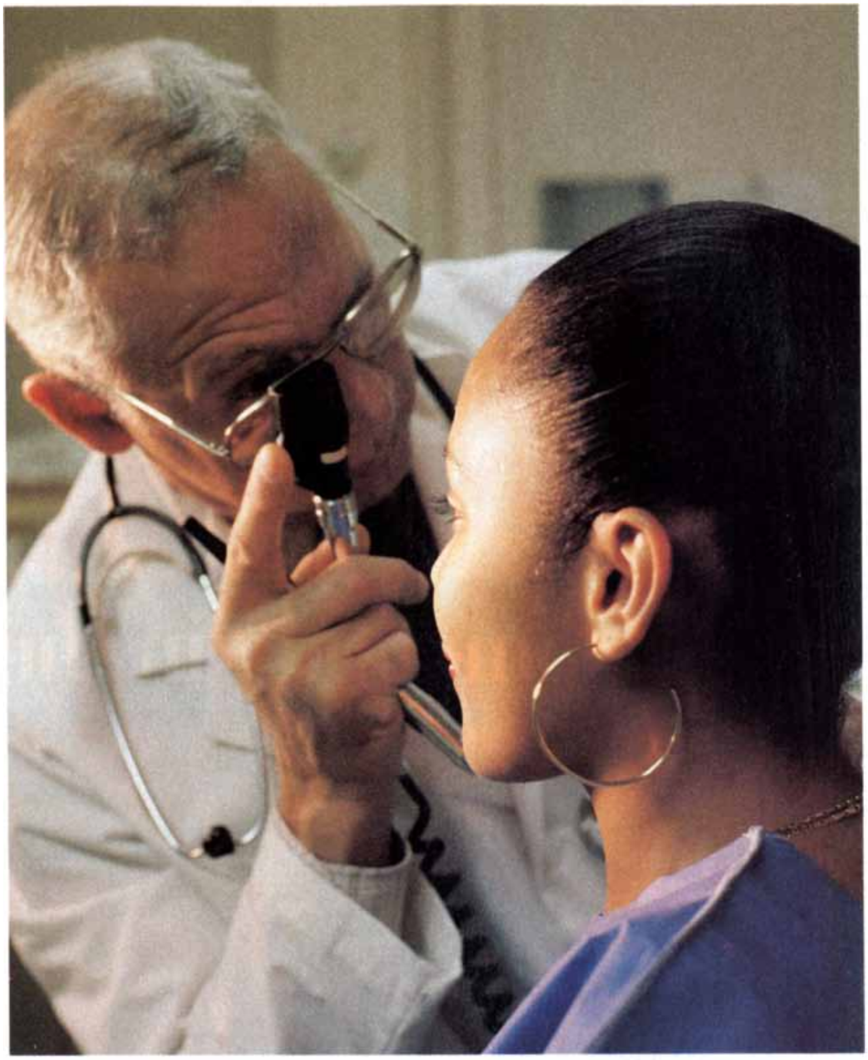
Health care begins to address needs of women and minorities

Until the 1940s, it was medical doctrine that blacks did not suffer heart attacks. Chest pains were often thought to be a symptom of something else—indigestion perhaps. And although heart diseases kill more than 250,000 American women every year, two studies that appeared last July in the *New England Journal of Medicine* pointed out that women are not diagnosed or treated as aggressively as men, unless their condition is dire. "Primary care professionals don't think women get heart disease," says Joanne Howes of the Society for the Advancement of Women's Health Research in Washington, D.C.

Prejudice and preconceptions may well play a role in the different treatment women and minorities receive at the hands of physicians, but health care providers are increasingly recognizing another, significant factor. Despite the risk of accusations of fueling discrimination, researchers are examining differences between men and women, whites and minorities. "I used to think I was going to have rotten tomatoes thrown at me," says Elaine Eaker of the division of surveillance and epidemiology at the Centers for Disease Control, referring to meetings where she discussed differences between the sexes. Now the idea that "women are different, not only physically but psychologically," is more readily accepted, she says.

Some of these changes were initiated several years ago, when congressional hearings brought attention to women's health. A report by the General Accounting Office (GAO) found the National Institutes of Health unresponsive to the medical needs of minorities and women. As a result, the offices of Research on Women's Health and of Minority Programs were created at the NIH.

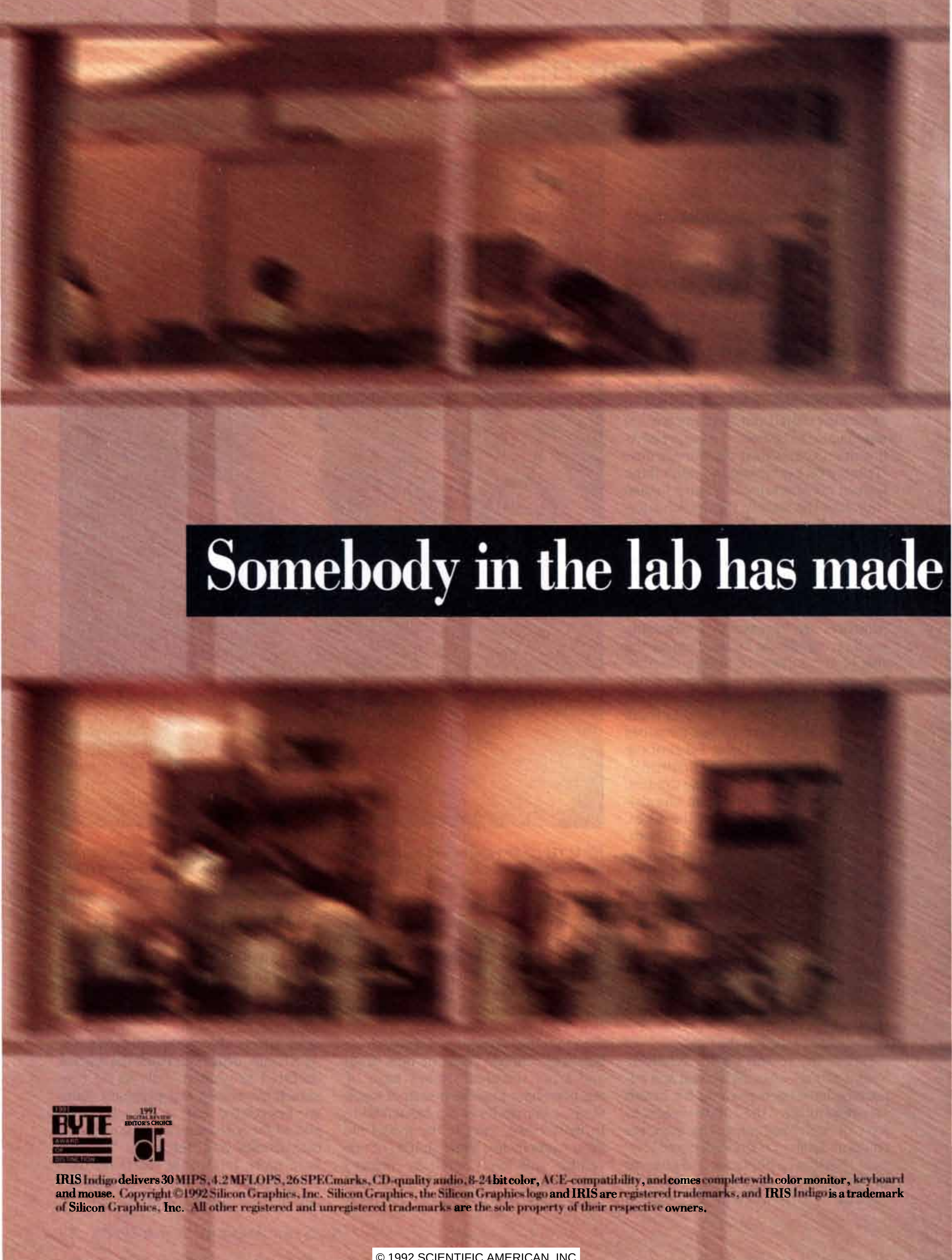
Part of the reason for the relatively abrupt action—after all, women and minorities have been pushing for health care reform since the 1960s—seems to have been demographic. The numbers could no longer be ignored, notes Florence Haseltine, a physician and director of the center for population research at the NIH. By 2010 more than 21 million women will become menopausal. Haseltine says these women have economic clout and are politically savvy. "Something is going on here; women are getting a voice," she says.



HEALTH CARE reform will continue to be slow until "legislators really deal with the problems on all levels," says Irving S. Rust of Planned Parenthood. Photo: Jason Goltz.

Responding to the political directive, the medical community is beginning to redress traditional research imbalances. Clinical trials of new drugs and large research studies have often been based on the notion of sameness—a human body is a human body. In most cases, though, the model has been white males. Many studies of prevention and intervention in coronary heart disease—including the U.S. Physicians Study, which examined the benefit of taking an aspirin a day, and the Multiple Risk Factor Intervention Trial—focused exclusively on men.

In addition, Phase I clinical drug trials have often excluded women of reproductive age because of concerns about liability and the potentially damaging effects of a compound on a fetus. As a result, drugs can reach the market untested in women—only to be consumed by women and pregnant women. Sex-related effects have been reported for some antidepressants, anticonvulsants and sedatives. "Women have more fatal drug reactions than do men, and they seem to be greatest during the reproductive years," says Su-


Somebody in the lab has made
BYTE
1991 EDITOR'S CHOICE
IRIS Indigo delivers 30 MIPS, 4.2 MFLOPS, 26 SPECmarks, CD-quality audio, 8-24 bit color, ACE-compatibility, and comes complete with color monitor, keyboard and mouse. Copyright ©1992 Silicon Graphics, Inc. Silicon Graphics, the Silicon Graphics logo and IRIS are registered trademarks, and IRIS Indigo is a trademark of Silicon Graphics, Inc. All other registered and unregistered trademarks are the sole property of their respective owners.

Tuning in Hot Signals

Creating ghostly pictures from radiated heat is usually work for infrared detectors made from semiconductor materials [see "Infrared Video Cameras," by Jerry Silverman, Jonathan M. Mooney and Freeman D. Shepherd, page 78]. Yet the sensitivity of these detectors is limited.

Researchers at the National Institute of Standards and Technology (NIST) in Boulder, Colo., believe they can significantly improve the sensitivity of infrared detectors by changing the design. At the heart of their nascent system are two unusual components: detectors made from low-temperature niobium superconductors and a spiral antenna no bigger than a grain of sand.

These new detectors, researchers say, should enable them to tune in a wide range of infrared frequencies, in much the way channels are selected on a radio. Moreover, they hope the tunable devices will evolve into a new generation of sensors for analyzing contaminants in the atmosphere or the composition of galactic clouds by detecting their characteristic emissions.

NIST physicists Donald G. McDonald, Erich N. Grossman and Joseph E. Sauvageau have spent the past two years developing the devices. Although superconductors have long promised to be sensitive to infrared radiation, "no one had made a useful infrared detector before," McDonald recalls.

The workers knew that very small superconducting detectors would be highly sensitive to infrared radiation, particularly if the size of the detector was smaller than the wavelength of the incoming radiation, say, about one micron across. (In this way, the incoming photons would make the greatest perturbations in the electronic structure of the superconductor.) The catch: at that size, the detector's ability to absorb radiation suffers.

Radio receivers must also pick up wavelengths too large for their detectors—and do so by using antennas. As a result, McDonald and his colleagues designed a spiral antenna about 60 microns wide. They used conventional lithographic techniques to lay down the gold pattern on top of niobium.

The gold antenna conveys 50 percent of the signal power it receives to the niobium detector—an efficiency comparable to that of a high-quality radio antenna. The superconductor, meanwhile, responds to a sizable range of signals, from five to 30 microns in wavelength. By embedding the detector in "the most sensitive of all amplifiers," a superconducting quantum interference device, or SQUID, McDonald reckons that the overall system will be far more sensitive than current infrared detectors.

That enhanced sensitivity, McDonald says, means that the devices may prove particularly handy for atmospheric or galactic spectroscopy. "Every molecule has a frequency associated with it, often more than one," McDonald points out. To scout out molecular contaminants in the atmosphere, the researchers plan to add an oscillator to the detector-antenna, then hook the system to a telescope. By tuning the oscillator until it is synchronized with incoming frequencies, they should be able to pick out specific molecules.

The researchers are still far from building a complete detection system, which would include an array of hundreds of antennas and superconducting detectors. By the end of the year they hope to have more experimental data on the sensitivity and bandwidth of their devices. But nature already offers them some cause for optimism, they say. Insects, too, rely on spiral antennas for detecting infrared radiation. "So you can ask the question, 'Is infrared detection important?'" McDonald offers. His answer: "It must be, because biological evolution has produced it." *—Elizabeth Corcoran*

san J. Blumenthal, chief of the behavioral medicine program at the National Institute of Mental Health and member of the board of directors of the National Women's Health Resource Center.

Many scientists argue that researchers and companies may have hidden behind the question of fetal damage. "The phantom fetus is not an issue," Haseltine says. "What's wrong with giving an abortion to a woman who gets pregnant during treatment? Abortion is available, and no one wants to use the word."

Some steps have already been taken to assure that future studies include a cross section of society. Research grant applications to the NIH must now include women and minorities or else provide a good reason why they do not. The Food and Drug Administration may loosen its guidelines for some Phase I clinical studies. And the Institute of Medicine is examining some of the ethical concerns about including women of childbearing age in studies. These moves are "already having a very profound effect," Blumenthal notes. "Both groups have not been well represented in research studies, and sharpening the focus on both populations is very important."

Indeed, researchers are just beginning to recognize that women and minorities may have disease etiologies or responses to therapies unlike those of white males. For instance, studies have found that when levels of estrogen fall after menopause, levels of low-density lipoproteins—those correlated with heart disease—rise.

Disorders such as fibroids, which affect a third of all premenopausal women, autoimmune diseases such as lupus, which for the most part afflict black women, and depression, which is twice as common in women than in men, are still not well understood, says Cindy Pearson, program director at the National Women's Health Network. One in nine women dies of breast cancer every year, and black women have a much poorer survival rate than white women. Yet a recent GAO report on breast cancer notes that "there has been no progress in preventing the disease" since 1971.

In addition, women make up the fastest-growing segment of the population with AIDS, Blumenthal notes. Most studies of the disease and of related treatments, however, have been conducted in men. Not all these data can be extrapolated to women. For instance, "AIDS appears to have a more virulent course in women; they die sooner after being diagnosed," Blumenthal says.

Disparities between ethnic groups are not as obvious as those between

men and women. Most discrepancies found in studies can be attributed to environmental factors, such as socioeconomic status and diet. And—except for well-known genetic diseases such as sickle cell anemia and Tay-Sachs—researchers are often careful about raising the issue of race. "There's been historical mischief with putative genetic differences with populations," says Herbert W. Nickens of the Association of American Medical Colleges.

Certain differences, however, seem to be biological. For instance, hypertension in black Americans occurs two to six times more frequently than in whites, says Richard A. Williams of the Minority Health Institute in Beverly Hills, Calif. Blacks are 10 times more likely to die from related complications.

A study released last summer suggested that genes may be responsible for the prevalence of glaucoma in blacks. The condition strikes black Americans four to six times more frequently than it does whites. The trait may also be responsible for different responses to treatment. Jonathan C. Javitt of the Center for Sight at Georgetown University notes that surgeons are now aware that there is a greater tendency in black patients to scar during glaucoma operations. William W. Stead of the Arkansas Department of Health suspects that blacks may have a genetic susceptibility to tuberculosis that causes the disease to progress rapidly. In whites, tuberculosis tends to be chronic rather than acute.

Still, many researchers argue that pervasive changes in the health care system may not be realized until all groups are well represented in the medical and research establishment. Women now make up about 38 percent of medical students—up from 25 percent a decade ago—but they are rarely directors of laboratories, and at present no dean of a medical school is a woman, according to Janet Bickel of the Association of American Medical Colleges. Minorities fare even worse. Blacks between the ages of 20 and 29 make up 13.2 percent of the population but only 5.7 percent of all medical students. "We need more input from minorities," says Irving S. Rust, medical director of the Bronx Planned Parenthood clinic.

The new offices at the NIH say they are working to promote research on and by women and minorities. A study of more than 60,000 postmenopausal women has been designed to answer questions about the effect of diet and estrogen replacement therapy on cardiovascular health and the prevention of breast cancer and osteoporosis. And $1.5 million has been set aside to study hypertension and coronary disease in blacks. "Increased research brings increased hope for the future," Blumenthal says.

—Marguerite Holloway and Philip Yam

Relative Hunger

A recent experiment with a cannibalistic creature has provided experimental support for the extension of Darwin's theory of natural selection known as kin selection. Tadpoles of the spadefoot toad, when fed on whole-animal prey, develop into a specialized form that eats members of its own species. Yet they tend to avoid feeding on their close relatives.

Kin selection was proposed in the early 1960s by the British evolutionist William D. Hamilton. He argued that natural selection should favor genes that make animals act more altruistically toward their close relatives than toward unrelated animals. Aiding relatives could help genes spread through the population because close relatives are likely to carry copies of the same genes.

Hamilton's theory, which he expressed mathematically, is one of the principal planks of sociobiology. It spurred many biologists to examine how animals behave toward their kin, and researchers soon found that animals indeed often behave differently with close relatives than with other individuals. Many creatures, for example, will choose to be near their kin, as opposed to unrelated individuals.

Although these demonstrations have often been taken as evidence for kin selection, Alan Grafen, a theoretical biologist at the University of Oxford who is a scourge of uncritical experimenters, has chastised investigators for not ruling out other explanations. Grafen argued recently that some purported proofs of kin selection were nothing more than demonstrations that animals tend to associate with others that smell like themselves—a phenomenon that could have evolved for any number of reasons. Simply associating with relatives does not prove kin selection, Grafen noted.

David W. Pfennig of Cornell University thinks his experiment with the cannibalistic tadpoles stands up to Grafen's critique and proves kin selection has tempered their appetites. Spadefoot tadpoles not fed on a whole-animal diet, he points out, become omnivores and do not eat members of their species. The omnivorous form, like many other creatures, shows a preference for associating with relatives. The carnivores, in contrast, prefer to associate with nonrelatives. When hungry, they swim around, "nipping" at any other tadpole they come across. A nonrelative unfortunate enough to be nipped is likely to be eaten, but a sibling is usually let go.

Pfennig believes the carnivores use chemical cues to recognize relatives. Because their preference for eating and associating with nonrelatives runs counter to the preference of the omnivorous form, even though they are genetically similar, Pfennig asserts kin selection is the most likely explanation.

Then why isn't cannibalism more common? Although kin selection might prevent the evolution of animals that eat close relatives, it would seem to favor the consumption of weak nonrelatives that would easily be subdued. Pfennig says he has one possible answer. In studies on larvae of tiger salamanders, he found that individuals that indulged in cannibalism were more likely to die from the effects of pathogenic bacteria and parasitic worms. That risk, he suggests, is one good reason not to take up the practice. "I wouldn't eat another human," he says.

—Tim Beardsley

Van B. Hooper
President of Hooper International
and Author of Hooper Software

I'll Show You How To Increase Your Income and Manage Your Money Like a Financial Whiz!

I PERSONALLY GUARANTEE COMPLETE SUCCESS WITH MY FREE UNLIMITED TELEPHONE SUPPORT AND NINETY DAY REFUND OFFER.

"Americas No. 1 Check Writing and Accounting Software For Non Accountants."

FREE 90 DAY TRIAL OFFER!

NOW, SAVE THOUSANDS OF DOLLARS A YEAR IN ACCOUNTING FEES AND AVOID I.R.S. AUDITS WITH THIS USER FRIENDLY CHECK WRITING AND ACCOUNTING SOFTWARE.

Most Executives, Professionals and Business Owners simply don't have the time to learn the necessary accounting principals for effective management of their affairs. On top of it all, the I.R.S. makes endless demands for more comprehensive records.

With Hooper Accounting Packages, you are immediately more productive. All you have to do is follow the easy-to-understand instructions. You control your finances instead of allowing your finances to control you. This enables you to spend more time at what you do best — **Making Money**!

Hooper packages are inexpensive and are written in plain, simple, language. A friendly and helpful support staff is always available to courteously answer any and all of your questions, free of charge. Hooper takes the mystery and awe out of using accounting software!

Take advantage of the 90 day free trial offer today. You have all to gain and nothing to lose. Act now!

"Hooper Software Does Much More Than Quicken, Is More Powerful Than Pacioli and is Immensely More Easy To Use Than Dac Easy."

90 DAY FREE TRIAL.

YOUR GUARANTEE OF SATISFACTION!

Your absolute satisfaction is our number one priority. That's why we invite you to try Hooper Software for a full 90 days without risk or obligation. If you are not completely satisfied, you can return the software for a full and cheerful refund. No questions asked.

WE SHIP ALL ORDERS BY FEDERAL EXPRESS

CHEQUE-IT-OUT™

The user friendly personal accounting software that helps you manage your affairs with complete accuracy. Ideal for executives, professionals and small business.

Sample Listing of Hundreds of Cheque-It-Out Features:

- Works under Windows or DOS.
- Handles and feels like your checkbook.
- Know your account balances at all times.
- No accounting experience necessary.
- Full general ledger.
- Automated Double Entry System.
- Complete range of reports instantly available.
- Intuitive pop up windows.
- Mouse compatible.
- Over 3,000 screen color selections.
- Unlimited transactions.
- Context sensitive help screens.
- Tracks outstanding checks.
- Pop-up calculator.
- Operates in calendar or fiscal year mode.
- Unlimited free telephone support.

Only $69.95 Plus $5 Shipping & Handling

TAKIN' CARE OF BUSINESS!™

The simple-to-use software that will help you succeed in business. It will enable you to operate more efficiently and effectively, improving your bottom line.

Includes All Features of Cheque-It-Out Above, Plus:

- Leading Edge Accounts Receivable Operating & Reporting Package.
- Advanced Accounts Payable Operating & Reporting Package.
- No accounting experience necessary.
- Unlimited free telephone support. And much more.

Only $149.95 Plus $6 Shipping & Handling

PAYROLL-MASTER™

Provides everything you need to professionally manage your payroll function.

Sample Listing of Dozens of Payroll-Master Features:

- Designed to be fully compatible with all other Hooper software.
- Automatic Tax Collections.
- Prints W-2's and 1099's.
- No accounting experience necessary.
- Manages employee vacations, sick time, overtime and holidays.
- Supports up to 1,000 employees.
- Unlimited free telephone support.

Only $69.95 Plus $5 Shipping & Handling

ORDER TOLL-FREE 1-800-525-9200

VISA MasterCard DISCOVER

or send check or money order to:

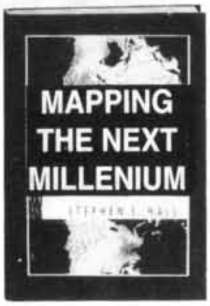
HOOPER INTERNATIONAL, INC.
"The Accounting People"
P.O. Box 49711, Dept. 10
Colorado Springs, CO 80949, USA

All Hooper software operates with IBM PC-DOS or compatible computers.
Specify 5-1/4" or 3-1/2" floppy disk.

TAKE ANY

3 BOOKS FOR ONLY $1 EACH

(Publishers' Prices Shown.)

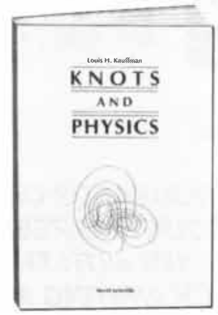


60508 $30.00

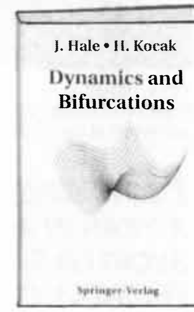
48686 $24.95

58656 $29.50

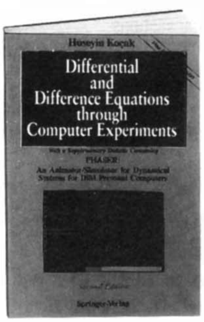
37504 $24.95



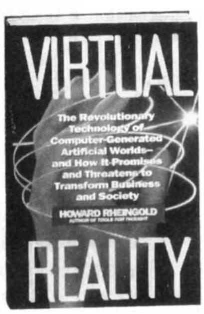
57373 $28.00



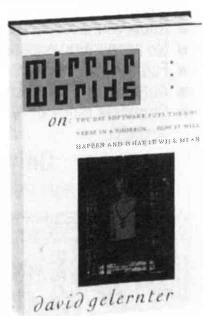
42903-2 $49.00
(counts as 2 choices)



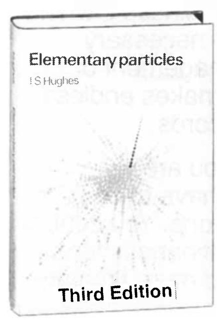
42260-2 $45.00
(counts as 2 choices)



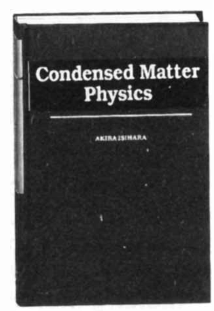
85835 $22.95



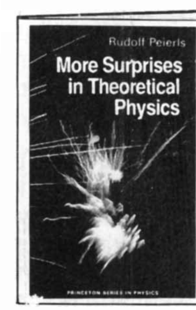
62790 $24.95



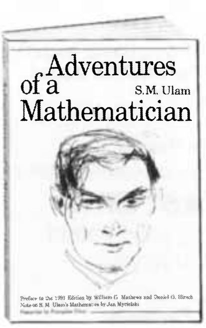
44188 $29.95



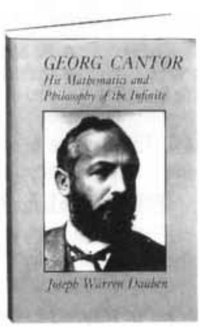
40141-2 $59.95
(counts as 2 choices)



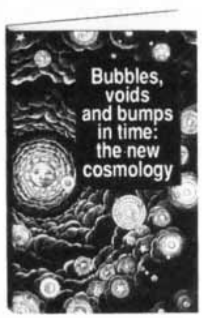
63080 $27.50



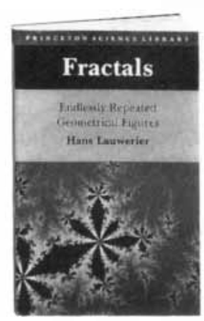
32295 $15.95



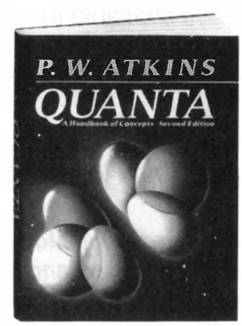
50236 $14.95



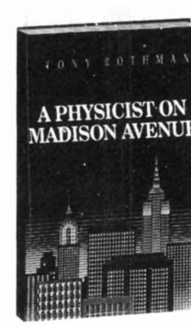
36899 $14.95



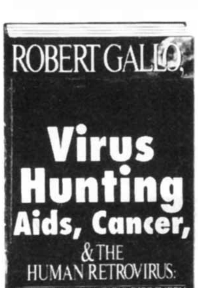
48842 $14.95



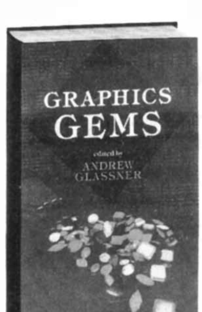
73050-2 $45.00
(counts as 2 choices)



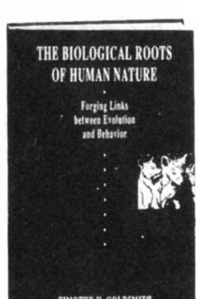
68770 $19.95



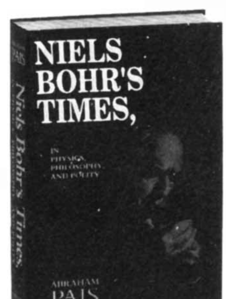
85831 $22.95



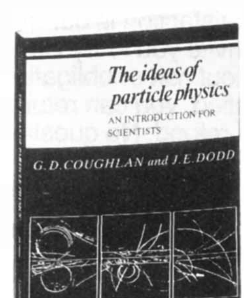
50572-2 $49.95
(counts as 2 choices)



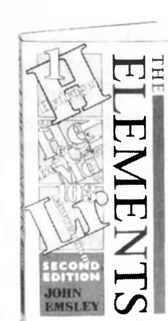
36275 $24.95



64704 $35.00



55083 $29.95

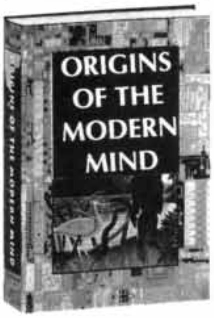


44200 $22.95

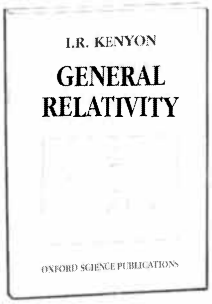
as your introduction to **The Library of Science.**

VALUES TO $94.95!

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

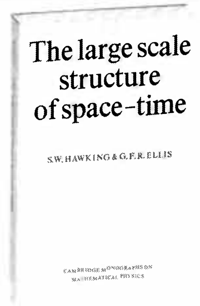


66260 $27.95

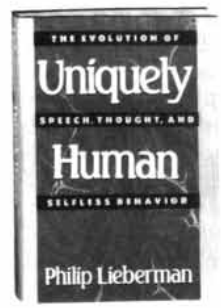


50200 $29.95

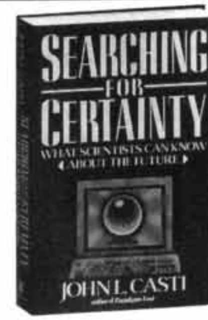
32490-2 $39.95 (counts as 2 choices)



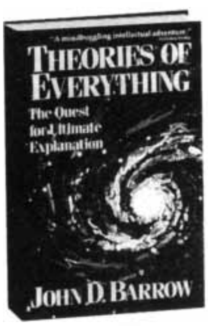
57765-2 $37.50 (counts as 2 choices)



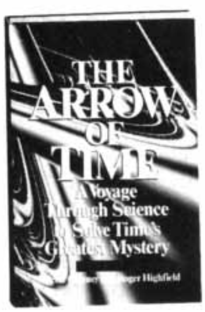
85346 $27.95



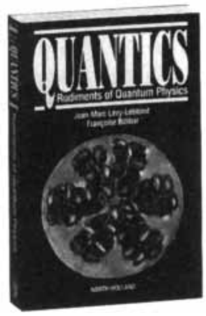
77506 $22.95



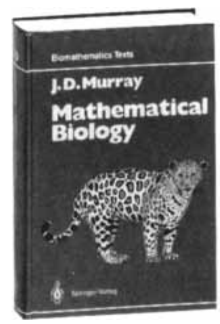
83254 $22.95



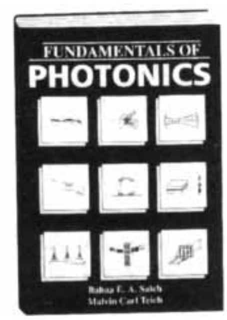
34192 $22.50



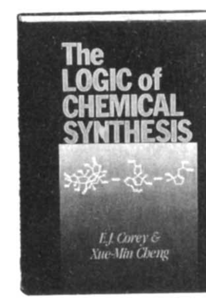
73183-2 $44.95 (counts as 2 choices)



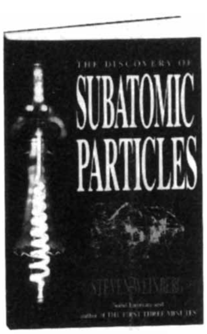
60734-2 $59.00 (counts as 2 choices)



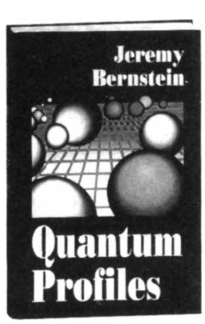
49332-2 $64.95 (counts as 2 choices)



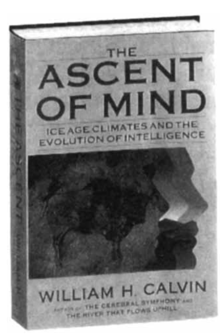
58760 $29.95



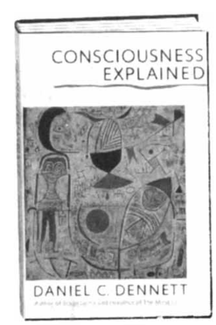
42315 $13.95



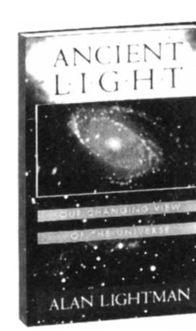
73238 $17.95



34531 $21.95



40170 $24.95



33460 $18.95

4 Good Reasons to Join

1. The Finest Books. Of the hundreds of books submitted to us each year, only the very finest are selected and offered. Moreover, our books are always of equal quality to publishers' editions, never economy editions.

2. Big Savings. In addition to getting 3 books for only $1.00 each when you join, you'll receive discounts on books you choose to purchase. Discounts generally range from 15% to 30% off the publishers' prices, and occasionally even more. (For example, your total savings as a trial member—including this introductory offer—can easily be over 50%. That's like getting every other book free!)

3. Bonus Books. Also, you will immediately become eligible to participate in our Bonus Book Plan, with savings of more than 50% off the publishers' prices.

©Newbridge Communications, Inc.

4. Convenient Service. At 3-4 week intervals (15 times per year), you will receive the Library of Science News, describing the Main Selection and Alternate Selections, together with a dated reply card. In addition, up to three 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-DK8, 3000 Cindel Drive, Delran, NJ 08370-0001, for membership information and an application.
Scientific American 3/1/92

Unlikely Places

Planets show up where nobody expected to see them

Ever look everywhere for something only to find it in the least likely place imaginable? Astronomers seeking signs of planets around other stars know how you feel. For decades, they have scrutinized mild, sunlike stars for signs of planetary systems like our own. Their search has come to naught. Instead the first strong evidence of distant planets has come from observations of pulsars, those tiny, ultradense, rapidly spinning stellar corpses that remain after supernova explosions.

In January, Alexander Wolszczan of Arecibo Observatory in Puerto Rico and Dale A. Frail of the National Radio Astronomy Observatory reported that two or perhaps three planets are orbiting around a pulsar known as PSR 1257+12 (the unglamorous name refers to its position in the sky). The pulsar, located about 1,500 light-years away in the constellation Virgo, was discovered by Wolszczan in February 1990.

Soon afterward Wolszczan noticed some odd behavior from PSR 1257+12. As pulsars rotate, their powerful magnetic field causes sharp bursts of radiation to sweep by the earth. Radio signals from pulsars usually remain highly stable. But sometimes this pulsar's radio blips arrived slightly ahead of schedule, other times a tiny bit late. Moreover, the timings did not vary randomly. Rather the signals showed two superimposed cycles having periods of 98.2 and 66.6 Earth days.

Wolszczan inferred that the pulsar was being tugged back and forth by two unseen orbiting objects. When the pulsar nudged closer to the earth, its radiation took a little less time to travel here, so that the signals arrived early, and vice versa. Based on the tiny back-and-forth motion of the pulsar (only about 900 kilometers), Wolszczan and Frail concluded that the invisible bodies are not stars but planets, each about three times the mass of the earth.

The history of searches for planets around other stars is littered with spurious discoveries—even some involving pulsars with planets. Last July a team of radio astronomers at the University of Manchester announced the discovery of a planet around another pulsar, PSR 1829–10. But on January 15 Andrew G. Lyne, one of the members of the team, informed the American Astronomical Society that their data had been thrown off by a slight error in the measured position of the pulsar.

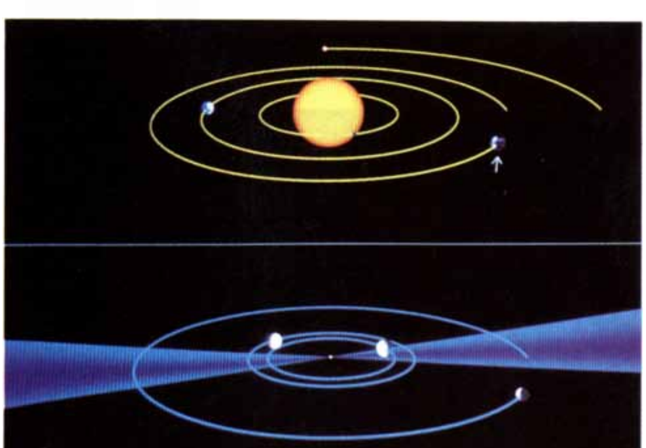
Nevertheless, astronomers are taking Wolszczan and Frail's planets seriously. Frederic A. Rasio, who has collaborated with several co-workers at Cornell University to devise a test to prove the reality of the planets, comments that "none of us questions the existence of the planets." Rasio points out that it will be fairly easy to determine whether the pulsar planets do indeed exist. Gravitational interactions between the planets should cause their orbits to shift, producing a clear and unique variation in the arrival times of signals from the pulsar. Such changes should show up within the next one to three years. If they do, "the reality of the planets will be irrefutable," Rasio says.

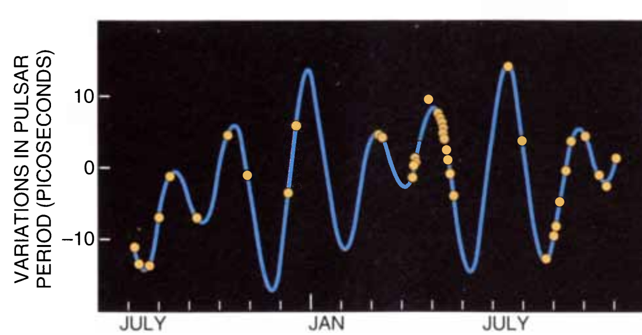
There remains the question of how planets could come to circle a pulsar. Many theorists actually breathed a sigh of relief when Lyne retracted the claim of a planet around PSR 1829–10. The supernova explosion that transformed the central star into a neutron star would certainly have destroyed or jettisoned any previously existing planets. And conjuring up a plausible explanation of how a planet could have come together around PSR 1829–10 proved a tricky task. The pulsar is young (which means there has been little time for planets to coalesce) and rotates slowly (which implies that it has not interacted closely with another star since the supernova explosion).

Wolszczan's pulsar, on the other hand, is one of a class of extremely rapidly spinning objects known as millisecond pulsars; it rotates every 6.2 milliseconds, or 162 times per second. These are thought to be old pulsars that have pulled material from a nearby stellar companion, causing the rotation to increase. Leftover, unconsumed gas around these pulsars could coalesce into planets.

PSR 1257+12, like many millisecond pulsars, now travels without a stellar companion. Many theorists think energetic electromagnetic radiation and particles emitted from a pulsar could heat its larger, less evolved companion and, in effect, boil it away. In fact, two pulsars recently have been observed in the process of vaporizing their stellar neighbors. Marco Tavani of Lawrence Livermore National Laboratory has run simulations on a supercomputer showing that some of the liberated gas could collect in a stable disk around the pulsar, providing raw material for possible planetary formation. Alternatively, the planets might have coalesced out of the debris from the final distintegration of the companion star.

Clearly, some pulsars do interact with nearby orbiting stars. Nobody knows for sure whether these interactions do in fact create planets. But if the planets around PSR 1257+12 prove real, it would suggest that pulsar planets may be a fairly common phenomenon. "It's an exceedingly exciting discovery," says Stanford E. Woosley of the University of California at Santa Cruz. "Perhaps planets are easier to form than we previously thought." —*Corey S. Powell*





***PLANETS AROUND THE PULSAR* (bottom left) *are distributed on about the same scale as planets in the inner solar system* (top left). *The existence of pulsar planets is inferred from variations in the pulsar's radio emissions* (above).**


BOMBAY SAPPHIRE. POUR SOMETHING PRICELESS.
PRODUCT OF ENGLAND
BOMBAY
SAPPHIRE
Distilled
LONDON
DRY GIN
Distilled from 100% Grain Neutral Spirit from a 1761 recipe.
1 Litre 94° Proof
Bombay® Sapphire™ Gin. 47% alc/vol (94 Proof). 100% grain neutral spirits. © 1989 Carillon Importers, Ltd., Teaneck, N.J.

To Those Who Have Ta
To Advance, We Offer The

Most likely your days are filled with ringing telephones, mounting pressures and impending deadlines. All of which make your commute that much more crucial to your peace of mind.

Accordingly, the cabin of the LS 400 was designed to serve as your sanctuary.

Step inside. The door seals shut with a solid hush. The steering column automatically lowers itself into your hands.

You reach for the ignition. Under the hood before you lies a 250-hp, V8 engine, powerful yet efficient. With the twist of your wrist it springs to life. At least you think it

Pointing Fingers

DNA identification is called into question

DNA fingerprinting has been admitted as evidence in more than 2,000 court cases in the U.S. since 1988. The technique, which is employed extensively by the Federal Bureau of Investigation as well as by private companies, has been decisive in kidnap, rape and murder trials. But some population geneticists are questioning the accuracy of the tests. The controversy has led to charges that the FBI is attempting muzzle the critics.

The arguments erupted with the publication in *Science* of a paper by Richard C. Lewontin of Harvard University and Daniel L. Hartl of Washington University that spells out how differences within the population undermine the assumptions made in the calculations. Their misgivings are likely to be confirmed by a forthcoming report on DNA technology in forensic science by a committee of the National Academy of Sciences.

The technique involves analyzing DNA samples for the number of copies of certain crucial fragments that vary from individual to individual. If the numbers of copies of several such sequences in a suspect's DNA—the "fingerprint"—match those in a sample from the scene of a crime, a simple calculation is done to estimate the odds of the match being due to chance alone. Often the odds seem vanishingly remote—in one case cited by Lewontin and Hartl, one in 738 trillion.

In their paper, Lewontin and Hartl contend that such calculations are based on false assumptions that could make them wrong by two or more orders of magnitude. They argue that existing techniques for calculating the odds of a DNA match fail to account adequately for the distorting effects caused by the ethnic and racial variation. An innocent suspect racially or ethnically similar to that of a criminal could have an inflated chance of matching a forensic sample—and thus be wrongly convicted. The paper suggested techniques that would minimize the chances of declaring such a false match.

Even before their paper appeared in *Science*, Lewontin and Hartl had testified as expert witnesses in several cases, and their critiques of DNA fingerprinting calculations had been widely circulated among "the defense community," says John W. Hicks, who heads the FBI's crime laboratory. These reports have led to the rejection of the FBI's DNA fingerprinting evidence in a number of recent court cases.

Among them was a consolidated hearing on DNA evidence for 15 separate cases in the District of Columbia. As in most jurisdictions, scientific evidence in the District has to be examined in a "Frye hearing" and found to be generally accepted among experts before being admitted. The judge in the D.C. case ruled that the underlying genetic ideas were not generally accepted. (The U.S. Court of Appeals for the Second Circuit in Manhattan, in contrast, has accepted DNA evidence, but it employs a less stringent standard.)

Not surprisingly, Lewontin and Hartl's paper has been bitterly criticized by proponents of DNA fingerprinting. Attempts were also made to persuade Lewontin and Hartl not to publish their paper. An assistant U.S. attorney in Cleveland, Ohio, James Wooley, called Hartl and told him publication would

Going through Proper Channels

By this time next year, says neuroanatomist Rodney Douglas of the University of Oxford, it is likely that one of his graduate students will be preparing to study nerve impulses in a silicon fly. Others may be delving into the neural intricacies of electronic ants, silicon slices of mammalian visual cortex and other models of biological information processing.

The reason for Douglas's enthusiasm is a neuron on a chip he has developed with Misha Mahowald of the California Institute of Technology. Artificial neurons are an essential component of so-called neural nets in which researchers try to duplicate the structure and function of nerve cells in silicon. Although neural nets are being designed to create computers that may mimic human thought, they are also becoming important models of biological processes. Douglas says the device brings a new level of verisimilitude to simulations of nerves because it emulates the flow of calcium and potassium ions that triggers the behavior of biological neurons.

Until now, Douglas says, researchers have had to rely on digital simulations of neurons, which require large, fast computers and expensive programming. Even then, the results still felt "slightly removed" from reality, he says. The silicon version, which should be both cheaper and faster than software simulation, provides the feel of working with living cells—even some of the noise, randomness and variability that real neurons display. In addition, experimenters can vary the properties of the silicon neuron so that it matches the behavior of particular types of neural cells. For example, the parameters that correspond to membrane permeability for different ions can be adjusted to reflect the voltages and current flows of specific cells.

Previous attempts to re-create axons and synapses in silicon have run into trouble. For example, it is difficult to connect any significant number of neurons. Long wires on a chip, necessary to duplicate axons, attenuate or distort signals. Differences in the fabrication of chips can make a voltage level "mean" one thing on one chip and indicate something different on another.

The new neuron, however, produces not a steady voltage but rather a train of pulses at a rate that depends on its activation. "A spike is a spike," Mahowald says, explaining that the simplest circuitry can detect its presence or absence. The neurons are connected by a high-speed bus so that when a neuron generates a spike, it flashes its address to all the others. Any neuron that has the firing neuron as one of its inputs recognizes the address and changes its activation level accordingly. This system should allow experimenters to wire together hundreds or even thousands of neurons on dozens of chips, Douglas claims. Furthermore, the address circuits of each neuron are programmable, so that their interconnections can be changed without rewiring.

Douglas predicts that chips based on the neural microcircuit will be widely used both by artificial intelligencers trying to build simple autonomous systems and by neuroanatomists studying the function of the brain. The circuit can test theories of neural function and interconnection directly, a luxury "not usually available to neuroscientists," he says. "You think you understand it," he comments, but the interesting questions arise when "you see it work in the world."

—*Paul Wallich*

be misunderstood and could harm prosecutions.

Then the paper, which had survived peer review, was delayed for several weeks at *Science* in the galley stage when a reviewing editor, C. Thomas Caskey of Baylor College of Medicine complained about it. Caskey was also a member of the National Academy's forensic DNA committee. Daniel E. Koshland, Jr., the editor of *Science,* published a rebuttal by Ranajit Chakraborty of the University of Texas at Houston and Kenneth K. Kidd of Yale University in the same issue, dated December 20, 1991.

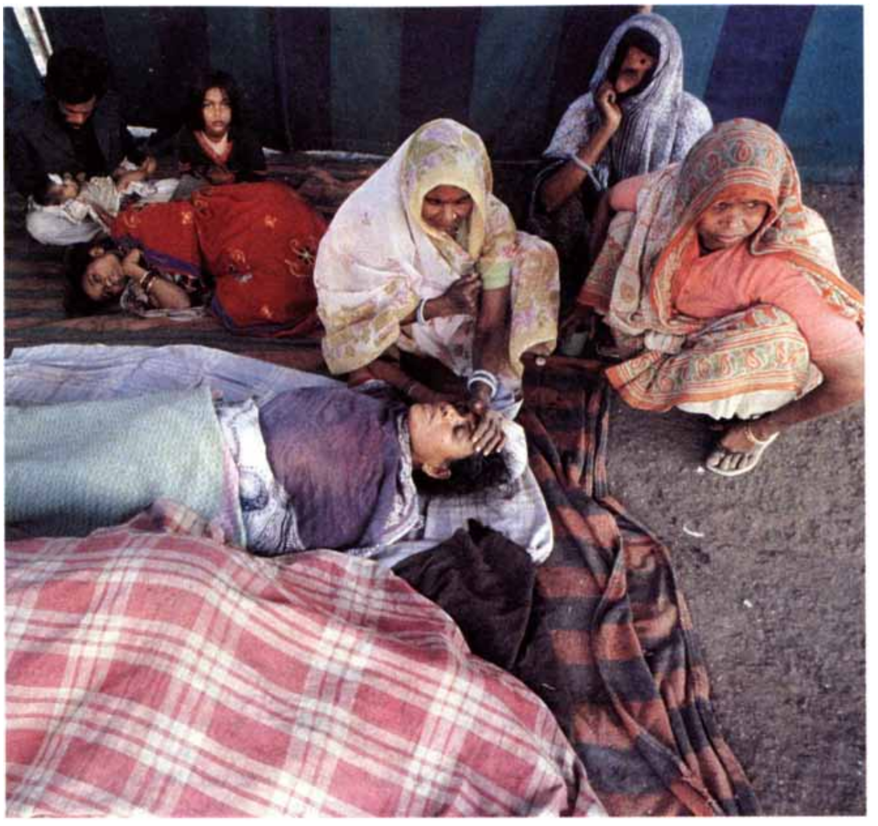
The rebuttal concluded that current techniques for calculating the odds of a match in a mixed population, which draw on data banks on gene frequencies in different races, will not lead to significant error. "It is the general frequency in the total population that is desired," they wrote. Chakraborty says, "If I were on a jury it wouldn't matter to me if the odds were eight million or 40 million."

Defense lawyers find that line of reasoning unacceptable. And Hartl, who with Lewontin plans to publish a response to Chakraborty and Kidd, insists they are arguing the wrong point. "We are not talking about averages," Hartl says. "We are talking about the chance there is someone else in the world who matches."

The day after the publication of Lewontin and Hartl's paper, Caskey resigned from the National Academy's forensic DNA committee. A committee member says the committee became concerned about possible conflict of interest. Caskey has substantial financial interests in DNA fingerprinting technologies he has developed, and his laboratory has a $200,000 grant from the National Institute of Justice. The institute is, like the FBI, an arm of the Department of Justice.

The FBI was also one of the sponsors of the academy report, but academy staff were still surprised when Hicks sent a lengthy unsolicited critique of a leaked chapter of the report to an academy official, John E. Burris. "We were very concerned there wasn't a balanced perspective," Hicks declares.

Hicks says he was given a draft chapter of the report by two members of the committee. The academy, which asks committee members to keep draft reports confidential, declined to pass on Hick's letter to the committee in the interests of preserving impartiality. The academy's report, which has been two years in the making, may yet add a much needed objective viewpoint to the debate. —*Tim Beardsley*



***SURVIVORS IN BHOPAL** receive medical aid after the 1984 chemical spill. Could the accident have had health effects beyond burns? Photo: Baldev/Sygma.*

Trojan Horse

Did a protective peptide exacerbate Bhopal injuries?

Soon after midnight on December 3, 1984, more than 30 tons of methyl isocyanate, a gaseous toxin, escaped from a Union Carbide pesticide factory in Bhopal, India. The gas reacted furiously with moisture in the exposed tissues of the sleeping residents, burning them. Many victims suffocated as body fluids poured into their scorched lungs. Respiratory damage and infection eventually claimed others. According to current estimates, more than 3,500 people have died, and tens of thousands of others are believed to have been acutely injured by the methyl isocyanate.

Seven years later medical authorities are still plumbing how deeply the toxin may have scarred the survivors. Some physicians have found that, aside from the obvious burns on lungs and eyes, many patients also suffer from unexplained cardiovascular, gastrointestinal, neuromuscular, reproductive and immunologic ailments.

Organs throughout the survivors' bodies seem to have been damaged—tissues that, in theory, the chemical should not have touched. Some chemists now suspect the toxin was helped by glutathione, a peptide usually believed to protect cells from chemical attack. Recent work is tarnishing glutathione's reputation: at times, it may act as a Trojan horse, carrying poisons into the body.

At first, health authorities had generally assumed that methyl isocyanate was too reactive to have diffused far into the tissues of the Bhopal victims. In 1988, however, two groups, led by Yves Alarie of the University of Pittsburgh and by B. K. Bhattacharya of the Defence Research and Development Establishment in India, demonstrated independently that the chemical could penetrate tissues. When laboratory animals inhaled radioactively labeled methyl isocyanate, the molecules quickly entered their blood, urine and bile and continued to circulate through the body for several days.

How could the poison move through tissues without reacting? Thomas A. Baillie of the University of Washington and J. Greg Slatter of the Upjohn Company in Kalamazoo, Mich., think methyl isocyanate may hitch a ride with glutathione, which is found at high concentrations in the lining of the lungs.

Normally glutathione helps to purge cells of dangerously reactive radicals and other compounds by binding with them. The resulting conjugate molecule is then swept away and excreted.

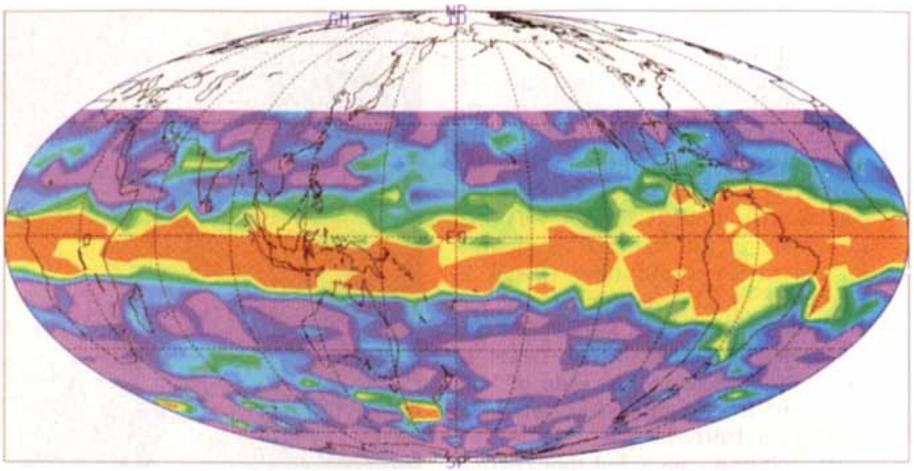
Glutathione binds with methyl isocyanate, too, Baillie says, but it does so reversibly. At Bhopal, glutathione might have ferried a quantity of methyl isocyanate out of the lungs and occasionally released it deep within the body, where it could have attacked the surrounding tissues, leading to the multifarious illnesses seen in survivors.

Experiments conducted by Baillie and Slatter support the idea. They have shown that animals injected with methyl isocyanate do excrete its glutathione conjugate. In the test tube, they have also demonstrated that the conjugate can break down under certain physiological conditions and that methyl isocyanate delivered in this way can damage proteins and membranes located on cells. "There's no question that inhaled substances like methyl isocyanate would come in contact with glutathione," Baillie says. "What we don't actually know at this point is, when the conjugate is generated, whether it can cross over from the lungs into the bloodstream."

"Baillie has certainly come as close as anybody to showing that this can happen physiologically," says William E. Brown of Carnegie Mellon University, who has studied methyl isocyanate for two decades. He notes, however, that biochemists still debate whether the methyl isocyanate conjugate does decompose inside the body and what the consequences might be.

Although Baillie's work represents a good working hypothesis, Brown cautions, alternative mechanisms might be at work. Metabolic pathways for detoxifying pollutants in the lungs, for example, might transform methyl isocyanate into longer-lived radicals that could enter the blood. Such reactions are known to convert some substances into carcinogens, he says.

"People are starting to reconsider the accepted mechanisms of toxicity and detoxification," Baillie believes. As he and Slatter pointed out in a recent review in *Accounts of Chemical Research,* glutathione can reversibly conjugate with several classes of toxins found in nature and industry, including isothiocyanates—compounds found in many vegetables. Further research should clarify how adversely glutathione's Jekyll and Hyde personality affects human health in Bhopal and elsewhere; it may also determine whether anything can be done to remedy the molecule's uglier aspect. *—John Rennie*



Volcanic Disruption

A giant eruption frays the tattered ozone layer

It has become an annual tradition, like Ground Hog Day. Every October the National Aeronautics and Space Administration releases a satellite image of the ozone hole lurking over Antarctica. The media remind us that ozone shields the earth from ultraviolet radiation, which can wither crops and cause cancer, cataracts and weakened immune systems in animals, including humans. Then shifting wind patterns dissipate the hole, and ozone is forgotten, for the most part, until the hole reappears during the next Antarctic spring.

But this year the story did not dissipate. On October 6, a NASA satellite recorded the lowest levels of ozone ever detected above Antarctica, suggesting that the annual hole is growing larger. Days later the United Nations Environment Program announced that the ozone is steadily thinning not just over the poles but at populated mid-latitudes during summer months, when crops grow, people are outdoors and sunlight peaks in intensity.

Furthermore, evidence has begun accumulating that Mount Pinatubo, the gigantic Philippine volcano that erupted last June, is accelerating the erosion of stratospheric ozone even in tropical regions, where intense sunlight normally suppresses ozone-destroying reactions. Based on the preliminary data, Susan Solomon of the National Oceanic and Atmospheric Administration (NOAA) suggests that Pinatubo could trigger ozone losses of more than 30 percent this spring in mid-latitudes.

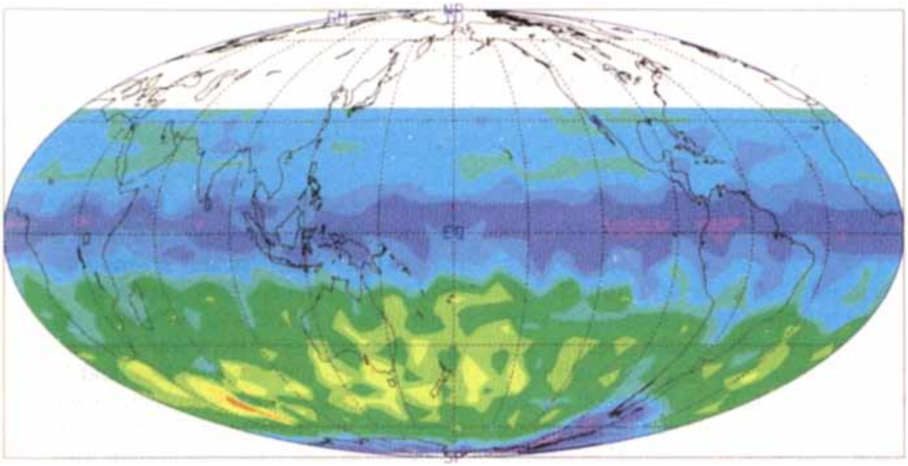
If Solomon is right, regions such as the U.S. and Europe could receive double their usual springtime dose of DNA-damaging ultraviolet radiation, according to Sasha Madronich of the National Center for Atmospheric Research. He says such increases could have a statistically significant impact on the health of plants and animals. Owen B. Toon of the NASA Ames Research Center notes that as the volcanic cloud disperses over the next two or three years, its impact will fade. Nevertheless, he says, "this is not good news."

The idea that volcanic eruptions could intensify ozone depletion was first proposed in 1989 by Solomon and another NOAA scientist, David J. Hofmann. Most ozone loss is thought to stem from chlorine-catalyzed reactions occurring on the surface of ice crystals found in polar stratospheric clouds. (The major sources of chlorine in the stratosphere are artificial compounds called chlorofluorocarbons, or CFCs.) The NOAA researchers argued that sulfur-based aerosols spewed forth by volcanoes would create, in effect, temporary polar stratospheric clouds.

To support their hypothesis, Solomon and Hofmann pointed out that ozone levels in the Northern Hemisphere dipped by some 10 percent during the year following the 1982 eruption of El Chichón in Mexico. After Pinatubo erupted, Solomon and Hofmann warned that it could cause much more ozone depletion than El Chichón, particularly in middle and upper latitudes.

According to the data available so far, their warnings were, if anything, understated. Pinatubo, the largest eruption of the century, coughed up 20 million tons of sulfur dioxide, three times as much as El Chichón. It created a cloud estimated to be as large as all the stratospheric clouds at the North and South Poles combined.

Perhaps the most dramatic results come from the Microwave Limb Sounder on NASA's *Upper Atmosphere Research Satellite,* which was launched just last September. The MLS has observed ozone levels in the tropics 10 percent lower than any previously measured,



GLOBE-GIRDLING CLOUD emitted by Mount Pinatubo was tracked by the Microwave Limb Sounder on the **Upper Atmosphere Research Satellite** *last September. The cloud was rich in sulfur dioxide* **(orange band, left)** *and low in ozone* **(violet band, right).** *Images from Joe W. Waters of the Jet Propulsion Laboratory.*

according to Joe W. Waters of the Jet Propulsion Laboratory, the microwave instrument's principal investigator. Solomon and Hofmann had expected tropical losses to be negligible.

Recently, Waters says, the MLS has revealed "tongues" of ozone-poor air reaching north and south from the tropics into higher latitudes, a reversal of the usual course of ozone depletion. In January the instrument also detected "extremely high" values of ozone-destroying chlorine monoxide—high enough to reduce ozone levels by 1 percent *every day*—in a region extending from the North Pole well into northern Europe. Waters emphasizes that his data must be considered preliminary, especially since the instrument was deployed so recently. "But my feeling is," he added, "that this is very significant."

Balloon measurements support that assessment. Last summer balloons lofted above Hawaii by NOAA detected low ozone within patches of cloud from Pinatubo, according to Samuel J. Oltmans of NOAA In December, he says, another NOAA balloon recorded the lowest ozone levels ever observed that time of year above Boulder, Colo. Balloons released at the extreme northern border of Canada have detected dips in ozone in conjunction with sulfurous aerosols that were "almost certainly" from Pinatubo, says Hans Fast of Canada's Atmospheric Environment Service.

Privately, NASA sources reveal that Pinatubo's effects have been detected by the satellite-borne Total Ozone Mapping Spectrometer (TOMS) and by an ER-2 aircraft. Publicly, NASA officials insist the findings are too preliminary for release. Michael J. Kurylo, manager of NASA's upper atmospheric research program, notes that the TOMS data may have been distorted by volcanic debris. He contends the link between Pinatubo and ozone depletion remains hypothetical. "It is dangerous to ascribe cause and effect," he says.

Solomon counters that "it begins to stretch credibility" to assert that all the correlations observed so far could be coincidental. She calls the apparent confirmation of the predictions she and Hofmann made in their 1989 paper "extremely exciting." She adds, however, that "if ever there was a paper that I wish was wrong, it's this one."

Ironically, the cloud from Pinatubo is also expected to cool the globe for the next few years by preventing infrared radiation and visible sunlight from reaching the earth. But scientists say the volcanic cloud does not block out nearly enough ultraviolet radiation to offset its ozone-depleting effects.

Reducing stratospheric chlorine levels would blunt the ozone-depleting effects of volcanoes, but chlorine levels are headed the other way. "If all CFCs and other chlorine sources were cut off today," remarks Jay R. Herman of the NASA Goddard Space Flight Center, "chlorine levels would continue to rise in the stratosphere for another 20 or 30 years." The next large volcano, in other words, may destroy even more ozone than Pinatubo has. *—John Horgan*

Add Ozone to the Global Warming Equation

Ozone is a paradoxical molecule. In the stratosphere, it plays a crucial role as a sunscreen filtering out ultraviolet radiation. Its relentless decline there has contributed to global anxiety about increases in skin cancer and genetic damage to organisms at the base of the food chain. At ground level, ozone is a dangerous pollutant, causing respiratory diseases and damaging forests.

Then there's the rest of the troposphere, above the ground but below about 10 kilometers. Observations made from balloons suggest that tropospheric ozone levels have been increasing rapidly—by about 10 percent per decade, according to an assessment for the United Nations Environment Program.

The news should not make anyone breathe more easily. Because the absolute amount of ozone in the troposphere is far less than that in the stratosphere, the increase won't help mop up ultraviolet radiation leaking through from above. Rather ozone is a "greenhouse gas," and the buildup will add to the warming effect of increasing carbon dioxide.

Emissions of nitrogen oxides from aircraft engines are believed to be one reason for the increase. While nitrogen oxides cause reactions that deplete ozone in the stratosphere, they participate in others that form it in the troposphere. In *Nature,* Colin Johnson and his colleagues at the Harwell Laboratory in Didcot, England, estimated that because of the ozone they produce, emissions of nitrogen oxides at altitudes frequented by passenger aircraft have 30 times more effect on surface warming than would the same emissions at ground level. Johnson believes aircraft emissions might be as important in global warming as all other emissions of nitrogen oxides combined.

The results could complicate attempts by the National Aeronautics and Space Administration to assess the effects of a future fleet of supersonic passenger aircraft. NASA's latest studies seem to alleviate concerns that a supersonic fleet would decrease stratospheric ozone. But they make it more likely that a supersonic fleet would contribute to tropospheric ozone and so add to global warming.

Researchers caution that ozone is far less important as a greenhouse gas than carbon dioxide. Nevertheless, "this is definitely an issue we're watching," says Joel M. Levy of the Environmental Protection Agency. *—Tim Beardsley*

The Lonely Odysseus of Particle Physics

It should come as no surprise that Murray Gell-Mann, a master theoretician, has an explanation for the astonishing range of his own interests. Whereas most people are either "Apollonian," that is, detached and analytical, or "Dionysian," meaning involved and intuitive, he combines both traits. "Some people call that 'Odyssean,'" Gell-Mann says, distinctly enunciating each syllable of the word, perhaps to help me spell it. "It's a lonely thing to be."

Gell-Mann, a 62-year-old professor at the California Institute of Technology, is best known as a mere particle physicist. A Nobel laureate, he is one of the primary architects of the Standard Model, the elaborate theoretical edifice that accounts for the behavior of subatomic particles and the forces ruling them. Among other contributions, he proposed that neutrons and protons, the major constituents of ordinary matter, are made of more fundamental particles called quarks.

But physics is too small a realm to contain Gell-Mann's restless intellect. His 17-page "personal statement" lists such interests as "natural history (especially bird study), historical linguistics, archaeology, history, depth psychology and creative thinking." Then there are "policy matters related to world environmental quality (including conservation of biological diversity), restraint in population growth, sustainable economic development and stability of the world political system (including the strategic arms situation)."

Gell-Mann admits he is overextended. In fact, he has calculated just how overextended he is. "I take on about 50 times more than anyone can do, and I work at about 2 percent efficiency. So every day I'm fighting a factor of 2,500, and every day I fall eight years more behind. Ninety-eight percent of my time is wasted, just wasted." He sighs.

Naturally, there is little time for interviews. One week Gell-Mann is serving as host and keynote speaker at a celebration of Caltech's 100th birthday. The next week he is flying off to Europe to attend a physics symposium. Spare moments are devoted to writing a book, entitled *The Quark and the Jaguar,* about a major current interest: the interaction in nature between simple phenomena, such as quarks, and complex ones, such as jaguars. But Gell-Mann does have an unscheduled morning at the end of a trip to New York. I can join him for breakfast and then see him off at La Guardia Airport.

AMATEUR ANTHROPOLOGIST Murray Gell-Mann, in his office at Caltech, explains how to play a West African game known as "Wari." Photo: James Aronovsky.

Chewing a lox-draped bagel in a packed Manhattan restaurant, Gell-Mann, a compact man with a white crew cut and black glasses, seems at home. And so he should. Like several other prominent physicists of his generation, he was born and raised—by Jewish immigrants from Europe—in New York City. His manner of speaking, though a paragon of erudition and cosmopolitanism, still retains a New York edge.

Perhaps it is the way he serves up opinions. Science writers, and journalists in general, are "ignoramuses" and "a terrible breed." British scientists have a tendency "to be more concerned with being clever and paradoxical than with being right." Gell-Mann also offers less than glowing assessments of some prominent scientists. "But I don't want to be quoted insulting people," he says. "It's not nice. Some of these people are my friends."

We skim over Gell-Mann's youth. He entered Yale University in 1944, just after turning 15, and graduated four years later along with another high achiever, George Bush. Although they occupied different circles at Yale, Gell-Mann and Bush became acquainted decades later while serving as regents for the Smithsonian Institution. Bush is "a nice fellow" who gives "very good parties," Gell-Mann says. "I just wish someone would find him a better job than running the country."

Gell-Mann was heartbroken when he failed to win a financially acceptable offer from an Ivy League graduate physics program. He feared that the Massachusetts Institute of Technology, which offered him a job as assistant to the physicist Victor F. Weisskopf, would be a "grubby place," a breeding ground for engineers rather than true scientists. "But I went there, and it was fine," Gell-Mann says. Weisskopf turned out to be "a wonderful man and an excellent physicist." Moreover, many of his colleagues were Ivy League graduates, "so it wasn't like I was separated from these people."

After obtaining his doctorate from M.I.T. early in 1951, Gell-Mann strove to find order beneath the bewildering variety of hadrons springing up in particle accelerators. Hadrons are particles subject to the strong nuclear force, which clamps neutrons and protons together in the nuclei of atoms. Gell-Mann eventually invented a quantum property, which he termed strangeness, that accurately predicted the behavior of the exotic new hadrons.

Building on this insight, Gell-Mann erected a particle classification system that he called the eightfold way, after the Buddhist path to enlightenment. The term encouraged the notion, popular in the 1960s, that particle physics and Eastern mysticism are profoundly linked. Gell-Mann calls the idea "rubbish." He now says his allusion to Buddhism (another interest) was just "a joke that some people took seriously."

Gell-Mann points out that he could easily follow tradition and concoct "proper, pompous names for things, based on Greek and so on. I know how to do that. But usually they were based on ideas that turned out to be wrong. 'Proton,' for example, meaning first. 'Atom,' uncuttable. All these things turned out to be wrong! So I figured it would be better to come up with something playful."

Gell-Mann won the Nobel Prize in 1969 for his work on strangeness and the eightfold way. Meanwhile, five years earlier, he had invented a much more comprehensive theory (also proposed independently by a former student, George Zweig). It holds that neutrons, protons and all other hadrons consist of more elementary particles called quarks. Perhaps the most radical property of quarks is that they carry fractional charges. A proton, which has a charge of +1, is made from two +2/3 quarks and one –1/3 quark. A chargeless neutron consists of two –1/3 quarks and one +2/3 quark. Quarks are bound together by force-carrying particles dubbed (not by Gell-Mann) gluons.

Gell-Mann is eager to dispel two common misconceptions about the history of quarks. First, he did not simply swipe the neologism from *Finnegans Wake*. Rather he initially decided that a sound like "kwork" (pronounced to rhyme with "fork") might make a nice tag for the new particle. Only then, while perusing James Joyce's gobbledygookian tome—as Gell-Mann has done often since its publication in 1939—did he come on the phrase "three quarks for muster mark" and decide to adopt Joyce's spelling. (Gell-Mann notes that "quark," as used by Joyce, evokes both a quart of ale and the cry of a gull.)

Then there is the matter of Gell-Mann's having suggested in an early paper that quarks are "mathematical" rather than "real" entities. He made this distinction, he says, because he believed quarks would prove to be permanently "trapped" inside hadrons by the gluons and thus would be impossible to isolate and detect individually. "I didn't want to get into an argument with philosophers, who are an unbelievable pain in the neck, many of them, saying, 'What? Real? But it can't come out! What do you mean saying it's real!'"

Experiments in the 1970s provided indirect yet still compelling evidence for Gell-Mann's basic trapped-quark model (and for a more detailed quark-based theory developed by Gell-Mann and others, called quantum chromodynamics). "But virtually everybody repeats in book after book by science writers and people like that that I never really believed in quarks because I said they were most probably 'mathematical,'" Gell-Mann says. "I've suffered terrible insults as a result of being right!"

Gell-Mann regrets having once proposed that quarks are "mathematical" rather than "real" entities.

For the last decade or so, Gell-Mann has been less involved in day-to-day particle physics. He is a cheerleader for—rather than active developer of—superstrings, extremely small and hypothetical particles whose wrigglings supposedly give rise to all the observable particles and forces of nature, including gravity. Gell-Mann says he feels paternal toward superstrings, since his work on something called dispersion theory in the 1950s helped to inspire the first string model. He calls critics of superstrings "crazies": "They should try to work out the implications of the theory instead of attacking it."

Gell-Mann has been more directly involved in devising a version of quantum mechanics that can be applied to cosmology. The work stems in part from a proposal by Stephen W. Hawking of the University of Cambridge and James B. Hartle of the University of California at Santa Barbara that the entire universe can be described by a quantum equation, or wave function. There is a problem, however: a wave function describes a range of possible paths for a particle, and the particle "chooses" one path only when someone observes it. If the particle is the entire universe, who and where is the observer?

In the late 1950s, Gell-Mann notes, Hugh Everett III of Princeton University invented an observer-free version of quantum mechanics called the many-worlds interpretation. It held that a particle follows all the paths described by its wave function—in different universes. With Hartle, a former student, Gell-Mann is developing a modified version of Everett's concept, called the many-histories interpretation.

In choosing the word "histories," Gell-Mann says, he and Hartle sought to distance themselves from Everett's "flawed" belief in the actuality of alternate universes. Gell-Mann recalls a colleague's remark that "if you believe in Everett's interpretation, you should play Russian roulette for high stakes, because in one of the worlds you will win a lot of money."

Gell-Mann has also become an enthusiastic student of complexity, which he says embodies principles underlying virtually all his other pursuits and interests. In the early 1980s he helped to found the Santa Fe Institute in New Mexico, where researchers study such "adaptive complex systems" as languages, brains and national economies. He has also sought to make his mark on the field by proposing a new term, "plectics," which refers to the study of simplicity and complexity in all their myriad manifestations. (The term has not caught on yet, perhaps because it has a Greek root.)

We have arrived at the airport and are running out of time. We have scarcely touched on Gell-Mann's political persona, which extends back to his childhood. In 1944 he entered Yale late so he could work as a volunteer in the reelection campaign of President Franklin D. Roosevelt. In the 1960s, as a science adviser to the Pentagon, he helped to formulate the policy that led the U.S. and U.S.S.R. to ban all but very limited defenses against each other's nuclear missiles. Gell-Mann says he was shocked in 1983, when "that third-rate actor Ray-gun" proposed a "Star Wars" system that would abrogate the ban.

After he became a director of the MacArthur Foundation in 1979, Gell-Mann began establishing programs to promote conservation, population control and "sustainable activities" worldwide. Recently he has succeeded in diverting some MacArthur funds into research on how the subconscious affects mental health. Intrigued by the role his own subconscious played in his creative achievements, Gell-Mann had urged psychologists and psychoanalysts to study the subconscious as early as the 1960s—but was told, basically, to mind his own business.

I do manage to get in a question about historical linguistics. "Linguistics is something I'm very good on," Gell-Mann confides. In fact, he is co-editing and writing a preface and chapter for a book on that topic. Gell-Mann's preface briefly argues for a theory, proposed by Joseph H. Greenberg of Stanford University, that one can construct a family tree for all the world's languages by analyzing similarities between them. Some linguists—full-time, professional linguists—argue that Greenberg's linguistic evidence cannot support his conclusions. Gell-Mann finds this view "so silly on the face of it that you wonder how adult human beings can adopt it."

As Gell-Mann checks in at the ticket counter, he begins worrying aloud that he may not have enough cash to take a taxi from the airport in California to his home. I lend him $40, and Gell-Mann writes me out a check. As he hands it to me, he suggests that I consider not cashing it because his autograph might be "rather valuable."

Gell-Mann then rushes toward the departure gate. "I like to be the first one on the plane," he says. Trailing behind him, I finally muster enough courage to mention that some scientists, while acknowledging that he is very smart, also suggest that he is, well, sort of a know-it-all. "I don't know what that means," Gell-Mann snaps over his shoulder. A few seconds later he declares, "I do know a lot about a lot of things, it's true!" *John Horgan*

SCIENTIFIC AMERICAN

Building a Market Economy in Poland

How is a previously centralized economy to be transformed? Poland's experience thus far offers lessons for other Eastern European nations and a challenge for the West

by Jeffrey Sachs

During the past two years, the nations of Eastern Europe have abandoned an economic system based on state ownership and the centralized control of production and prices. In its place they hope to construct free-market economies like those that have brought such visible prosperity to Western Europe during the postwar period. How is this transformation to be accomplished?

The changeover was made possible by the democratic revolutions that ended communist control in Eastern Europe, but the change in economic direction reflects more than political developments. By 1989 the perverse incentives of communist rule had led to near economic collapse. State bureaucrats expanded outmoded heavy industries far beyond sensible limits while neglecting the provision of consumer goods and services. The push to expand industrial production at all costs had led to environmental degradation well beyond that which sparked the environmental movement in the West. The average Czech, Pole or Hungarian enjoys—though that is hardly the word—a living standard far below the levels of even the poorer countries of the European Community and about on a par with the some of the developing nations of Latin America.

Although citizens want the benefits of a market economy, national leaders still face political and logistic challenges in implementing reforms. Last fall's parliamentary elections in Poland, for example, show the limits of people's patience for unemployment and lower real wages during the period of economic transition. (Architects of other restructurings, such as Germany's postwar economic miracle, were also widely criticized in early years for the turmoil they allegedly caused.) Meanwhile the sheer bulk of state-owned enterprises poses a task of unprecedented dimensions in turning them over to private hands.

Western aid will be essential to this restructuring. Not only are vast resources required to rebuild worn-out industries and crumbling cities, but the West must make it clear that the newly free economies, once established, will be welcomed into the community of European and global trade. Failure to do so could threaten not only the economic health of Eastern Europe but also its nascent democracy.

I have been most closely involved with the restructuring of Poland. Its economy is widely recognized as one of the worst-off of Eastern Europe. Its heavy industrial sector is overgrown, inefficient and completely protected from international competition. Light industry, services and distribution have been neglected, as has the nation's financial system. Millions of poor peasant farmers form the core of the agricultural sector. They work tiny plots of land, typically no larger than five hectares (2.5 acres).

To remedy these conditions, the country is aiming to put in place a political and economic system similar to that of members of the European Community. Its fundamental goal, like that of the other former Soviet bloc countries, is the closest possible integration with Western Europe. By the turn of the century this course should lead to full membership in the European Community. There is to be no "third way"—the mix of continued socialist ownership with market forces rather than central planning that some have called for as a compromise between the old system and modern capitalism.

Former finance minister Leszek Balcerowicz put it bluntly: "Poland is too poor to experiment. We will therefore follow working models. Let the rich countries experiment if they want."

The urge to harmonize with and eventually join the European Community has deep roots. Poland desires to regain its place in the mainstream of European society and culture. The slo-

JEFFREY SACHS, who serves as an economic adviser to several governments in Latin America and Eastern Europe, is one of the architects of Poland's economic reform program. Between 1986 and 1990 Sachs helped to design and implement a stabilization program for Bolivia that reduced annual inflation from 40,000 percent to 15 percent. He is currently leading a team of economic advisers for Russian president Boris Yeltsin. Educated at Harvard University, Sachs now holds the Galen L. Stone Chair in International Trade there.

gan of the revolutions of 1989, after all, was "return to Europe." In addition, Poles are virtually unanimous in their admiration for the social and political accomplishments of postwar Western Europe in creating wealthy and largely equitable societies based on private ownership, social welfare systems and stable parliamentary democracies.

Not only do the Poles regard European Community nations as a model to emulate, but they also believe that free trade and financial relations with the European Community will directly help them catch up with Western living standards. Strong evidence indicates that the close ties of the European Community have helped poorer countries such as Portugal and Spain close the gap in living standards between them and richer nations such as Germany and France. More generally, the economic evidence suggests that outward-oriented trade regimes stimulate growth, and Western Europe is the logical trading partner for its Eastern neighbors.

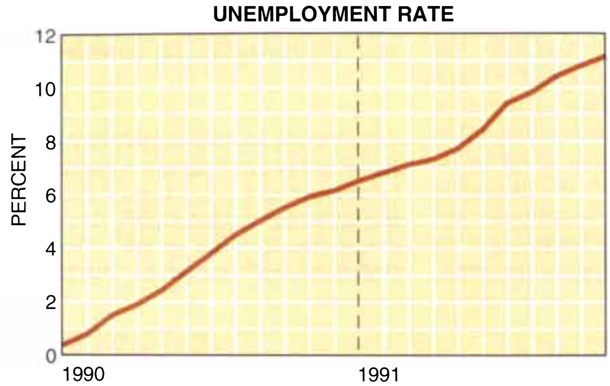
Another advantage of early commitment to eventual European Community membership is symbolic. Many potential investors naturally fear that the enormous economic turmoil of restructuring could lead to domestic political demands for protectionism or other actions inimical to long-term growth. Treaty agreements with the European Community could irrevocably set Poland on a free-trading course and so make investment there more attractive. After Portugal and Spain joined the European Community in 1986, the rest of Western Europe increased its investment in those countries significantly, at least in part because of the increased assurances on long-term policies.

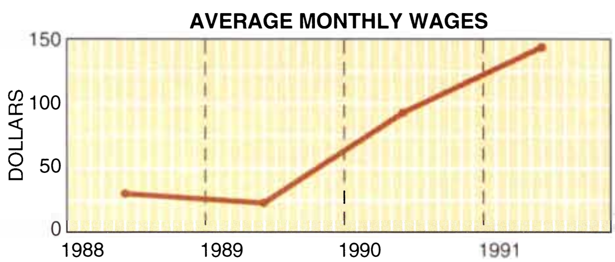
Poland's economic transformation has two critical components: the reforms that convert a command economy to a market-driven one; and the actual restructuring of industrial and

OUTDATED HEAVY INDUSTRY is a major weak spot in Poland's economy. Newly unleashed market forces are shrinking industrial output rapidly. In contrast, employment in the nation's previously neglected service sector is burgeoning.

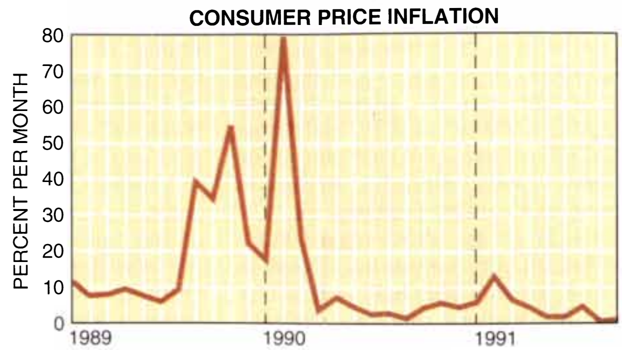
The State of the Polish Economy

The breakdown of the communist regime during 1989 produced sharp inflation and a drastic fall in the value of the zloty (*right*), both arrested by the economic reforms put in place in January 1990. Since then, unemployment (previously illegal) has increased rapidly, as have the wages of those still employed.

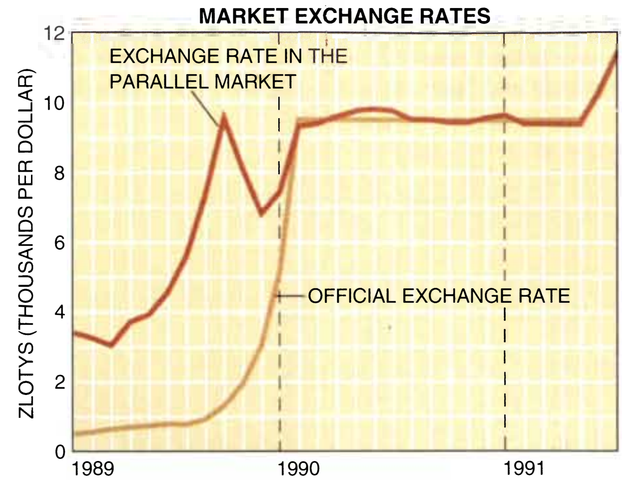




SOURCES: Polish statistical bulletin; International Financial Statistics of the IMF



SOURCE: Polish statistical bulletin (various issues)



SOURCE: Polish statistical bulletin (various issues)

business activity that will take place as a result of the reforms. The reforms can and should be introduced quickly—within three to five years at most. Restructuring, in contrast, will necessarily last a decade or more, as investors build up new sectors of the economy in response to the resulting market signals.

While restructuring takes place, the government must make certain that a safety net protects the most vulnerable members of society—among them, farmers, pensioners and those working in outmoded industries. Much of the safety net consists of measures held over from the earlier government, such as health care, housing and the like. Other parts, such as unemployment insurance, are uniquely capitalist requirements. As in other democracies, public demand for such spending can easily outpace budgetary realities. Maintaining a proper balance between social needs and budgetary possibilities is a key economic and political challenge that has no easy solution in sight.

Although the Polish government faces pressure to modify policies that cause unemployment and other social dislocations, it is pressing forward with reforms. Three major kinds of measures dominate the policy landscape: liberalization, stabilization and privatization. Economic liberalization means introducing market competition and creating a legal framework for private property and privately owned businesses. Stabilization fosters a climate in which enterprise of any kind can survive. It involves limiting budget deficits, reducing the growth of the money supply and establishing a realistic, uniform currency exchange rate to promote stable prices and foreign trade.

Privatization, certainly the trickiest area of reform, transfers existing state property, such as factories, to the private sector. Under the communist regime, 3,000 government-held industrial enterprises accounted for 90 percent of industrial output. Another 5,000 state enterprises operated in areas such as transport, communications and trade. In addition to the privatization of state enterprises, the formation of entirely new firms made possible by economic liberalization will also act to privatize the economy.

As the economy moves toward market capitalism, comprehensive, rapid reforms can vastly reduce public confusion about the "rules of the game." Gradual introduction of new measures would create ongoing uncertainty, internal inconsistency and political resistance. The "big bang" approach, in contrast, sets clear incentives for the new economic system in place immediately. As one wag put it, gradual reforms would be the rough equivalent of the British shifting from right-hand to left-hand driving by just switching over the trucks at first.

In a strategy now being emulated more or less closely by other nations of Eastern Europe, Poland set in place the three main pillars of reform as rapidly as possible. On January 1, 1990, it introduced a convertible currency and other substantial elements of liberalization and stabilization. It has also begun the more complex process of privatization.

Before the big bang at the beginning of 1990, many economists and others expressed concern about liberalizing prices while the bulk of industrial production was in the hands of a few state-owned monopolies. Some insisted that price controls should remain until competitive producers had been established. Fortunately, they were wrong: free trade and the resulting availability of reasonably priced foreign goods proved a much quicker and more effective route to price competition in Polish markets than any attempt at administrative demonopolization could have offered.

Poland also achieved macroeconomic stabilization fairly quickly by eliminating its budget deficit and tightening monetary policy. The final breakdown of the old regime led to extraordinary financial instability and a brief period of hyperinflation, but as part of the big bang the government cut subsidies that had amounted to more than 5 percent of the gross national product (GNP), increased tax collections, put on wage controls and sharply devalued the currency. The budget shifted from a deficit of 7.4 percent of GNP in 1989 to a surplus of 3.8 percent in 1990.

These measures, together with the elimination of most price controls, caused the consumer price level to double during January 1990. Soon after this one-shot jump in prices, however, the new economic environment brought an end to high inflation and, even more remarkably, a rapid end to decades of chronic shortage and lines.

Hundreds of thousands of small businesses started operation in 1990, creating in a matter of months a network of shops, service establishments and firms engaged in international trade, construction and transport. The development of the service and trade sectors was particularly dramatic in an economy that had long been starved of them. In the first few months, most start-up capital came from unofficial sources rather than from the official banking sector. Households invested hard currency (as opposed to zlotys) that they had accumulated in earlier times. Entrepreneurs' neighbors and families contributed money, as did relatives living in the West. In addition, capital was spun off either formally or informally from state enterprises: for example, municipalities leased tens of thousands of shops to their workers. Many operations started informally (for instance, by selling meat from the back of a truck) and quickly expanded to full-fledged small businesses. By the end of 1990 and throughout 1991, the formal banking sector began to provide an increasing flow of credit to new small private businesses.

The economy also adjusted quickly to the opening of markets to international trade. Thanks to the new convertibility of the zloty and the favorable exchange rate after devaluation, Polish exports to the West surged from around $8.5 billion in 1989 to $11.5 billion in 1990. This surge continued in 1991 and has helped pay for a sharp rise in imports. The import boom reflects the eager purchase of Western durable goods previously unavailable and a substantial increase in imported machinery as well as energy to offset the loss of supplies from the Soviet Union.

Although efforts at liberalizing and stabilizing the Polish economy have been largely successful, privatization, especially of large enterprises, lags dangerously. This is the nation's key problem: unremedied, it could lead to the return of profound economic instability.

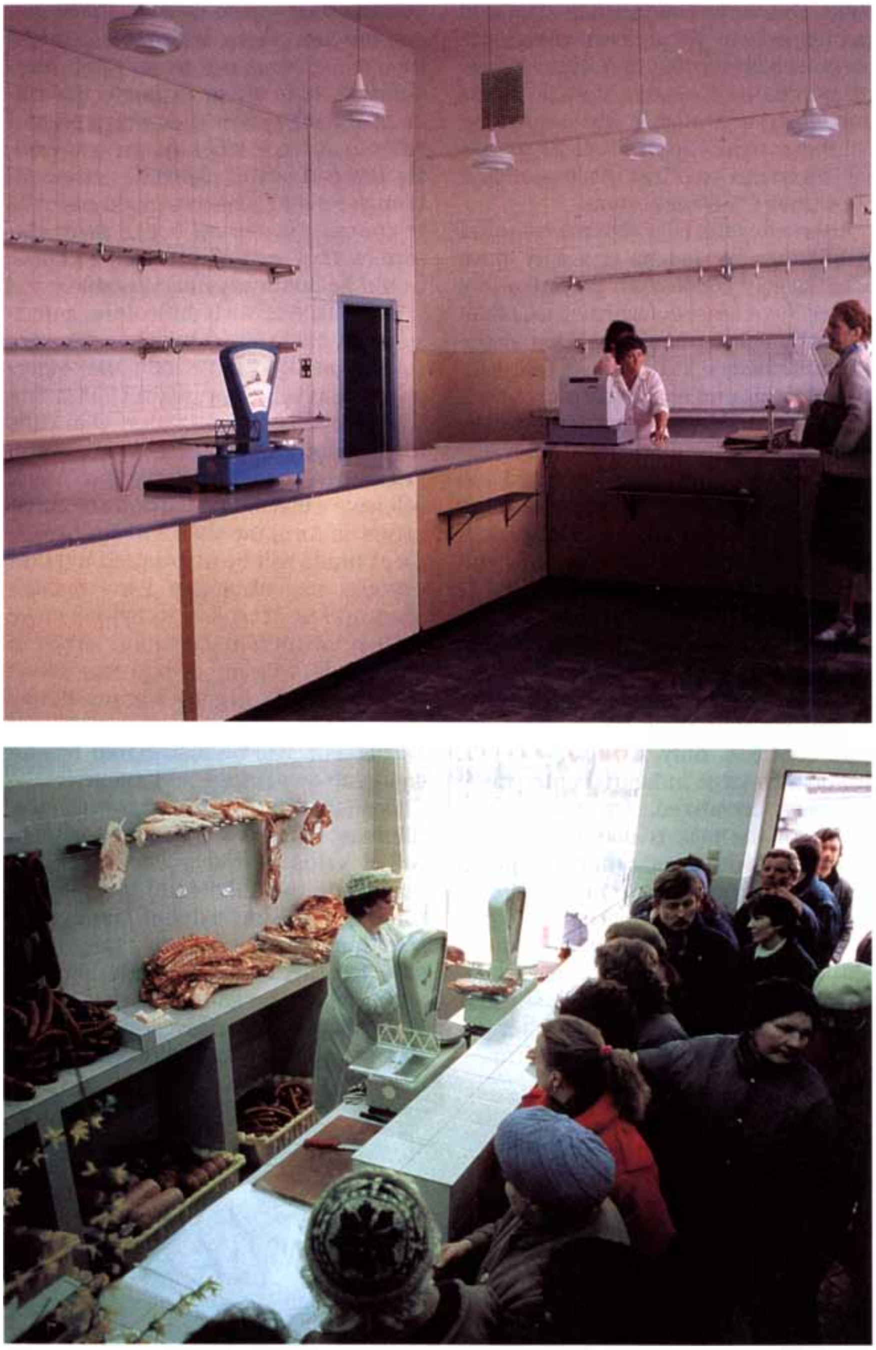
The privatization of small enterprises (those with fewer than 500 employees) has moved forward rapidly. Trucking, construction and small industrial units are now largely in private hands, most either auctioned off or sold to their employees. By mid-1991 more than 40,000 shops had been sold or leased to private operators; combined with the growth of private firms in the service sector, this has privatized an estimated 80 to 90 percent of all retail trade.

In contrast, only a small fraction of Poland's large industrial enterprises has been privatized. The sheer magnitude of the task is daunting. Under the aggressive privatization campaign of the Thatcher regime in the 1980s, for example, the U.K. succeeded in privatizing about five state-owned enterprises a year. Even at that rate—which in the U.K. relied on a sophisticated capital market and a private sector capable of absorbing state-owned firms—privatization would take the Poles several hundred years. Without a long history of performance in an open market, it is difficult to estimate the value of the state-owned enterprises, and without a large stock of private capital, few can buy at plausible prices. Although foreign investors might have the resources to snap up Polish firms (assuming they wanted to), such a course would be politically unacceptable.

In the face of such difficulties, a decision was made last summer to privatize several hundred large industrial enterprises by giving shares to all Poles. The so-called Mass Privatization Plan calls for distributing shares among several private investment funds, each of which will have a diversified portfolio of enterprises. In turn, the shares of the investment funds will be distributed to adult citizens, free of charge. Foreign managers will be recruited to help oversee each of the investment funds, although the funds will be owned by Polish households. At the same time, a fraction of each enterprise's shares, around 10 percent, will be distributed free of charge to its workers and managers.

Once the shares are in private hands, the reasoning goes, a market for them will develop. Workers and managers may trade their shares to other investors, and the investment funds may also trade the shares among one another or to other investors, both domestic and foreign. The households may trade their shares of the investment funds, thereby putting market pressure on the

POLISH AGRICULTURE is still inefficient and labor intensive. Most farmers work plots no more than five hectares in size. Sharp cuts in subsidies are expected to force consolidation into larger, more efficient farms.



ECONOMIC LIBERALIZATION has led to a wider availability of consumer goods (albeit at higher prices). Photographs above show a butcher shop in 1989 (*top*), when price controls were still in force, and in 1990 (*bottom*), after they had been lifted.

investment funds to manage their portfolios wisely. Meanwhile the investment funds will appoint members to the board of directors of each enterprise that they own. These board members will scrutinize management as they do in the West.

Enterprise managers will therefore face several kinds of scrutiny and incentives that do not exist for state-owned firms. In addition to being motivated by their own share holdings, the managers will be watched by the directors of the enterprise, many of whom will be appointed by the investment funds. They will also face the market test of the value of the shares traded in the stock market. Enterprises whose prospects are bleak will find their shares falling in value, rendering them liable to takeover, or at least to a change of management. The value of well-run enterprises, in contrast, will climb, potentially making additional capital available for investment. Furthermore, once firms are in private hands, the political consequences of restructuring will be reduced, because the private sector rather than the government will be doing the work. Call for subsidies and protection should also be easier to resist.

This plan was still moving toward implementation at the end of 1991, although the inconclusive results of the parliamentary elections rendered its future uncertain. At least some funds will be established, but the method of distributing their shares to households remains a matter of debate. The number of firms that will eventually participate in the program is also open to question. Around 200 have been designated, and policymakers intend to distribute shares in additional groups of firms if the program shows signs of success.

In the meantime, most of Poland's large state-owned industry is left rudderless or worse. Under the communist regime, managers were controlled by state administrators or the Communist party, their commands backed up by the threat of state violence. In the West, and under the plans of mass privatization, managers are governed by a board of directors that is legally obliged to act on behalf of shareholders. (Although that does not always happen, at least the basic legal responsibilities and economic incentives are in place.)

As matters now stand in Poland, however, most managers answer neither to administrators nor to their nominal owner, the Treasury. They operate on their own behalf or under the control of a council of workers. Only in a small minority of cases has the government appointed a supervisory board to represent the interests of the Treasury as the legal owner of enterprises, a measure known as commercialization.

As a result, managers are all too ready to pay out enterprise profits in the form of higher wages, making industrial plants even less competitive than they already were. What wage control there is in Poland comes from highly unpopular centralized income policies, which are both economically inefficient and politically debilitating.

In addition, reports indicate widespread self-dealing and conflicts of interest among managers of state-owned enterprises. Managers have been accused of intentionally bankrupting firms in order to buy them out cheaply or establishing private firms that then receive preferential contracts with the state-owned enterprises. They have accepted unfavorable joint-venture and takeover offers that provide personal benefits while rejecting better offers that might put their jobs at risk. In short, all the ills of unbridled laissez-faire capitalism that plagued the West in the 19th century are running amok in Poland.

There are also the "zombie" enterprises, whose incompetent managers have failed to adjust to new market conditions. Many of the Polish firms hit by the collapse of the Soviet market, for example, simply continued to produce goods at normal rates and build up inventories while waiting for a miracle. In mid-1991 the money ran out, and a spate of politically and economically troubling insolvencies struck the country.

As long as the privatization of large industrial enterprises is delayed, wider use of commercialization is imperative. Although appointing a supervisory board to represent the interests of the Treasury does not establish private ownership, it does bring in some modicum of oversight of managerial behavior, and it does subject the enterprise to normal rules of corporate behavior.

Even this step, however, is being applied only slowly. Some policymakers underestimate the importance of establishing a clear legal framework for state enterprises, whereas others point to the real (but surmountable) logistic difficulty of finding qualified individuals to sit on supervisory boards. Some critics of commercialization offer concerns that abuse of power by supervisory boards could in effect return enterprises to centralized government control.

The structure of the Polish economy is already changing under the influence of the market forces that have been unleashed during the past two years. As additional reforms such as large-scale privatization take effect, those changes will accelerate. Ultimately, the pressure of international competition will probably play the largest role in restructuring.

Although it is impossible to predict long-term effects precisely—which industrial sectors will flourish and which will decline—some basic trends are already emerging. The industrial sector as a whole is shrinking, while the heretofore neglected service sectors (notably shops, restaurants, banking and personal services such as plumbers and mechanics) are growing. Indeed, although Poland is often characterized as being in a sharp recession, that is true only of the industrial sector. The service sectors of the economy are booming. Between December 1989 and June 1991, individuals started 460,000 businesses, operating mostly in services, and the payrolls of all small private firms grew by an estimated 860,000 workers.

Major changes are also afoot in agriculture. Poland never completed the great demographic shift from agriculture to industry and services that has been traversed by nearly every advanced industrial country. Agriculture still accounts for about 20 percent of total employment and 10 percent of GNP. Many small farmers are now under financial pressure as a result of the sharp cuts in Polish agricultural subsidies during the past two years. In the coming decade or two, most smaller farms will be consolidated into larger, more efficient operations.

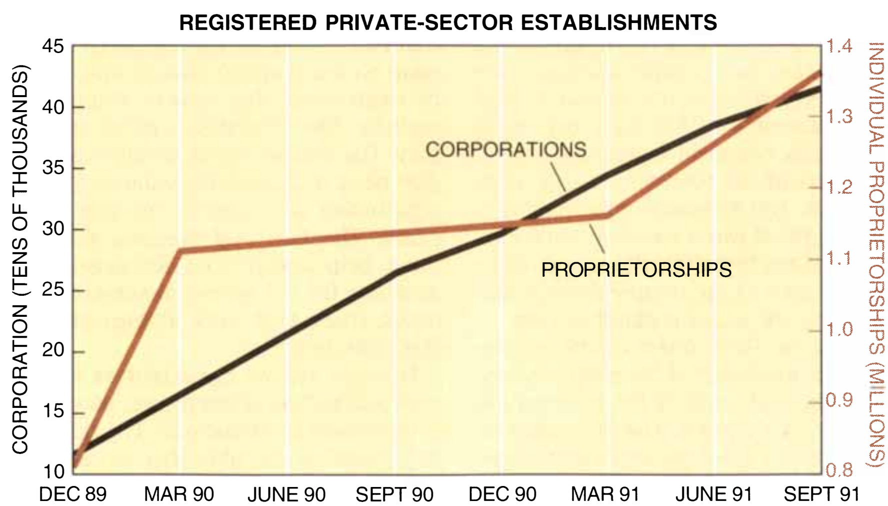
Most important for the long term, however, is a complete restructuring within what remains of the industrial sector, away from energy-intensive heavy industries such as steel and shipbuilding and toward more labor-intensive and skill-intensive industries that can compete in world markets. Poland offers low-cost skilled labor, particularly in engineering, that should make it an attractive place to produce high-value-added goods for export to Western Europe. Already some important restructuring is under way: the number of private industrial corporations grew by about 25,000 between December 1989 and June 1991.

Western firms are likely to set up operations in Poland to manufacture for export in the same way European firms are now investing in Spain, American firms in Mexico and Japanese firms in Korea, Taiwan and the rest of Southeast Asia. Thus far, however, concerns about unclear property rights and the political sustainability of Polish economic reforms have kept foreign direct investment at a low level. If the reforms are sustained, investment should increase markedly in the coming years.

Despite the accomplishments to date, Poland's situation remains decidedly fragile. Average citizens appear less taken by the fact that the burgeoning of firms promises future economic growth than by the rate at which existing industrial enterprises are going bust. Although they appreciate the end of shortages and queues, they are having trouble making low monthly earnings cover the higher market prices. Many are confused—they do not understand why the end of communism has not brought prosperity, why they have both free markets and low living standards and why economic reform's most visible product is unemployment. In a survey last summer, approximately 60 percent of Poles said they feared losing their jobs in the near future.

Significant parts of the population have lost ground in the first two years of reforms. The hundreds of thousands of farmers working tiny, inefficient plots, for example, have lost the subsidies that helped them stay afloat. Similarly, hundreds of industrial enterprises employing hundreds of thousands of workers are unable to compete under the new market conditions. The unemployment rate had risen to around 11 percent at the end of 1991.

Not all of the measured rise in unemployment is a result of reforms. Some people were always unemployed, but their numbers were not recorded by a regime that considered unemployment a crime rather than a misfortune meriting compensation. Many labor economists also suspect that a fair proportion of the unemployed hold jobs in the informal economy. Furthermore, the collapse of the Soviet economy and the breakdown of trade relations with the Soviet Union (now the Commonwealth of Independent States) have led to a rapid decline in its imports from



EXPLOSIVE GROWTH OF PRIVATE BUSINESSES has in effect privatized significant parts of Poland's economy even as large enterprises remain in state hands.

the Eastern bloc. Hundreds of enterprises in Poland and elsewhere are being forced to close their doors or scale back production dramatically. Insofar as enterprises can shift their exports from the East to the West, the reforms may in fact help cushion Poland from such external shocks. Nevertheless, much of the increase in unemployment is real, and the fears of unemployment are indisputable.

Although Poles blame the reforms for an alleged steep fall in living standards, this decline is almost surely overstated. Average wages have risen less than prices since the start of the reforms, thereby reducing measured real take-home pay. But, in general, goods were never actually available at pre-reform prices, at least not without the buyer waiting in a long queue or paying a bribe or paying extra amounts in the black market. According to survey data, average Poles are now consuming somewhat more meats, fruits and consumer durables than they did before, suggesting that greater availability has more than compensated for higher prices. Furthermore, people are saving millions of hours per week that used to be wasted standing in queues.

Some kinds of expenditures, however, do seem to have been cut back on average. One important area of reduced consumption is clothing. Particular social groups such as industrial workers and farmers have been hard hit by the reforms. And with the collapse of Soviet trade with Poland in 1991 and the end of subsidized sales of Soviet petroleum, the society suffered a significant loss of income, albeit not as the result of the economic reforms themselves.

In the current turbulent climate, the fate of policies for economic restructuring is uncertain. In postwar Germany, for example, Ludwig Erhard, now remembered as the father of the postwar German economic miracle, was widely criticized as the creator of high unemployment rather than high living standards. Not until several years after the start of his reforms did prosperity arrive. The Adenauer government of which Erhard was a member just barely survived the first difficult years, in large part thanks to the money flowing into Germany through the Marshall Plan.

In Poland, Balcerowicz will be remembered as the father of his country's economic miracle only if his reforms are given time to work. And time may be very scarce indeed, given the limited patience of the population and the powerful political forces supporting protection and subsidy of failing industries.

The greatest risk in Poland is that populist concern over short-term hardships, confusion over property rights and a splintering of political power in parliament will lead to a weak government unable to take the last decisive step to private ownership. In that case, the economic situation will almost surely deteriorate further, and as has happened in Latin America, democratic government could be as thoroughly discredited as its totalitarian predecessor.

The "valley of tears" that intervenes between the institution of reform and later prosperity has been observed in nearly every country that has undergone a radical economic transformation, from postwar Germany and Japan to Chile and Mexico in the 1980s. Passing quickly through the valley depends on political leadership and a social consensus to maintain a stable set of policies. If there is inconsistency or wavering, it is easy to get lost—Argentina, for example, has been wandering for 45 years.

Peculiar electoral traditions and social divisions increase the chance that the country will "wander in the wilderness." Poland organized its parliamentary elections in October 1991 according to a system of nearly strict proportionality that gives representation even to small parties receiving a small fraction of the overall vote. The election law amplified inherent divisions between peasants and workers, church followers and anticlericists, urban and rural voters, and so on; as a result, Poles elected representatives of no less than 29 parties. No party received more than 15 percent of the vote. The new government will therefore surely be based on a large coalition, and history suggests that such governments are easily paralyzed by internecine disputes.

And even if the Poles can forge a solid internal consensus, that will not be enough for the new economic policies to succeed. Attractive as the hope might be that Poland and the other Eastern European nations can get back on their feet by themselves, that view is simply not realistic. The West has a crucial role to play. The former Soviet satellites are in dire need of financial assistance for restructuring now and in the next few years. The U.S. and Western Europe must help assure that resources are available for the modernization of factories, roads and other utilities during this delicate period.

Perhaps just as important as ready cash and technical assistance, however, is the West's symbolic role. The "return to Europe" is the glue that so far has held the social consensus for reform together. If Western Europe shows no interest in the return of Poland (or the rest of Eastern Europe), then the intense pressures already facing the new democracy are likely to undermine the course of reform and restructuring. Worse yet, given the stingy attitude of the West to date and the undoubted cases of rapacious Western investors intent on profiteering, a xenophobic backlash may already be taking shape.

To aid restructuring, the European Community should signal Poland that it can indeed return to Europe. It should make it absolutely clear that Poland can expect European Community membership after it has passed the key hurdles of economic reform. Nothing would so much channel energies and political passions for reform.

The signing of association agreements between the European Community and Czechoslovakia, Hungary and Poland at the end of 1991 marked an important move in this direction. These agreements foresee free trade between the three countries and the European Community in industrial goods by the end of the decade. Unfortunately, the agreements stop short of what is needed, both in regard to membership and on the important issue of agricultural trade. The European Community made some concessions, but its protectionist framework remains intact. Clearly, Western Europeans still have deep anxieties about closer links with the East, and these misgivings could lead to considerable political trouble.

Europe stands at a crucial point in its history. The cold war between democracy and totalitarianism that divided the continent for half a century is over. Communism has lost. But unless all Europeans work together to help build market economies in the nations of Eastern Europe and to integrate them on an equal footing in international trade, the victory over communism could be followed by chaos and confusion rather than economic renewal and growth.

FURTHER READING

CREATING A MARKET ECONOMY IN EASTERN EUROPE: THE CASE OF POLAND. David Lipton and Jeffrey Sachs in *Brookings Papers on Economic Activity*, No. 1, pages 75–147; 1990.

THE ROAD TO A FREE ECONOMY: SHIFTING FROM A SOCIALIST SYSTEM: THE EXAMPLE OF HUNGARY. Janos Kornai. W. W. Norton & Company, 1990.

ACCELERATING PRIVATIZATION IN EASTERN EUROPE: THE CASE OF POLAND. Jeffrey Sachs. Washington, D.C., World Bank Annual Conference on Development Economics, 1992.

STRUCTURAL ADJUSTMENT AND INTERNATIONAL TRADE IN EASTERN EUROPE: THE CASE OF POLAND. A. Berg and J. Sachs in *Economic Policy* (in press).

The One Time Of Year When Those Who Know The Joy Of Making Money Are Consumed By The Misery Of Losing It.

IN THE WORDS of the famous poet, T. S. Eliot, "April is the cruelest month."

Indeed it is. The thought of writing big checks for income tax does little to quiet one's soul.

After all, it is *your* money we're talking about. You earned it. You paid taxes on it all year long.

And now (as if that weren't enough), you also get to pay taxes on your investments.

The human mind is quite ingenious in the ways it finds to cope with such financial dilemmas. But none are as practical as the way we're about to suggest: a tax-free* investment from John Nuveen & Company.

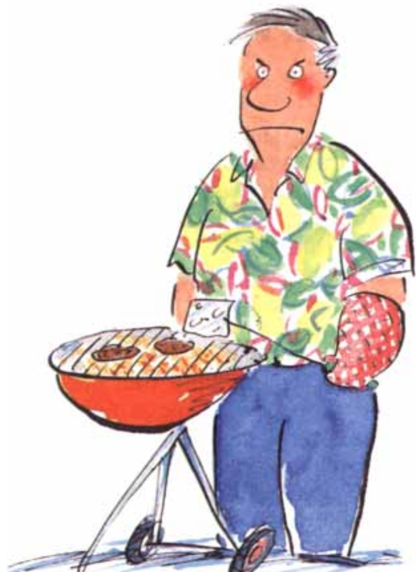
It will help you keep the money you earn. It is also our specialty.

With nearly a hundred years experience, you'll find we have an insight into the municipal bond business that, in our opinion, few other companies can match.

**Income may be subject to state and local taxes. Capital gains, if any, will be subject to capital gains taxes.*

This being the case, it should come as no surprise that we review billions of dollars worth of bonds each year. And only after a painstakingly thorough process of research do we invest in a select few (including some that other investment companies have overlooked).

Francine became frazzled.



Orville became ornery.

Myron became morose.

All of which are scrupulously managed for the long term with one goal in mind: to provide a high level of tax-free income over time, while still protecting your initial investment.

So if you would like to learn more about how it's possible to keep more money (and who wouldn't), we suggest you ask your financial adviser or your personal banker about Nuveen tax-free investments (for current and future income).

Or you can call us toll-free at 1-800-524-6500. We'll send you a free booklet that just might help next April be a joyous one.

For more complete information on Nuveen tax-free open end mutual funds, including charges and expenses, call for a prospectus. Read it carefully before you invest or send money.

NUVEEN

Specialists In Tax-free Investments Since 1898.

The Patch Clamp Technique

A simple procedure can easily isolate ion channels on cell membranes. Its Nobel Prize–winning developers explain what the technique has revealed about cellular signaling

by Erwin Neher and Bert Sakmann

The transmission of signals within and between cells is mediated by ion channels, pore-forming proteins embedded in the plasma membranes of nearly all cells. Within the body, ion channels produce the flickers of electrical activity that stir neurons and muscle cells. In sensory organs the channels translate physical or chemical stimuli into electrical signals for the nervous system. Even cells not connected to the central nervous system, such as those in the blood, immune system, liver and other organs, use ion channels for signaling processes.

Since the 1950s biologists have been able to study the electric currents arising from these ion fluxes at a macroscopic level. Only since the 1970s, however, could they examine the individual ion channels themselves. The patch clamp technique, for which we recently received the Nobel Prize for Physiology or Medicine, made these details accessible for the first time.

Although its development and refinement took several years, the technique is fundamentally simple: a thin glass pipette of the proper shape is tightly sealed against a cell membrane, thereby isolating a small patch of the membrane and the ion channels it contains. These channels can then be chemically or electrically manipulated and their properties deduced. A researcher can even remove a patch of membrane from a cell or carefully open a window into a living cell to alter its cytoplasmic constituents. In all these various applications the patch clamp technique makes it possible to probe how ion channels affect membrane voltage and cell processes such as secretion and contraction.

Several hundred laboratories throughout the world have adopted and extended the patch clamp technique since we introduced it in 1976. This summary of what the procedure has revealed about ion channels and their functions draws on their many contributions and not just our own.

Much of our early interest in ion channels was inspired by two fascinating papers published in 1969 and 1970, when we were doctoral students of membrane biophysics at the Max Planck Institute for Psychiatry in Munich. The papers concerned artificial lipid membranes, which in their pure form are electrical insulators. Yet Ross C. Bean and his colleagues at the Ford Aerospace Corporation in Newport Beach and Steven Hladky and Denis Haydon of the University of Cambridge demonstrated that if trace quantities of certain antibiotics or proteins were inserted into the membranes, they became electric conductors. The discrete changes in the current passing through the membranes suggested that the proteins created porelike channels that opened and closed individually. Charged ions could then traverse the membrane through the open channels.

It became clear to many investigators that the electrical signals in neurons and other cells must be mediated by similar proteins in the lipid plasma membranes surrounding them. Alan L. Hodgkin and Andrew F. Huxley of Cambridge had previously invoked the concept of ion channels in their classic analysis of currents through nerve membranes, for which they received the Nobel Prize in 1963. All of us interested in membrane biophysics anticipated that if measurements at an appropriately fine resolution could be taken, a whole microcosm of signaling molecules would be found.

In the following years, evidence continued to accumulate that ion channels do operate in cell membranes. Most notably, in 1972 Bernard Katz and Ricardo Miledi of University College in London conducted a statistical analysis of voltage fluctuations at the neuromuscular junction (the synapse between a motor neuron and a muscle fiber). They concluded that the synaptic signals consisted of small electrical events of the same magnitude as those associated with the artificial channels.

But Katz and Miledi could only infer the properties of the channels from their analysis, which depended on several assumptions. No method was available for directly measuring the unit events constituting the synaptic signal. The background noise associated with the standard techniques for measuring the electric current passing through a cell membrane was only one ten-billionth of an ampere, but it was 100 times greater than the current of a unit event and drowned out its signal.

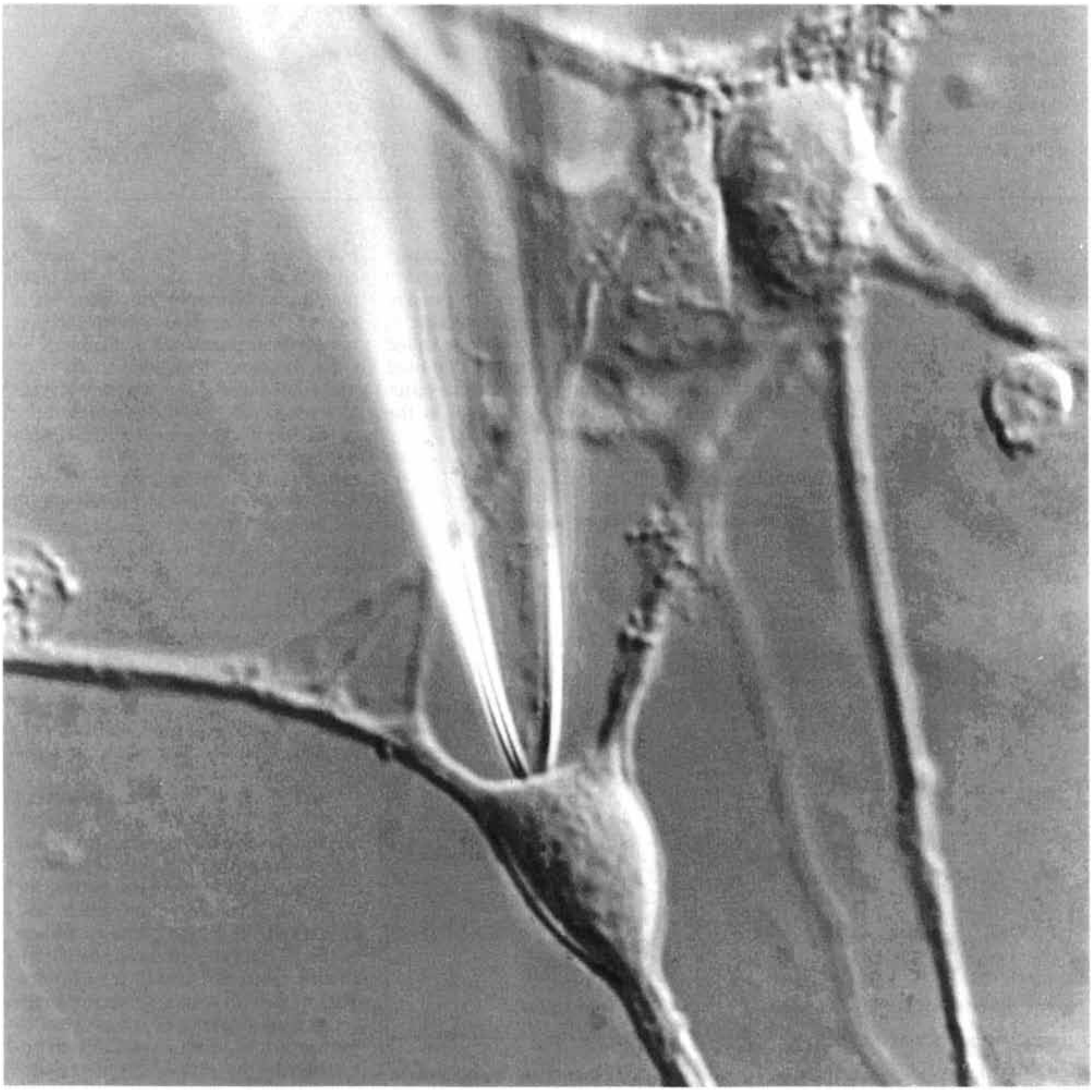
In 1973 we decided to tackle the problem of reducing the background noise. We knew that the available electronic components would have the reso-

ERWIN NEHER and BERT SAKMANN are the corecipients of the 1991 Nobel Prize for Physiology or Medicine for describing the function of single ion channels using their patch clamp technique. Neher is director of the membrane biophysics department of the Max Planck Institute for Biophysical Chemistry in Göttingen. After completing his undergraduate studies in physics at the Institute of Technology in Munich in 1965, he went on to earn a master's degree in 1967 from the University of Wisconsin, where he was a Fulbright Scholar. He then returned to the Institute of Technology and completed his doctorate in 1970. Neher has also worked as a visiting scientist at Yale University and the California Institute of Technology. He is a corresponding member of the Bavarian Academy of Sciences and a foreign associate member of the U.S. National Academy of Sciences. Sakmann is director of the cell physiology department of the Max Planck Institute for Medical Research in Heidelberg. He received his medical degree from the University of Munich in 1969. Between 1970 and 1973 he studied biophysics as a British Council scholar at University College in London with Nobel laureate Bernard Katz. Before moving to Heidelberg in 1989, he worked at the Max Planck Institute in Göttingen.

lution necessary for measurement only if a small patch of membrane could be insulated from the rest. The approach we took was to place a glass micropipette onto the surface of enzymatically cleaned muscle fibers. The nonconductive glass pipette, we hoped, would isolate a few ion channels, thereby providing us with a clear signal.

Unfortunately, obtaining a tight seal between a measuring glass pipette and a membrane was not easy. Like many others who had used extracellular pipettes before us, we had to put up with electrical leaks that connected the extracellular fluid with the interior of the pipette. Nevertheless, by carefully cleaning the cell surface and by optimizing the shape and size of the pipette, we succeeded in observing individual channel currents in response to acetylcholine, the transmitter (signal-inducing chemical) at the neuromuscular junction. That early experiment confirmed many previous inferences about the elementary currents through ion channels, in particular the assumption that they were pulselike events with constant amplitudes and varying durations.

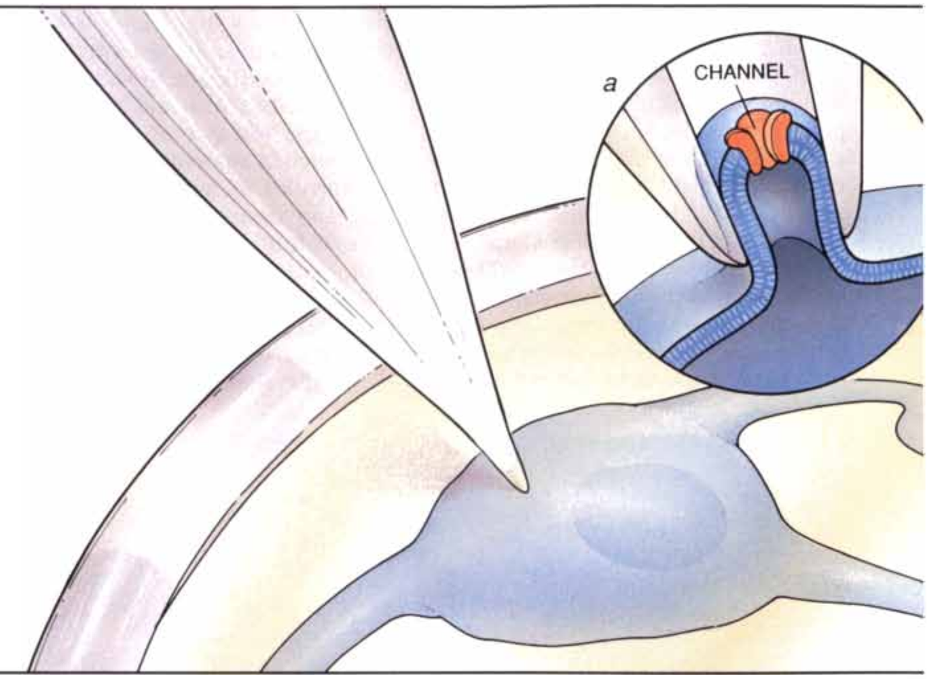
The low quality of the pipette-to-membrane seal and the attendant background noise initially precluded us from making detailed recordings of ion channels other than those found at the neuromuscular junction. A few years later we discovered by chance that the application of a slight suction through the pipette, together with some other procedural changes, could raise the resistance of the seal to more than a billion ohms—an improvement of several orders of magnitude. By gently pulling



TIGHTLY SEALING A PIPETTE against the surface of a neuron enables researchers to study the ion channels in the outer plasma membrane. The pipette, only 1/25,000 the diameter of a human hair, physically and electrically isolates the trapped channels. This technique, known as patch clamping, can record the opening and closing of ion channels. In laboratories around the world, the technique is being used to discover the details of signaling networks within cells.

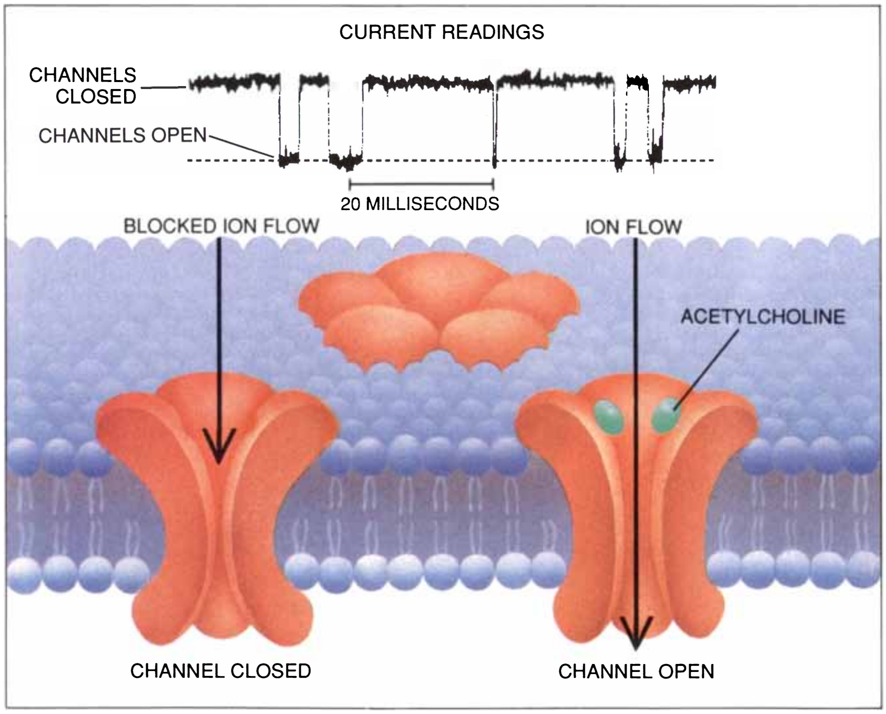
Three Forms of Patch Clamping

By pressing a patch pipette against the enzymatically cleaned surface of a cell and applying gentle suction, researchers can place a giga-ohm seal around a small patch of the cell membrane and the ion channels it contains (*a*). An experimenter can then apply various stimuli from within the pipette and measure the behavior of the trapped channels. Alternatively, the experimenter may detach the membrane patch from the cell, thereby exposing the cytoplasmic mouth of the channels (*b*). If the membrane patch can be ruptured without breaking the giga-ohm seal (*c*), the experimenter can alter the constituents of the living cell's cytoplasm.



with the attached pipette, we found it was even possible to excise microscopic patches of the membrane for study in isolation. Thereafter, the recording of single ion channels became a high-resolution technique.

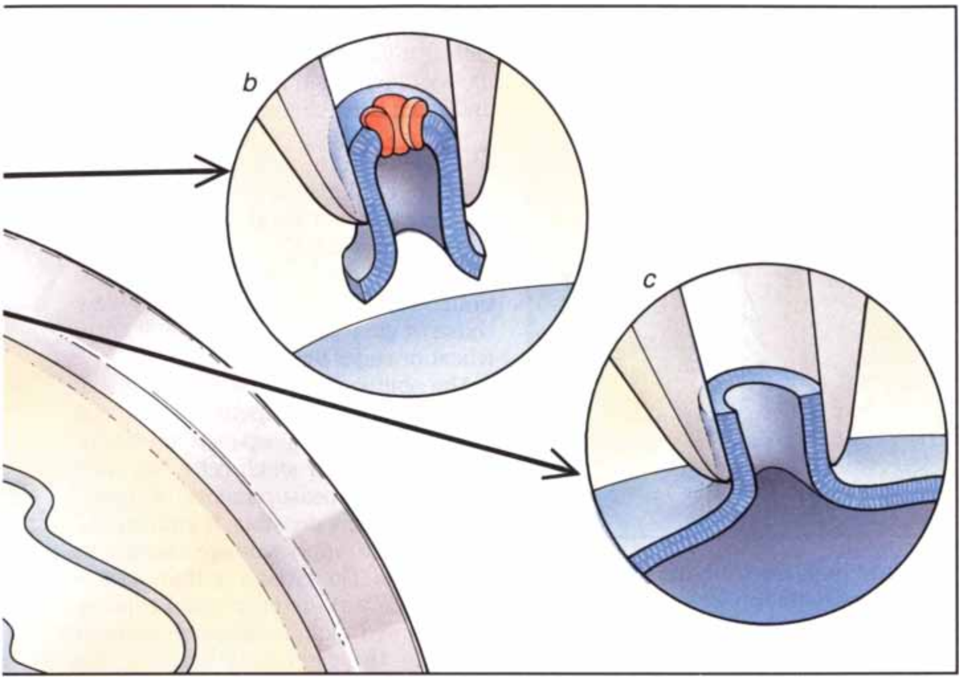
The most detailed information on the role of ion channels in synaptic transmission has been obtained from experiments on the neuromuscular junction of skeletal muscle. There the presynaptic neurons release acetylcholine in discrete multimolecular packets called quanta. Acetylcholine molecules bind transiently to acetylcholine receptor channels, specialized proteins in the postsynaptic membrane, and cause a current to flow across the end-plate membrane. This end-plate current is the sum of the elementary currents through hundreds of thousands of channels.



RECEPTOR CHANNELS found at the neuromuscular end plate open in response to the transmitter acetylcholine. When no acetylcholine is present, effectively no current passes through a channel. When acetylcholine binds to the receptor, an elementary current of a few picoamperes flows. The measured durations of the currents and the intervals between them vary because the interaction of acetylcholine molecules with the receptors is governed by probability.

To measure these elementary currents individually, one can press the tip of a patch pipette onto a muscle fiber at the end plate. This region of muscle surface contains the acetylcholine receptor channel. When a low concentration of acetylcholine is in the pipette solution, the current recorded with the pipette switches between two levels. At one level, effectively no current flows because all the ion channels in the membrane patch are shut. When one channel molecule flips to the open state, because of a voltage applied to the membrane, a current of about 2.5 picoamperes abruptly flows through it. After a variable period, the molecule flips back to the shut state, and the current switches off.

The opening of the receptor channel is triggered by its binding to acetylcholine; the unbinding of acetylcholine causes the channel to close again. The randomness of the lengths of time for which the channel stays open and closed reflects the probabilistic nature of the interactions between the acetylcholine molecules and their receptors. The amplitude of the current steps represents the capacity of the channel to transport ions, such as sodium or potassium in the case of the acetylcholine receptor channel. By comparing the measured current amplitudes and duration distributions with the predictions of various hypotheses, an investigator can determine how ions interact



cent. Given the high concentration of acetylcholine in the synaptic cleft, end-plate channels can act like rapid, chemically gated current switches.

Membrane channels are apparently heterogeneous: in the same membrane patch, two or more classes of closely related channels often seem to be activated. Pharmacologists have known for many years that different subtypes of receptors exist, but the sharper tools now available in the form of recombinant DNA techniques have shown that the diversity of subtypes is far greater than had been imagined. For each type of transmitter- or voltage-gated channel, there are several versions or subtypes with different conductance or gating properties that form a channel family. External stimuli and the developmental stage of an organism can alter the mosaic of channel subtypes found on cell membranes.

A primary aim in the understanding of membrane channels is to relate their functional properties to their three-dimensional structures. One way to attain this goal is to locate and change critical sequences of amino acids in the channel protein and then to observe the effect of the alterations on the channel's function.

After the DNA encoding each subunit of the acetylcholine receptor had been cloned and sequenced, it became possible to perform such experiments. In collaboration with Shosaku Numa of Kyoto University and his research group, we performed such studies on normal and genetically altered acetylcholine receptor channels in the membranes of host cells, such as eggs from the clawed frog *Xenopus*. From copies of DNA that encoded the acetylcholine receptor, we synthesized complementary messenger RNA molecules in the test tube. These RNA sequences could be injected into egg cells, which then translated the genetic information into receptor proteins and inserted them into the cell membranes. The functional properties of these recombinant receptor channels closely resembled those of end-plate channels at the neuromuscular junction.

This combination of techniques helped to elucidate the structural differences between subtypes of end-plate channels in muscle. Elementary end-plate current recordings had revealed that mammalian muscle produces two subtypes of end-plate channels: a low-conductance subtype that appears predominantly in fetal and neonatal muscle and a high-conductance subtype with different gating properties that is expressed in adult muscles. During postnatal development, the fetal sub-

with a channel molecule and how the interaction between a transmitter and its receptor controls the opening and closing of the channel.

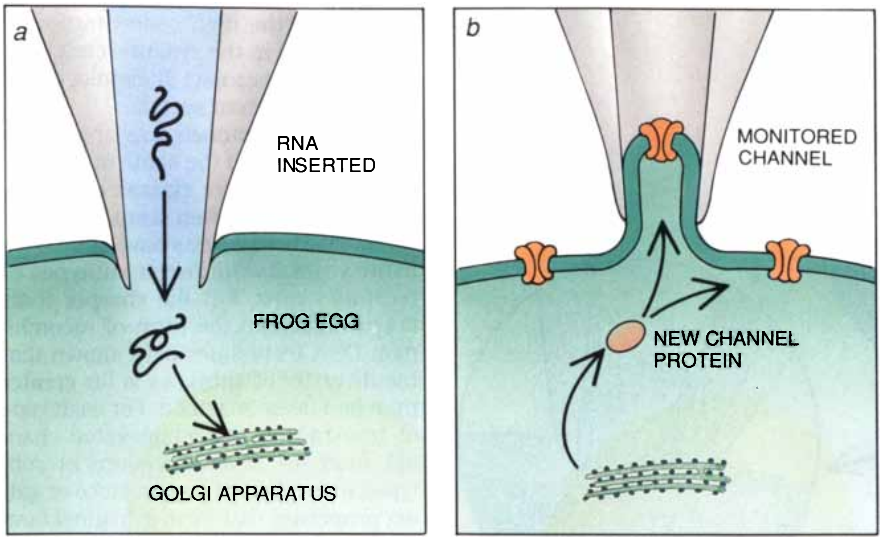
In synapses of the central nervous system, amino acids such as glycine, gamma-aminobutyric acid (GABA) and L-glutamate are the most prominent signaling substances for rapid communication. The pulselike shape of the currents measured for channels that bind these transmitters indicates that they, too, open and close randomly. The receptor channels may therefore work in essentially the same way as do the acetylcholine receptor channels at the end plate. Nevertheless, transmitter-gated channels in the central nervous system often show an additional complexity, in that some channels may be only partially open or closed and that different subtypes of the channels may occur in various brain regions.

The transmission of information from the central nervous system to the neuromuscular junction needs to be extremely rapid. In the myelin-coated neuronal axons that carry those signals in vertebrates, conduction is not mediated by transmitter-gated channels like those we have already described. Instead it is mediated by faster channels that respond to changes in membrane voltage—the difference in electrical potential between the inside and outside of the cell.

Voltage-gated channels for sodium ions underlie the rapid rise of the action potential in neurons. Analyses of the elementary sodium currents indicate that the voltage-sensitive channels switch between two states and have a high probability of opening shortly after the beginning of a voltage change. The currents supporting the nerve impulse are generated by the superposition of tens of thousands of such elementary sodium currents. Voltage-gated channels for other ions, such as potassium and calcium, seem to operate in much the same way as sodium channels do and share many of the same structural features.

Patch clamp investigations have revealed the dynamic molecular mechanisms of transmitter- and voltage-gated channels at a level of detail even finer than that we have discussed. In almost all the ion channels studied, a close look reveals that the elementary event is not a single current pulse but a series of transient pulses separated by brief gaps. When a receptor channel in the end plate of a neuromuscular junction binds with acetylcholine, for example, it opens and closes several times before the acetylcholine finally dissociates from it.

We investigated the structure of these transitions for end-plate channels in collaboration with David Colquhoun of University College in London. Using probability theory, we estimated the number of states an acetylcholine channel adopts during its succession of openings and closings and also the rates of the transitions from one state to another. Such detailed studies indicated that each receptor has two binding sites for acetylcholine. When both sites are occupied by acetylcholine molecules, the probability that the channel will open is close to 100 per-



GENETICALLY ENGINEERED FROG EGGS are useful tools for studying ion channel behavior. RNA sequences encoding subunits of ion channels can be injected into an egg (*a*). The egg will then synthesize the channel proteins and insert them into its surface membrane, where they are accessible for patch clamp experiments (*b*).

type gradually disappears and is replaced by the adult subtype. As experiments with recombinant DNA techniques revealed, the shifting mosaic of channel types is generated by changes in the expression of genes that encode the channel subunits.

Similar techniques were also used to locate the amino acid sequences in the acetylcholine receptor protein that form the inner wall of the membrane channel. An analysis of the amino acid sequences in the subunits had suggested that they each contained four membrane-spanning segments, designated M1 through M4. The channel of the acetylcholine receptor seemed to be lined by several transmembrane segments, like the individual staves of a barrel, each contributed by one subunit.

Through recombinant DNA techniques, we created the genes for chimeric channels, made up of subunits from species (cows and electric rays) whose own channels had different conducting properties. By analyzing the properties of these chimeric channels, we learned that the M2 segment and its neighboring regions contain important determinants of ion transport. Directed mutations that caused amino acid substitutions within these regions helped us refine our map of where these determinants are located.

We found that three clusters of negatively charged amino acids may form rings at the extracellular and intracellular mouths of the channel; a similar ring of polar amino acids is present in the transmembrane portion close to the intracellular mouth. (Polar amino acids have electrically polarized structures even though they have no net charge.) These negatively charged rings drastically influence the rate of current flow and may select particular ions for transport across the membrane: positively charged sodium and potassium ions would be drawn into the channel, and negatively charged chloride ions would be excluded from it. Our results also suggested that the polar amino acids located close to the intracellular part of the M2 transmembrane segment formed the narrowest part of the acetylcholine receptor channel.

The usefulness of the patch clamp technique is not limited to revealing the molecular details of channel function. Patch pipettes can also be used to study signaling mechanisms at a cellular level, a procedure called voltage clamp analysis. Indeed, when measuring events in small cells, the technique has advantages over the use of conventional microelectrodes.

Conventional voltage clamp analysis has contributed immensely to our understanding of signaling processes in the nervous system. It was introduced by Kenneth S. Cole of the Marine Biological Laboratory at Woods Hole in 1949; Hodgkin and Huxley used it to unravel the basic mechanisms of nerve excitability. The technique involved "taming the axon," as Cole put it, by forcing a transmembrane potential onto the axon of a neuron. The resulting membrane currents could then be measured and interpreted.

Unfortunately, most voltage clamp techniques require that either axial wires or at least two microelectrodes be inserted into a cell, which is generally possible only with the largest types of animal and plant cells. Mammalian cells, for example, typically have diameters of no more than 10 to 30 microns and can barely tolerate impalement with a single standard microelectrode. Before the patch clamp technique was developed, we therefore knew more about the action potential in the giant nerve axon of the squid than about the nerve impulses of the human brain; the best-understood signals in plant cells were those of giant algae, not those of corn, wheat or sugar beets.

The ability to create a giga-ohm seal between a patch pipette and a cell membrane literally opened up mammalian and other small cells. Not only did it make measurements of membrane channels possible, it enabled researchers to apply voltage clamps to small cells. The reason is that, with a bit of luck, a researcher could rupture the clamped membrane patch without breaking the tight seal between the pipette and the surrounding membrane, thereby obtaining access to the cytoplasm. That configuration, termed whole-cell recording, resembles a classic microelectrode impalement. It is tolerated by much smaller cells, however, and it provides much better control over the intracellular milieu.

Whole-cell recording with patch pipettes has become more or less the standard for work with mammalian cell cultures. Even human red blood cells and platelets—cells only a few microns in diameter—have been voltage-clamped with patch pipettes. Thus, most cell types of clinical interest have become amenable to a biophysical analysis, and many disorders, such as cystic fibrosis, have been traced back to defects in channel function.

The new accuracy that whole-cell recordings brought to investigations of electrical signals in the central nervous system was particularly welcome. Because conventional voltage clamp techniques are cumbersome, they can often be applied only to isolated neurons. Those results have limited biological significance because the cells are not in contact with their natural neighbors. Channels studied in cultured cells also may not have properties identical to those prevailing in the intact organ. These ambiguities are particularly troubling in studies of cells in the central nervous system, which has a highly elaborate architecture and specificity.

We therefore developed a way to apply the whole-cell recording method to brain tissue in which neurons are left in contact with their natural neighbors. Per Anderson of the University of Oslo and Tomo Takahashi of the University of Kyoto had devised techniques for re-moving a small, intact slice of brain tissue about one centimeter across. The surface of the neurons is freed of glial cells and other adhering tissues by gently blowing a stream of bathing solution onto the surface of the slice. Subsequently, the tip of a patch pipette can be sealed onto an exposed neural cell body, and the intervening membrane can be disrupted for a whole-cell recording. With such an arrangement, the synaptic currents can be recorded with a resolution 10 to 50 times higher than that possible using intracellular microelectrodes.

When one neuron releases glutamate or other transmitters from its presynaptic structures, it induces excitatory postsynaptic currents (EPSCs) or inhibitory postsynaptic currents (IPSCs) in a neighboring neuron. (EPSCs increase the probability that a neuron will fire an action potential and release transmitters of its own; IPSCs decrease that probability.) In whole-cell recordings, peak amplitudes of EPSCs and IPSCs fluctuate randomly in successive trials. This observation suggests that EPSCs and IPSCs result from the almost simultaneous superposition of quantal events, in much the same way quantal events create the end-plate current at the neuromuscular junction.

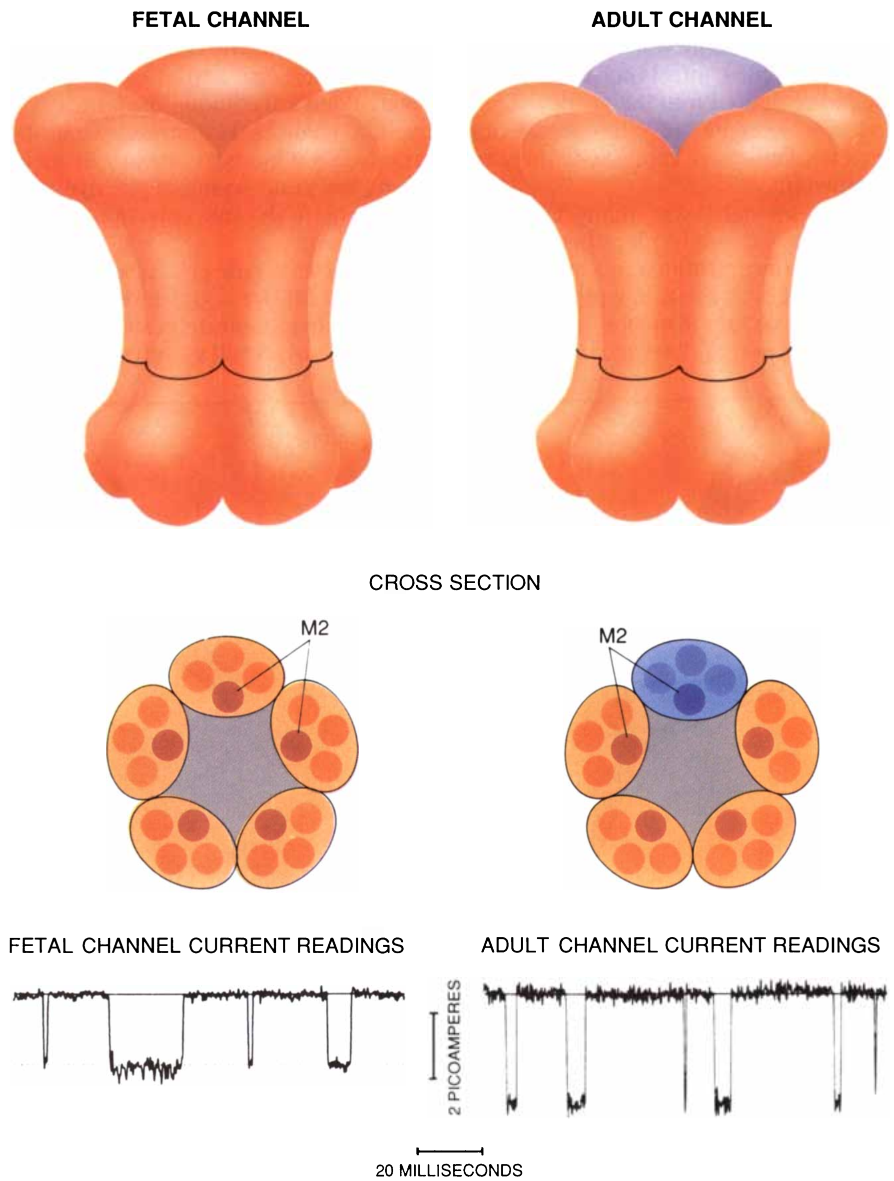
Surprisingly, we found that only about 20 to 30 postsynaptic receptor channels seem to open in response to a quantum of transmitter molecules released from a single presynaptic vesicle. This number is orders of magnitude smaller than at the neuromuscular synapse. Within the nervous system, that limited response probably allows the fine-tuning of excitation and inhibition signals impinging on a neuron.

The resolution of such quantal events in the central nervous system now enables us to distinguish the presynaptic and postsynaptic factors that underlie changes in synaptic efficiency. Such changes have been observed as a consequence of the usage or training of a neuron. Understanding synaptic efficiency may therefore enlighten us about learning and memory.

Ion channels are part of a complex network of signaling mechanisms that also involves second messengers—molecules capable of carrying signals from the surfaces of cells to their interiors by diffusing through the cytoplasm. The activities of many types of channels that respond primarily to external transmitters or voltage changes are influenced or modulated by these intracellular substances. Ion channels also have a reciprocal effect on the functions of second messenger molecules: virtually all the known second messengers interact with certain ion channels in some way, either directly or through other molecules, including kinases (which are enzymes that add phosphate groups to molecules) and G proteins (which couple receptors to enzymes that catalyze the formation of many second messengers).

For instance, a large fraction of the potassium ion channels in neurons are affected by both the membrane voltage and the concentration of free intracellular calcium ions. When intracellular calcium is low, the potassium channels do not open unless the depolarizations (changes in membrane voltage) are very large; when calcium concentration is high, they start to open at moderate depolarizations or even at the resting potential.

The channels that specifically gate the entry of calcium ions into a cell are themselves regulated by calcium ion concentrations. These channels open in response to depolarizations, but some types are deactivated by the calcium ions entering the cell. This mechanism can be viewed as feedback inhibition regulating the levels of calcium, in turn an important second messenger for many cellular activities.



STRUCTURE OF ACETYLCHOLINE RECEPTOR CHANNELS relates directly to their function. Each channel consists of five subunits, each containing four transmembrane segments. The M2 segment from each subunit lines the inside of the channel, like a stave in a barrel. Fetal mammals produce an acetylcholine receptor in their neuromuscular end plates that differs in just one subunit from receptors found in adults. Because a fetal channel has a subunit that differs in its M2 segment from that of an adult, the elementary currents that pass through it are smaller.

It is not surprising that such a complex network of mutual interactions, involving feedback loops and synergisms, has proved very hard to sort out. Recording the electrical activity that resulted from stimulating cells was not enough to uncover the links in the network—some of the intracellular regulators had to be controlled. The whole-cell method provided this control by permitting experimenters to manipulate the contents of a cell.

Because the passageway connecting the interiors of the pipette and the cell is fairly large, molecules diffuse rapidly from one into the other. That fact raised some experimental difficulties. In the early whole-cell studies, researchers found that subtle control mechanisms were simply washing out: many of the soluble control molecules in the cell were diluted into the nearly infinite relative volume of the measuring pipette. Only very robust channels remained intact.

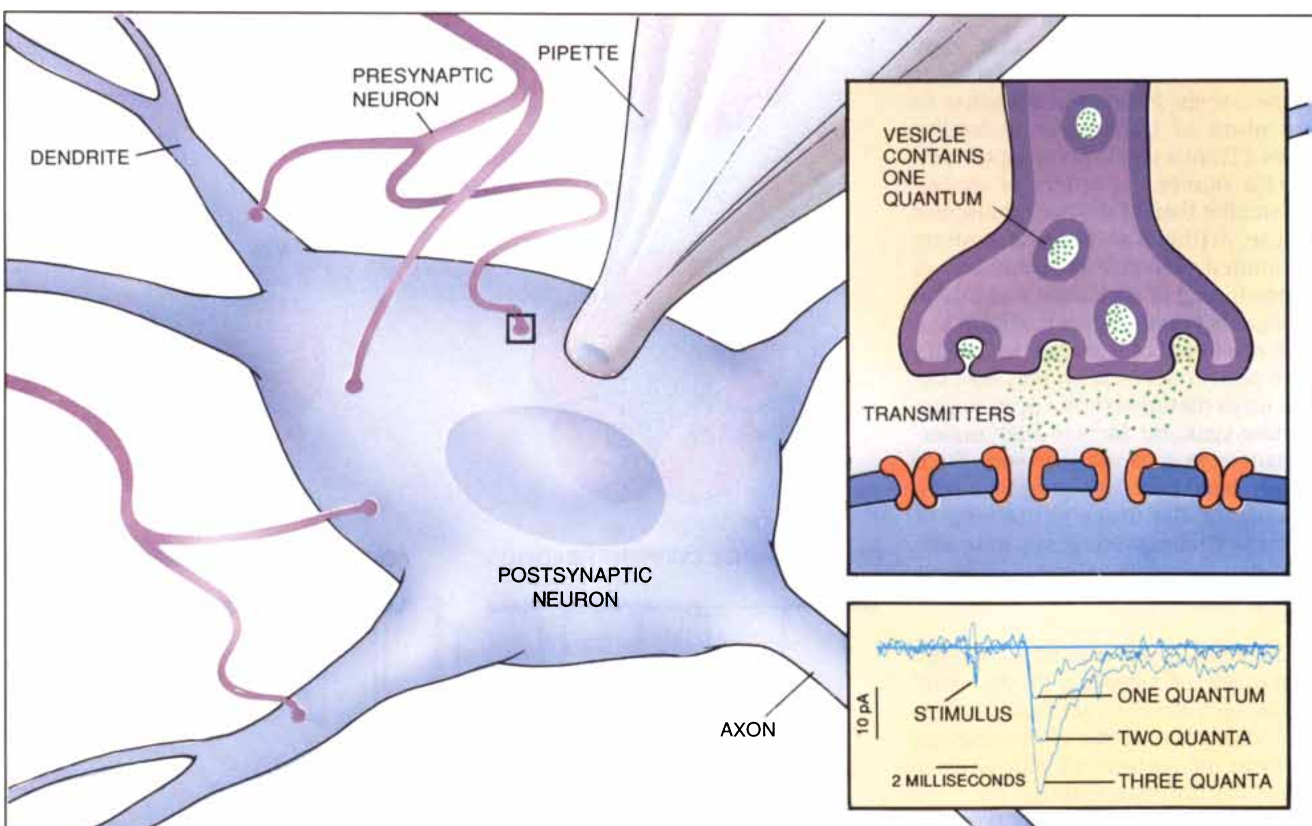
Over several years, many laboratories have gradually identified the required concentrations of substances that must be present in a pipette to retain the proper function of various channel types. In this way, the network of interactions between receptors, second messengers and ion channels has slowly become visible, making it possible to reconstitute some of the signaling pathways.

The message that ion channels convey to cells very often involves a change in intracellular calcium. We became interested in how this signal is passed on to elicit cellular responses and decided to focus on the role of calcium in secretion.

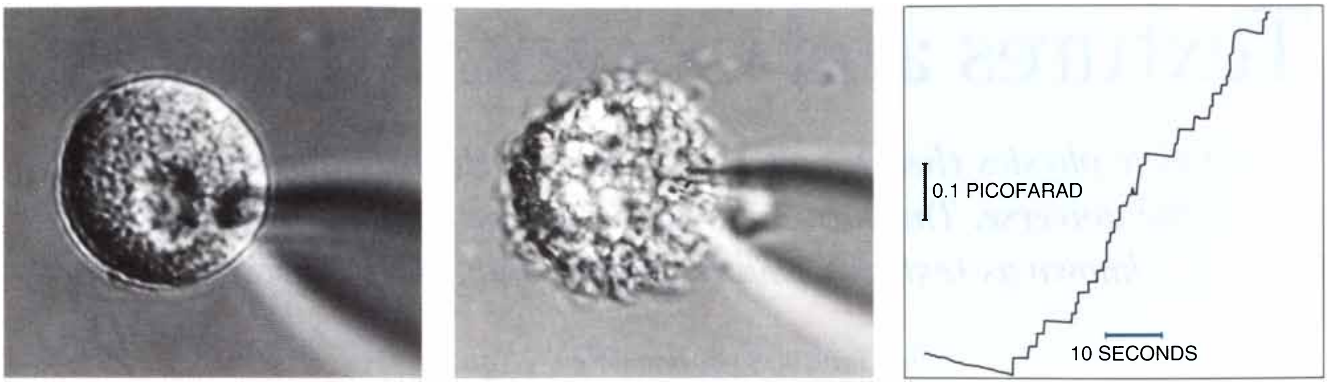
Ideally, for such studies, one would like to measure simultaneously the ion currents, the calcium signal and secretion in a single cell. Combining whole-cell recording with a method of calcium measurement developed by Roger Y. Tsien while at the University of California at Berkeley comes close to this ideal. Tsien's technique uses fura-2, a dye that changes its fluorescence properties when it binds to calcium. By measuring the fluorescence of the dye, one can calculate the concentration of free calcium ions.

Using the whole-cell patch clamp configuration, we could monitor both the ion currents and secretion in the following way. A cell secretes materials through exocytosis, a process in which storage vesicles inside the cell fuse with the plasma membrane and expel their contents. One consequence of the process is that the vesicular membranes are incorporated into the plasma membrane, which thereby increases in surface area. That increase can easily be detected as a change in the electrical capacitance of the membrane, which is proportional to its surface area. Although that change is sometimes transient or small, capacitance measurements are sensitive enough to detect the fusion of individual vesicles with the plasma membrane. In many cell types those signals can be used to correlate secretion with the intracellular calcium concentration and with electrophysiological events.

We studied chromaffin cells from the adrenal gland, which secrete the hormones adrenaline and noradrenaline in times of stress. Our earlier studies had shown that chromaffin cells possess voltage-dependent calcium channels much like those of nerve cells. When the cells are stimulated electrically or with acetylcholine, the calcium channels open and the intracellular concentration of calcium rises, an event



WHOLE-CELL RECORDING, in which a patch pipette takes the place of a conventional microelectrode, has revolutionized the study of mammalian cells, neurons in particular. Researchers can work on small slabs of neural tissue containing neurons that are still in contact with their neighbors. The technique is sufficiently sensitive to detect excitatory and inhibitory currents induced in the neurons by individual quanta (packets) of transmitters contained in single presynaptic vesicles (*inset*). The current traces were recorded by Peter Stern of the Max Planck Institute for Medical Research in Heidelberg.



MAST CELL (*left*) secretes histamine and other compounds in response to a stimulus delivered by a patch pipette (*center*). Using the patch clamp technique, researchers can measure the membrane capacitance changes that correspond to the fusion of just one secretory vesicle with the plasma membrane. Each fusion event appears as one step in the capacitance recording (*right*). Photographs courtesy of Julio M. Fernandes. Capacitance trace courtesy of W. Almers.

that can be measured readily by fura-2 fluorescence. The capacitance of the plasma membrane also increases, signaling exocytosis.

If a calcium chelator—a compound that binds with calcium ions—is present in the pipette, both the rise in the intracellular calcium concentration and the capacitance signal are suppressed in chromaffin cells. If the pipette contains a high level of calcium, then increases in the intracellular calcium and in the capacitance are induced. These results all suggest that secretion by chromaffin cells obeys the same principles that Katz and Miledi found for the release of acetylcholine at the neuromuscular junction.

When we tried to extend these studies to nonexcitable secretory cells, however, the relation between calcium and secretion was surprising. Our favorite cell type in this category is the mast cell found in connective tissues. When stimulated by hormones or antigens (molecules of foreign organisms), mast cells secrete the histamine and other compounds that mediate inflammatory and immune responses. Mast cells were useful for our studies because they are densely packed with large secretory granules, or vesicles. After stimulation, exocytosis in these cells is vigorous: their surface area typically increases twofold to fourfold within 10 to 20 seconds.

During degranulation, the exocytotic release of the storage granules, the capacitance of a mast cell increases in a staircaselike sequence, in which each step represents a single fusion event. The steps are not uniform, however: their amplitude is scattered around a mean value. The variation reflects the distribution of granule sizes.

To our surprise, raising the concentration of calcium inside mast cells did not induce them to secrete, as it did excitable cells. For some time, we believed this result was caused by the washout of essential molecules. Later we learned that the cells required another stimulus: certain hormones and antigens, when applied extracellularly, could cause vigorous secretion either alone or in synergy with an intracellular calcium signal.

All the stimuli invariably led to a prominent calcium signal—a momentary increase in the intracellular calcium concentration by more than an order of magnitude—after which secretion occurred. The transient calcium peak was largely caused by the release of ions stored intracellularly. There was also a prominent influx of external calcium ions through voltage-independent pathways, but it contributed little to the secretory response. Even if the cell's calcium concentration was fixed at low levels by including a chelating mixture in the pipette, degranulation proceeded to the full extent after external stimulation.

Nevertheless, the calcium concentration does influence mast cells. At very high concentrations, it induces secretion after some delay. This occurrence is probably the basis for many reports that calcium-induced secretion has been obtained with injections of calcium or with molecules that permit calcium to flow into cells. Also, after a hormonal or other stimulus, the cell seems to develop greater calcium sensitivity: subsequently raising the calcium level accelerates ongoing degranulations. In these cells, therefore, calcium is probably just one of several regulators of secretion.

All the recent findings indicate that the secretory process is under the control of a network of interacting signal pathways much like those that modulate ion channels. The components of this network seem to be the same: most prominently, calcium (particularly in neurons and some other cell types), second messengers, kinases and probably G proteins. The interactions that regulate secretion are still just emerging, however, whereas the outline of the network controlling ion channels has more or less been identified—largely by the patch clamp technique.

In conclusion, the extreme sensitivity of the patch clamp has revealed molecular details of how channels function. The technique has also enabled us to study the minuscule cells of mammalian tissue and to trace their signaling pathways by controlling the intracellular environment. Although it is powerful, the technique is also very simple. As more researchers adopt it, we hope and expect that it will continue to help unlock cellular secrets.

FURTHER READING

IMPROVED PATCH-CLAMP TECHNIQUES FOR HIGH-RESOLUTION CURRENT RECORDING FROM CELLS AND CELL-FREE MEMBRANE PATCHES. O. P. Hamill, A. Marty, E. Neher, B. Sakmann, F. J. Sigworth in *Pflügers Archiv: European Journal of Physiology,* Vol. 391, No. 2, pages 85–100; August 1981.

THE PATCH-CLAMP TECHNIQUE IN THE STUDY OF SECRETION. Reinhold Penner and Erwin Neher in *Trends in Neurosciences,* Vol. 12, No. 4, pages 159–163; April 1989.

A THIN SLICE PREPARATION FOR PATCH CLAMP RECORDINGS FROM NEURONES OF THE MAMMALIAN CENTRAL NERVOUS SYSTEM. F. A. Edwards, A. Konnerth, B. Sakmann, T. Takahashi in *Pflügers Archiv: European Journal of Physiology,* Vol. 414, No. 5, pages 600–612; September (I) 1989.

THE STRUCTURE OF ION CHANNELS IN MEMBRANES OF EXCITABLE CELLS. Nigel Unwin in *Neuron,* Vol. 3, No. 6, pages 665–676; December 1989.

STRUCTURE-FUNCTION STUDIES OF VOLTAGE-GATED ION CHANNELS. W. Stühmer in *Annual Review of Biophysics and Biophysical Chemistry,* Vol. 20, pages 65–78; 1991.

Textures and Cosmic Structure

Particle physics theories offer new clues to the origin of structure in the universe. The authors propose that primordial field defects known as textures seeded galaxies and galaxy clusters

by David N. Spergel and Neil G. Turok

Understanding the form and origin of structure in the universe is one of the grandest and most exciting challenges facing science today. New surveys reveal that galaxies lie in huge bubbles and sheets stretching hundreds of millions of light-years across. Astronomers have recently discovered extremely old, distant galaxies and quasars. Precision measurements by the *Cosmic Background Explorer* (*COBE*) satellite confirm that the sky is suffused with a remarkably even microwave glow, thought to date from the early days of the universe. To comprehend this array of phenomena, cosmologists have drawn on ideas derived from particle physics. The most popular such idea, called the inflationary universe scenario, successfully describes many aspects of cosmic structure. But inflation cannot fully account for the very large scale clumping of galaxies or for the existence of the oldest quasars.

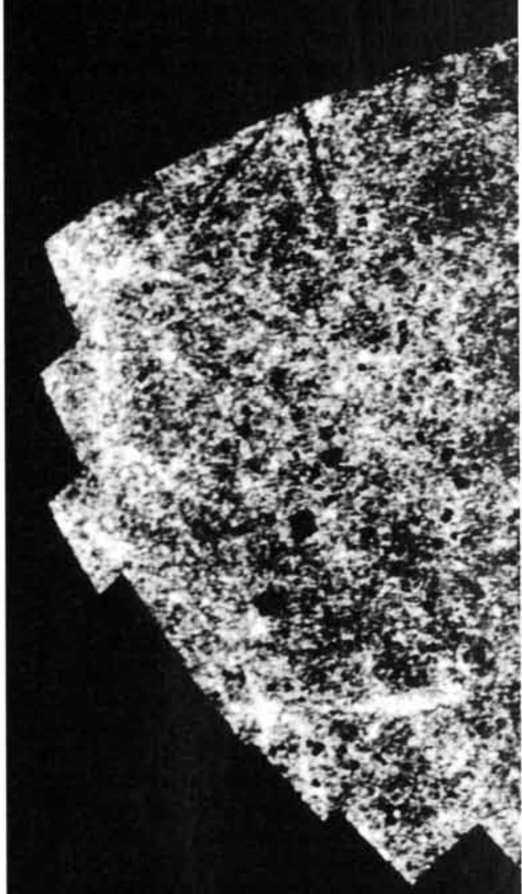
We have investigated an alternative idea that seems to match the explanatory triumphs of inflation while rectifying its major failings. Our hypothesis is built around the idea of symmetry breaking, a process central to unified theories of particle physics. Symmetry breaking leads to the formation of cosmic defects called textures, which would have appeared soon after the birth of the universe. Textures could have seeded aggregations of material that evolved into galaxies and galaxy clusters. The texture hypothesis yields a number of predictions that will soon be put to observational test.

Trying to understand events that occurred during the very first moments after the birth of the cosmos sounds like monumental hubris. Yet the big bang theory, the conceptual foundation of modern cosmology, has grown increasingly certain in recent years. It provides a simple and highly successful description of the physical conditions of the universe going almost all the way back to the instant of its origin.

In the big bang theory the universe—including all matter and all space—began as an infinitesimally small point some 10 to 20 billion years ago and has been expanding ever since. The primary evidence in favor of the big bang is the present observed expansion of the universe. Remote galaxies consistently appear redder than nearby ones. This pattern implies that the light they emit has been stretched (reddened) as galaxies move away from one another. Tracing this expansion back in time, one finds that the universe began in an intensely hot, dense state. Today the universe is flooded with a pervasive background of microwave radiation left over from these early times. *COBE* recently found that the energy spectrum of these microwaves perfectly matches that predicted by the big bang theory, dramatically confirming that the early universe was extremely hot and dense.

Most remarkably, the big bang theory accurately predicts the relative abundances of the light elements (hydrogen, deuterium, helium and lithium) in the modern universe. Based on those predictions, cosmologists deduced that there could be no more than four families of elementary particles; recent experiments have confirmed that only three families exist.

These impressive achievements have persuaded most physicists and astronomers that, to a first approximation at least, the big bang theory is probably correct. But the question of the origin of structure remains unsolved. The simplest explanation is that some event



CLUMPY UNIVERSE puzzles theorists who think matter was very smoothly distributed in the early cosmos. This map of two million galaxies (*white dots*) covers 10 percent of the sky and extends two billion light-years deep.

DAVID N. SPERGEL and NEIL G. TUROK have collaborated for the past two years on the texture theory of the origin of cosmic structure. Spergel received a Ph.D. in astronomy from Harvard University in 1985. In 1987 he joined the staff of Princeton University, where he is now assistant professor of astrophysical sciences. He has concentrated his research on galactic structure and dark matter. Turok earned his Ph.D. in theoretical particle physics from Imperial College, London, in 1983. In 1987 he moved to Princeton University, where he is now an associate professor of physics. Turok's main interests are large-scale structure, the origin of matter and phase transitions in the early universe.

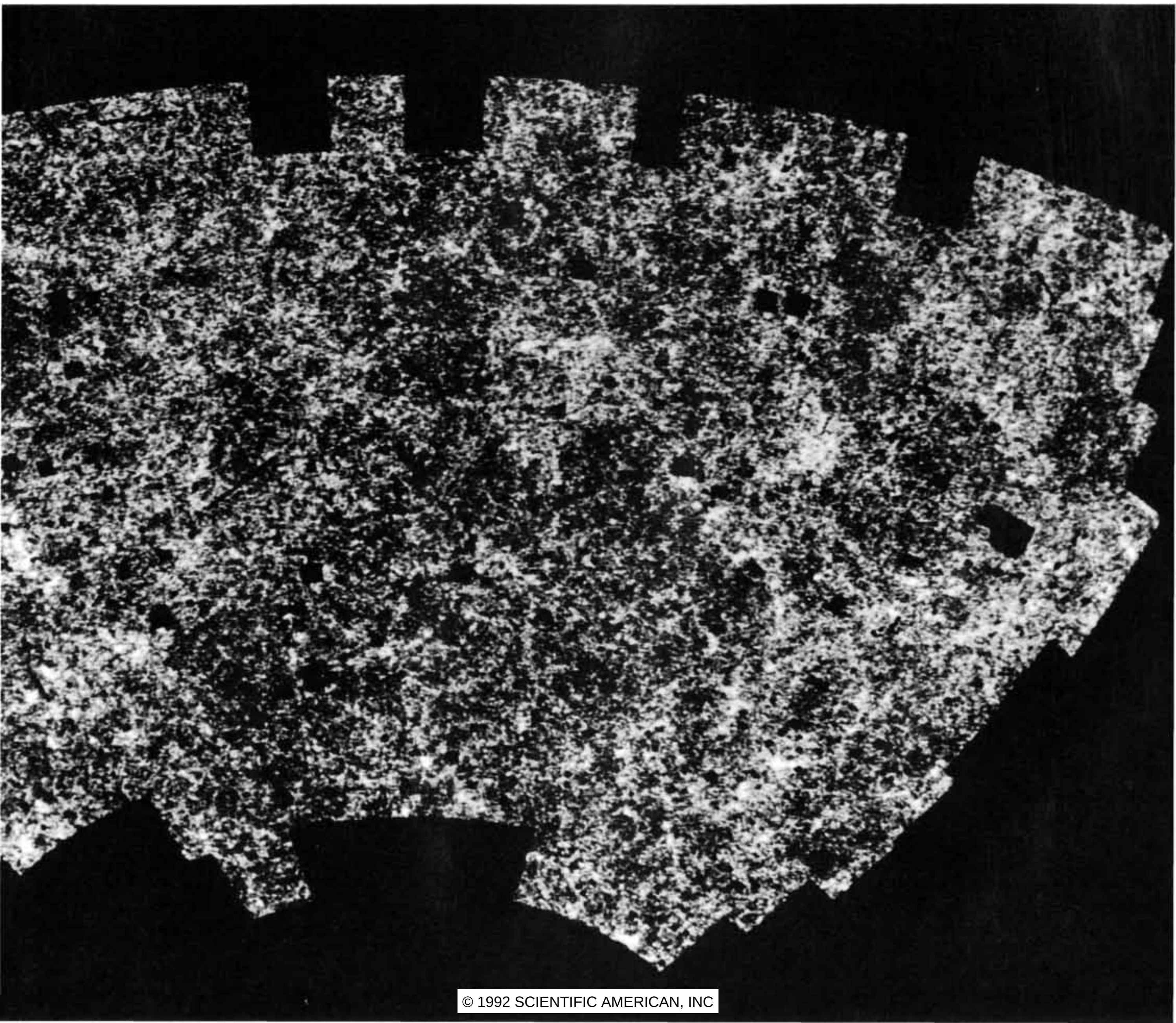
(such as the formation of textures) produced primordial density variations in the early universe. In overly dense regions, gravity more effectively resists the overall expansion of the universe. Eventually, gravity wins over expansion, and such regions collapse. As the universe aged, gravity could have magnified primordial density fluctuations as small as one part in 10,000 into structures such as galaxies, galaxy clusters and superclusters.

Extensive surveys of the large-scale distribution of galaxies show just how uneven the modern universe has become. Galaxies are seen clumping together in sheets and bubbles that surround relatively barren voids. The major problem in interpreting these observations is that the distribution of visible galaxies may not be the same as the overall distribution of matter. There are good reasons to think that most of the mass in the universe exists not in luminous stars or glowing gas clouds but in an invisible form, so-called dark matter.

Observations of the motions of stars and gas in orbit on the outskirts of galaxies indicate that each galaxy is surrounded by a halo of dark matter at least 10 times as massive as the visible component. Likewise, studies of the motions of galaxies in clusters, and of the motions of several thousand galaxies in our vicinity, indicate the presence of a gravitational pull from unseen matter.

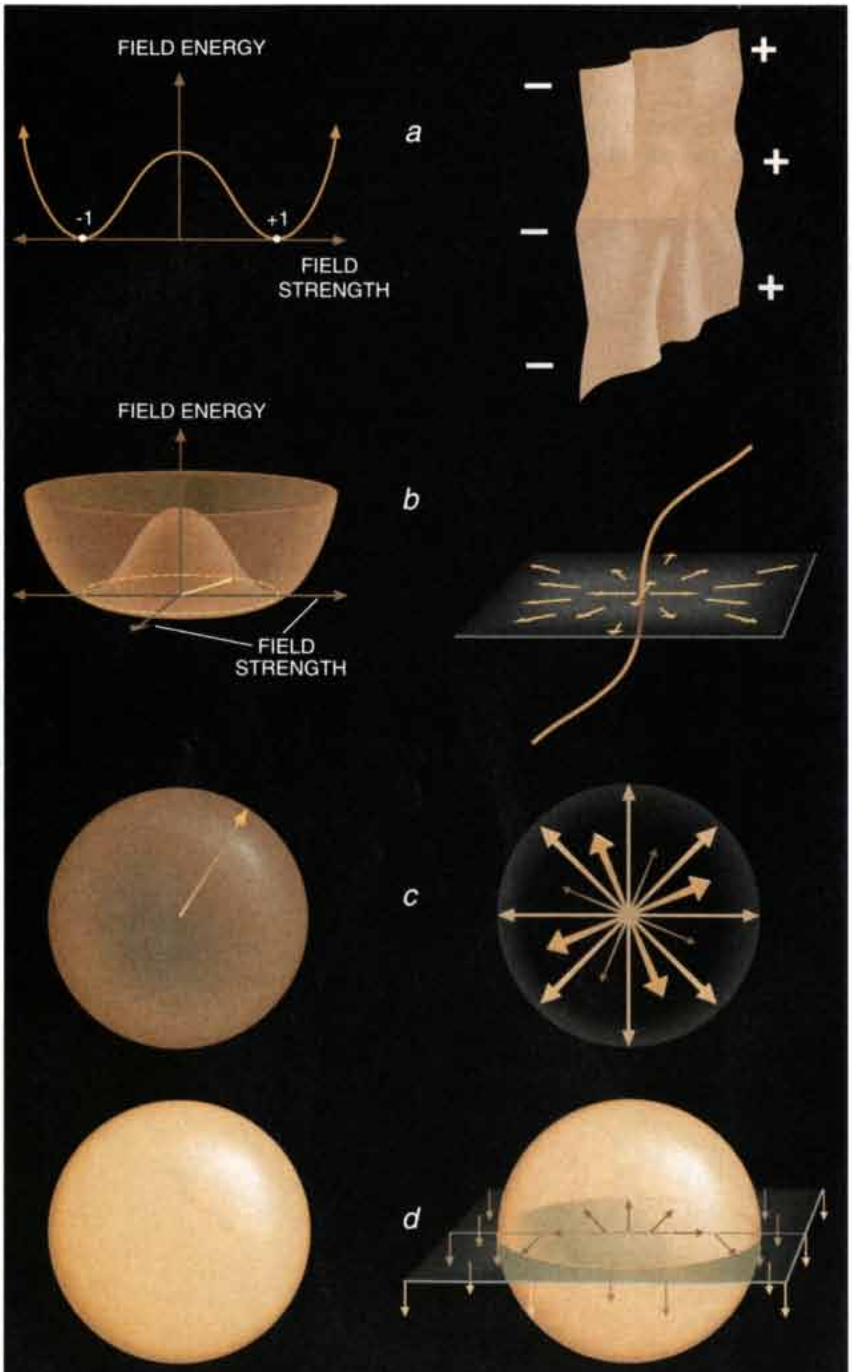
The nature of dark matter remains unknown; it is cosmology's biggest bugbear. Candidates for the unseen mass range from black holes or low-mass stars to various hypothetical subatomic particles. Many cosmologists think dark matter consists of material having a low thermal velocity, commonly called cold dark matter. Such sluggish material could cluster efficiently under the influence of gravity, forming objects ranging in size from individual galaxies to huge galaxy superclusters.

In their quest to understand the formation of galaxies and galaxy clusters, cosmologists have sought to detect evidence of lumpiness in the early universe. One place to look is in the cosmic microwave background. Local variations in the density of the early universe would have caused changes in the apparent temperature of the radiation across the sky. So far efforts to observe these temperature variations have failed. Recent measurements of the microwave sky indicate instead that the early universe was quite smooth. The density of the early universe seems to have been uniform to better than one part in 10,000.



Topological Defects

A hypothetical field called the Higgs field can support four kinds of deformations. Such defects are created during a phase transition, when a Higgs field breaks symmetry. When a discrete symmetry is broken, domain walls (*a*) separate regions within which the Higgs field takes on different values (for example, −1 or +1). Cosmic strings (*b*) form when, in the broken symmetry phase, the Higgs field can point in any direction in a two-dimensional space. As one travels around a cosmic string, the Higgs field rotates through all these values. Monopoles (*c*) appear when the Higgs field can point in any direction in a three-dimensional space. At the center of these defects, the Higgs field strength is zero. Textures (*d*) form when the Higgs field can choose any direction in a four-dimensional space. The illustration below shows a slice through a texture. Note that the arrows showing field orientation around the texture are parallel. As a result, the texture can unwind, permitting the surrounding field to reconnect smoothly.



Theories of the origin of cosmic structure must explain the observed clumping of both dark and visible matter while remaining consistent with the apparent smoothness of the microwave background. One way to account for the smoothness is a hypothetical mechanism called inflation. In the inflationary version of the big bang theory, the very early universe underwent a brief, extraordinarily rapid episode of expansion [see "The Inflationary Universe," by Alan H. Guth and Paul J. Steinhardt; SCIENTIFIC AMERICAN, May 1984]. During inflation, the universe would have become nearly uniform on large scales because local lumps would expand to enormous size. At the same time, microscopic quantum fluctuations would have been stretched and amplified into macroscopic density variations that could initiate the growth of large-scale structure.

Inflation is a beautiful idea, invoking quantum mechanics, whose rules govern the tiniest subatomic particles, to produce the biggest formations in the universe. It even agrees reasonably well with observed intermediate-scale structure, ranging from the size of a galaxy (tens of thousands of light-years) to the mean separation between galaxies (about 30 million light-years).

But the inflation theory has significant drawbacks. The quantum fluctuations typically are too large; consequently, certain parameters in the theory must be tuned to implausibly tiny values to make the rather delicate theoretical edifice hang together. Moreover, observations of very large scale structure increasingly conflict with it. For example, Steven Maddox and his colleagues at the University of Oxford surveyed the angular positions on the sky of more than two million galaxies. They found that galaxies cluster together much more strongly on large scales than the most popular version of the inflation theory predicts.

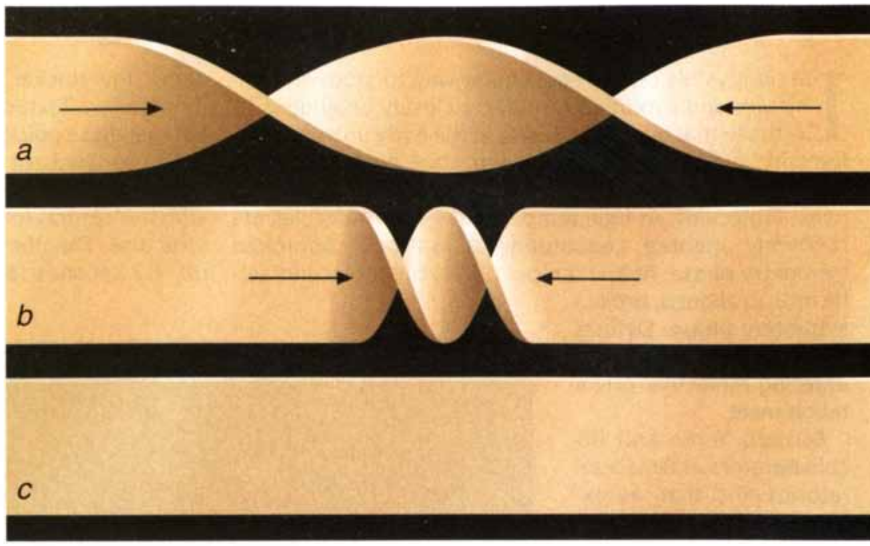
We have pursued a less ambitious approach to understanding the origin of cosmic structure. We began by asking what physical process could have created large-scale structure in an initially smooth, hot universe. Surprisingly, the most promising answer comes from researchers who study not distant galaxies but tiny subatomic particles. Particle physicists suspect that a fundamental process known as symmetry breaking determines the differing masses and charges of the elementary particles of matter and distinguishes the four forces through which they interact. Symmetry breaking in the very young universe could have led to the deviations from smoothness that are needed to explain the present appearance of the cosmos.

Although of considerable interest to cosmologists, symmetry breaking was first investigated by physicists seeking to unify the wide variety of subatomic particles and forces. Everyday matter is composed of electrons, protons and neutrons; protons and neutrons in turn consist of two kinds of quarks. These quarks and the electron and its neutrino constitute the first of three families of elementary particles. Interactions between these particles are mediated by four basic forces: electromagnetism, the weak force (which governs radioactive decay), the strong force (which binds protons and neutrons together in atomic nuclei) and gravity.

This complex assembly of particles and forces shows several striking regularities. The particles possess definite charges, all integral multiples of one third the charge of the electron. And corresponding particles in each family possess identical charges, differing only in mass. Most particle theorists think the repeated patterns in the particle families are manifestations of symmetry laws that are obscured at the low energies of everyday physics. These symmetries, which become evident in experiments conducted at high-energy particle accelerators, would similarly have been exposed in the first moments of the universe.

Abdus Salam of the International Center for Theoretical Physics in Trieste and Steven Weinberg of the University of Texas in Austin proposed that such an underlying symmetry exists between the electromagnetic force and the weak force, the electron and its neutrino, and the two quarks of each family. Numerous experiments in particle accelerators have confirmed to great precision the validity of the unified electroweak theory. Particle theorists strive to repeat that accomplishment by constructing "grand unified" theories that combine the electroweak force with the strong force and link the properties of electrons, neutrinos and quarks.

According to these unified theories, at very high temperatures matter enters unfamiliar so-called unbroken symmetry phases, in which the symmetries between particles and forces fully express themselves. The electroweak theory implies that at temperatures above 10^{15} kelvins, electrons appear identical to neutrinos, and the weak force is indistinguishable from the electromagnetic force. Grand unified theories predict that at temperatures above 10^{28} kelvins, electrons and neutrinos behave identically to quarks, and the electroweak force blends into the strong force. Finally, family symmetry theories hold that the three families of particles grow indistinguishable at a similar temperature.



TWISTED RUBBER BAND illustrates the unwinding and eventual disappearance of a texture. The orientation of the band at each point along its length is analogous to the direction of the Higgs field. A texture defect corresponds to a single, 360-degree twist in the band (*a*). Over time, the texture twist becomes smaller and smaller, stretching the band more and more taut (*b*). Finally, the twist becomes so tiny that the band breaks and reconnects without a twist (*c*). When a texture shrinks to subatomic size, the Higgs field likewise reconnects, eliminating the defect.

At birth, the universe was exceedingly hot and existed in a state of perfect unbroken symmetry. But as it expanded the universe cooled, passing through a series of abrupt changes in its physical state, called phase transitions. In each phase transition, some of the symmetry relations between forces and particles broke. Thus, the universe degenerated from a phase containing unified forces and identical particles to the modern, more familiar state of differing forces and particles.

Theory holds that the electroweak phase transition occurred when the universe was 10^{-12} second old and only 10^{-15} its present size. The grand unified transition would have occurred when the cosmos had attained the ripe old age of 10^{-36} second, at which time the present observable universe was no bigger than a grapefruit. The breaking of symmetries (such as family symmetry) at this very early time could have produced defects such as textures.

These symmetry-breaking phase transitions may sound abstruse, but analogous transitions take place in many familiar situations. The sudden changes in the physical properties of water when it freezes into ice or boils into steam are examples of phase transitions, and just as in the particle theories discussed above, such phase transitions are often linked to a change in symmetry properties. Liquid water is highly symmetric: it looks the same from any angle. Ice is less symmetric than water because it consists of a definite crystalline structure that appears the same only if shifted by certain discrete distances or rotated through certain discrete angles.

When water freezes suddenly, it does not form a perfect crystal; it is riddled with imperfections. These defects appear because as the water begins to freeze, tiny crystals of ice start growing throughout it. Each crystal assumes a random orientation as it forms. The crystals grow until they meet one another and fill all available space. Defects in the crystal structure appear where regions of different orientations meet. The quicker the liquid cools, the more defects form, because the ice crystals have less time to grow before they encounter one another. A similar process would have resulted as the newborn universe rapidly cooled and underwent symmetry-breaking phase transitions.

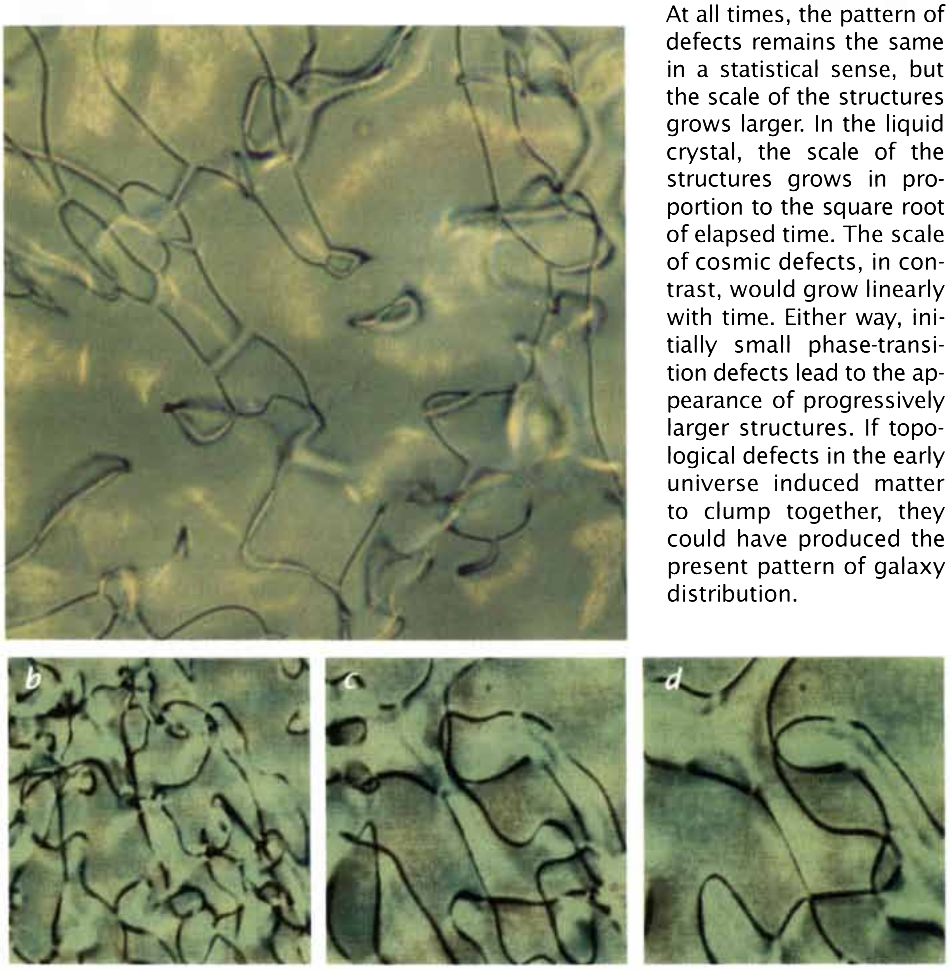
Phase transitions that occur in particle physics theories differ from those that occur in water in two important ways. First, defects in ice are held firmly in place by the crystalline structure, whereas the defects predicted in theories of particle physics can move freely. As soon as these defects form, they begin to untangle and straighten themselves in order to minimize their energy. This phenomenon of defect-ordering dynamics has been studied in the laboratory in materials such as liquid crystals [*see box on next page*].

Liquid Crystals

Liquid crystals offer a convenient way to study the formation and evolution of defects closely analogous to those that may have arisen in the early universe. Defects in liquid crystals show up as dark lines and loops (*top*). The liquid crystals shown here consist of rodlike organic molecules. At high temperatures, the molecules are randomly oriented, constituting an isotropic, unbroken symmetry phase. At low temperatures, the molecules settle into an aligned, broken symmetry phase. Defects appear where regions of differing molecular orientation meet.

Bernard Yurke and his collaborators at Bell Laboratories find that, as expected, rapid phase transitions in liquid crystals result in a dense tangle of defects. Remarkably, this simple substance can support three of the four possible kinds of defects. Other laboratory systems, such as superconductors or binary alloys, form only one kind of defect (strings or domain walls, respectively). The thin, dark lines seen here are analogous to cosmic strings. Pinches along the thicker, more diffuse strings correspond to monopoles. Textures appear as doubled loops of diffuse strings; these quickly decay into a pair of monopoles.

The subsequent evolution of defects in the liquid crystal (*bottom row*) beautifully illustrates the idea of scaling, which is central to defect theories of the origin of cosmic structure. The photographs show the crystal one second (*a*), 1.7 seconds (*b*), 2.9 seconds (*c*) and 4.8 seconds (*d*) after the phase transition. At all times, the pattern of defects remains the same in a statistical sense, but the scale of the structures grows larger. In the liquid crystal, the scale of the structures grows in proportion to the square root of elapsed time. The scale of cosmic defects, in contrast, would grow linearly with time. Either way, initially small phase-transition defects lead to the appearance of progressively larger structures. If topological defects in the early universe induced matter to clump together, they could have produced the present pattern of galaxy distribution.



Second, the phase of water is defined by the arrangement of its constituent molecules. In the realm of subatomic physics the phase of the various particles and forces is determined by the state of the fields that link them. Every particle and force in the standard model of physics is associated with a corresponding field. In addition to the well-known electric and magnetic fields, there is the electron field, the proton field and so on.

Particle theorists have proposed hypothetical fields, called Higgs fields, for the specific purpose of breaking the symmetry between forces and particles. Higgs fields follow equations of motion similar to those that govern electric and magnetic fields. Just like those fields, the Higgs fields tend to minimize their energy by assuming a uniform strength throughout space.

Unlike other fields, however, a Higgs field does not drop to zero in its lowest energy state. Instead it takes on a particular, nonzero strength throughout the universe. In so doing, the Higgs field breaks the symmetry between the forces and the particles that couple to it. Particle physicists have employed this so-called Higgs mechanism in the electroweak, grand unified, family unified and supersymmetric theories. In each case, extra Higgs fields are introduced to break the symmetry. For example, when temperatures fell below the electroweak phase transition, the electroweak Higgs field turned on and broke the symmetry between the electron and the neutrino and between the weak and electromagnetic forces.

Although there is convincing evidence that some form of symmetry breaking occurs, Higgs fields remain entirely hypothetical. Theorists introduce them solely to break symmetry, and the only indication of their existence is the fact that symmetries are broken.

One of the main goals for the next generation of particle accelerators is to search for the Higgs particle, which should be associated with the electroweak Higgs field. The discovery of the Higgs particle would be one of the most important breakthroughs in the history of physics. Physicists are particularly excited by this search because in electroweak theory the Higgs mechanism is responsible for giving all known particles their masses.

The Higgs mechanism is not the only possible process that could break the symmetry between forces and particles; it is merely the simplest and best understood. Many of the consequences of symmetry breaking discussed here hold true in general and do not depend on the details of the Higgs mechanism.

The ice analogy provides a conceptual understanding of how the Higgs field might produce textures and, through them, cosmic structure. Just as ice consists of crystalline planes having definite orientations, so the Higgs field can be pictured as taking on a certain direction. A magnetic field breaks the symmetry of space by telling a compass which way to point, thereby distinguishing one direction from all others. The Higgs field indicates a direction in a comparable but more abstract sense. For example, the direction of the electroweak Higgs field breaks the symmetry of electron/neutrino "space." The "direction" in which it points acquires mass and becomes the electron.

In 1976 Thomas W. B. Kibble of Imperial College, London, elaborated the concept. He pointed out that symmetry-breaking phase transitions early after the big bang could have caused the Higgs field to take on various directions in different regions of the universe as do different crystalline regions in a block of ice. Kibble noted that in many unified theories, defects in the Higgs field would form where regions of conflicting field orientation meet.

Such regions cannot settle into a uniform configuration. Instead a local, intense concentration of energy, known as a topological defect, forms. Such defects caught the attention of cosmologists because they could store huge amounts of energy, in essence retaining the conditions of the early universe. Because of the equivalence of energy and mass, these defects produce gravitational effects that would cause matter to clump together. Theorists therefore have exerted considerable effort studying topological defects and trying to learn whether they could explain the present structure of the universe.

Present particle physics theories allow for the existence of four different kinds of topological defects. Each has been subjected to close scrutiny to determine whether it could be the one that produced the cosmic structure.

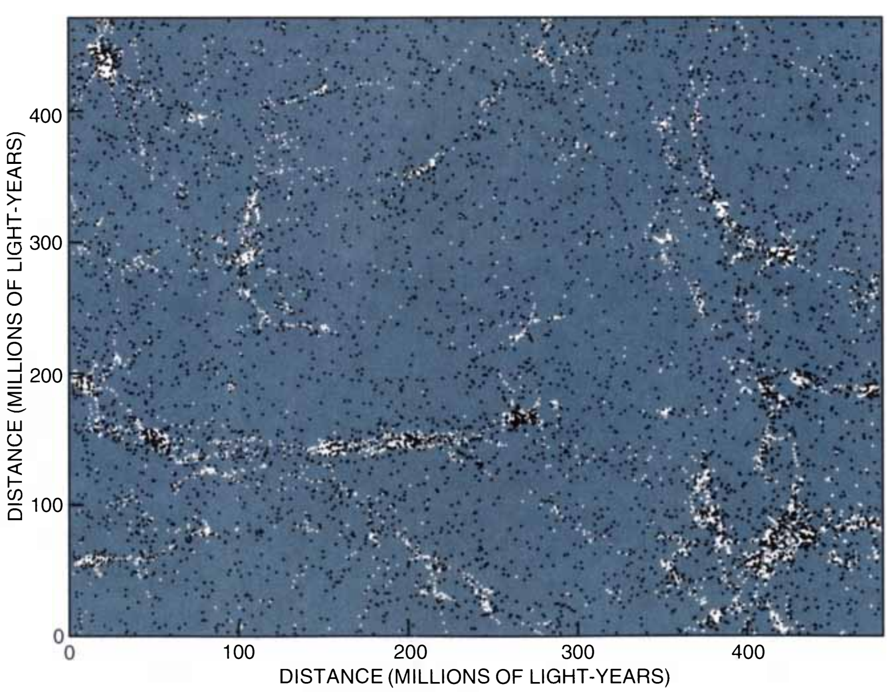
The simplest kind of topological defects, called domain walls, arises when a discrete symmetry (a mirror reflection, for example) breaks. Regions in space where the Higgs field takes on different values would be separated by surfaces at which it changes from one value to another. These domain wall surfaces cannot have edges, and so they must form closed bubbles or infinite sheets. Unfortunately, domain walls do a poor job of seeding cosmic structure. They store so much energy that if they formed soon after the big bang, they would have quickly dominated the mass of the universe, leaving little material from which to form objects such as stars, planets and living beings.

Continuous symmetries play a more important role in both cosmological and unified particle theories than do discrete symmetries. In this case, the Higgs field can point in a continuous set of directions. One kind of continuous symmetry defect has a linelike form. To understand how such a defect might arise, imagine that you are standing in the middle of a large crowd. As you turn around, the crowd looks similar in all directions; in other words, the directional symmetry is unbroken. The crowd is now told to lie down. Different people choose to lie in different directions. At first there is a lot of confusion, but gradually people adjust themselves to lie in the same direction as their neighbors. Now you find that people around you are lying in a definite direction. Directional symmetry has been broken.

If everyone lies down pointing directly away from you, you are forced to remain standing. You represent a defect in the broken symmetry phase. A defect of this kind, in which the Higgs field points radially outward, is called a cosmic string. Many grand unified theories predict the existence of cosmic strings. Cosmic strings would form lines through space that have no ends. They assume the shape of closed loops or infinite lines that wander across the universe; similar features are observed in liquid crystals.

In 1979 the Soviet physicist Yakov B. Zel'dovich suggested that cosmic strings offered a plausible mechanism for the creation of cosmic structure. During the following decade, theorists devoted considerable work to understanding the properties of cosmic strings. But Andreas Albrecht and Albert Stebbins of Fermilab National Accelerator Laboratory have recently shown that cosmic strings and cold dark matter would produce structure only on a scale much smaller than the broad clumping documented in recent galaxy surveys.

If linelike defects cannot do the job, what about pointlike ones? Such topological defects arise when the local directions of the Higgs field point radially outward in three dimensions from



SIMULATED STRUCTURE in a cosmos perturbed by textures broadly resembles the observed universe. As textures unwind, their energy density increases, and their gravitational perturbations intensify. Here visible matter (*white*) clumps far more tightly than does dark matter (*black*). If this finding holds for the real universe, astronomers may have significantly underestimated the amount of dark matter.

a single point. These defects, known as monopoles, resemble the outward-pointing quills of a frightened porcupine. One kind of monopole, which arises in many grand unified theories, presents a problem because, like domain walls, it tends to dominate all other matter in the universe. Physicists are still actively investigating other kinds of monopoles whose properties more closely resemble those of textures.

In 1989 one of us (Turok) recognized that many unified theories produce another variety of topological defect, called texture, which could initiate the formation of large-scale structure while avoiding the obvious inadequacies of the other kinds of defects. Texture is more difficult to visualize than the other defects because it results from a Higgs field pointing in four dimensions. Obviously, such an arrangement cannot be drawn on paper, but it is possible to depict a two-dimensional slice through a texture [*see illustration on page 54*].

Texture would be much more diffuse than walls, strings and monopoles. A texture is not localized at a point or on a line or a surface. Textures have no preferred size; they could be millimeters or light-years across. Once formed, a texture evolves by shrinking at the speed of light down to a subatomic size, at which point the Higgs field unwinds and reconnects in a smooth configuration. Along with William H. Press and Barbara S. Ryden of Harvard University, we have conducted numerical computer simulations of texture formation and evolution in the early universe. These simulations show that textures exhibit a simple scaling behavior like that shown by the string defects in liquid crystals. In both cases, the defects form a pattern that always looks the same (in a statistical sense) but in which the scale changes over time. For cosmological texture, the scale grows at the speed of light. This behavior seems to be just what is needed to produce the form of the observed cosmic structure.

The phase transition that creates textures takes place at the grand unification temperature, very soon after the big bang. At that time, textures of all sizes form. The smallest ones unwind first, then larger and larger ones. One hour after the big bang, textures that were initially one light-hour across begin unwinding; a year later textures initially one light-year across unwind and so on. In the universe today, there would be textures billions of light-years across still in the process of shrinking.

As a texture shrinks, its energy concentrates into a steadily shrinking region. This concentration produces a gravitational field that attracts nearby matter. Each texture leaves behind a bump in the mass distribution of the universe. Progressively larger bumps form as larger textures shrink down and unwind. When the universe becomes dominated by matter, around 10,000 years after the big bang, these bumps begin to grow under their own gravity. The broad range of sizes of textures leads to a similar variety of scales of gravitationally bound objects, from star clusters to galaxies to galaxy clusters to giant superclusters.

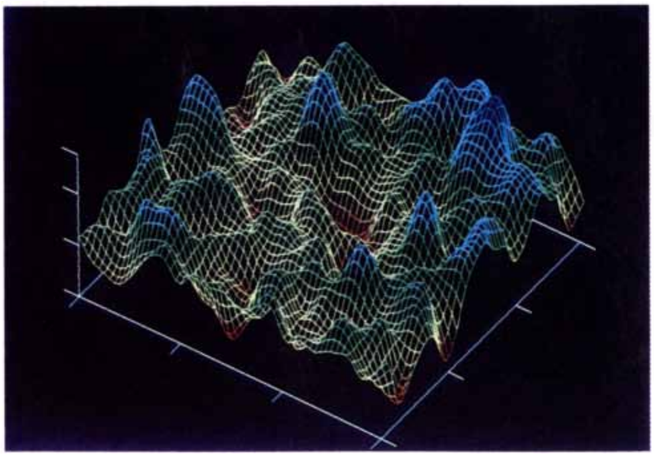
The energy density of the texture field, which determines how strongly it attracts the surrounding matter, is fixed by the strength of the Higgs field. The field strength in turn is set by the energy scale involved in the symmetry-breaking process in which the texture formed. Because there is, as yet, no uniquely compelling theory for producing texture, we have simply adjusted the strength of the Higgs field to get the best fit with the observed structure of the universe. When we do so, we find our theory predicts that symmetry breaking should happen at a temperature close to the grand unified temperature, an intriguing agreement.

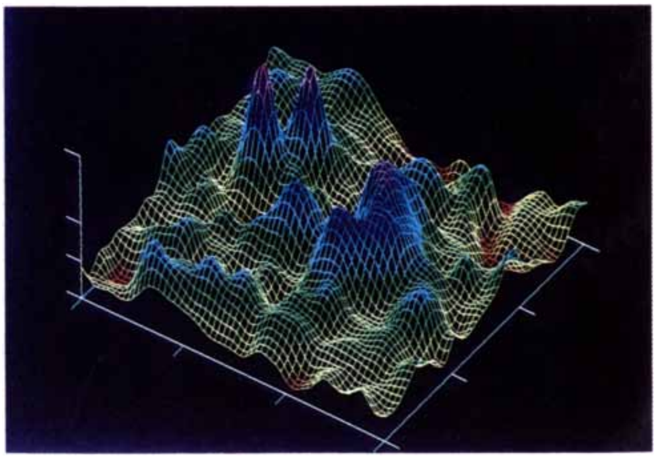
During the past decade, cosmologists have developed elaborate computer models to study gravitational clustering in an expanding universe. The newest models also take into account hydrodynamic effects such as shock formation, radiation and cooling. In part because of the rapid advances in computer hardware and software, during the past two years we have been able to develop our vague texture hypothesis into a highly detailed, predictive theory of cosmic structure formation.

The texture-perturbed model that best fits the observational data contains 95 percent of the mass in the form of cold dark matter; the remainder is ordinary matter.

One of the great successes of the inflation plus cold dark matter theory was its quantitative agreement with the observed clustering of galaxies on scales smaller than the mean separation of galaxies. In collaboration with Changbom Park of the California Institute of Technology, we performed detailed numerical computer calculations of gravitational clustering produced by textures. We found that the texture plus cold dark matter model also yields good agreement with the observed galaxy clustering.

In hindsight, the rough agreement of the two theories' predictions on scales smaller than the mean separation between galaxies is not surprising. At those distances, gravity dominates the

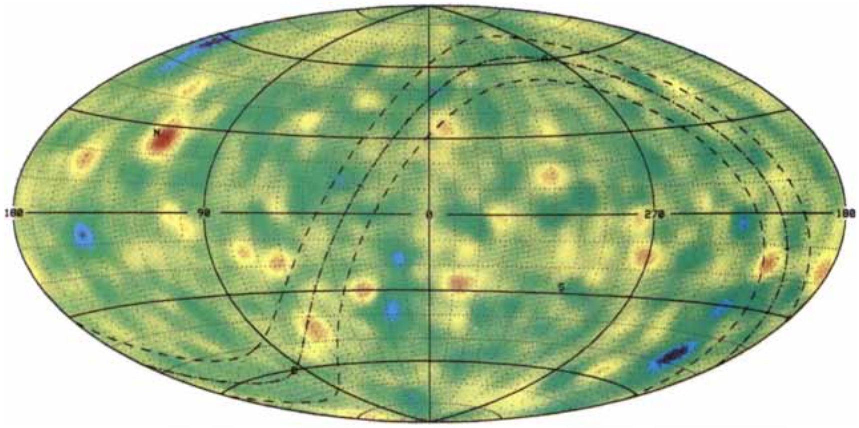




DENSITY FLUCTUATIONS can follow random (Gaussian) or nonrandom distributions. The inflation theory predicts random density fluctuations (*left*). Texture models predict more intense peaks and weaker dips in the density (*right*). In either scenario, gravity eventually causes dense regions to collapse into galaxies. Texture models produce some particularly dense regions that would collapse rapidly, explaining recent observations of extremely old quasars and galaxies.

clustering pattern, washing out the differences between competing theories. It is clear that more detailed measures of structure on large scales, where gravity has not appreciably altered the primordial clustering pattern, are required to test decisively the correctness of structure-formation models.

On these larger scales, the angular correlation function of galaxies (that is, their apparent clustering on the sky) is the best available statistic. As previously noted, the observed correlation conflicts with that predicted by inflation at large angular scales. Simulations we conducted with Park, combined with analytic work performed with Andrew Gooding and J. Richard Gott III of Princeton University, show that the angular correlation function predicted by the texture theory fits much more closely with the observations.



MICROWAVE BACKGROUND SIMULATION shows "hot" (*blue*) and "cold" (*red*) regions in the microwave sky, predicted by the texture model. These variations may be revealed in data from the *COBE* satellite that are now being analyzed.

There is another important difference between the inflation and texture theories. In the inflationary model, density fluctuations follow a random (Gaussian) distribution in the early universe. Texture models, however, produce density fluctuations in the form of strongly localized bumps that have very different statistical properties. In particular, high-density peaks are much more numerous in texture theory than in inflation.

Because of these plentiful dense regions, galaxy formation begins much earlier in the texture theory than in inflationary scenarios. In the latter, galaxy-size objects do not begin to collapse until the universe is a few billion years old. But in a texture-perturbed universe, galaxy formation begins after only 10 million years. This prediction is particularly significant in light of the recent discovery of several quasars whose light is highly redshifted, implying that they were shining when the universe was only about a billion years old. The existence of these quasars suggests that some gravitationally bound objects formed quickly after the big bang.

Over large distances, the intense density fluctuations produced by textures would lead to the appearance of more dense clusters and superclusters than inflation would predict. Gooding and Jim Bartlett of the University of California at Berkeley have found that clusters of galaxies form earlier and are denser in the texture theory than in inflation. Measurements of the densities of nearby clusters will constitute an important test of the texture model. The recent large-scale survey led by George Efstathiou of the University of Oxford reveals the existence of a number of galaxy superclusters hundreds of millions of light-years wide. These findings agree with the predictions of the texture theory, but they conflict with inflation.

We have worked with Renyue Cen and Jeremiah P. Ostriker of Princeton to run detailed computer models of how galaxies form around textures. We find that the texture plus cold dark matter theory leads to the development of collapsed objects whose properties closely resemble those of real galaxies. For example, the predicted number of galaxies of any given mass matches observations to within the computational uncertainty of the model.

Once we have identified where galaxies form in the computer simulations, we examine their overall distribution. Interestingly, galaxies cluster together far more tightly than does dark matter. This finding has important implications. Astronomers have traditionally assumed that galaxies follow essentially the same distribution as dark matter. Based on this assumption, measurements of the masses of clusters have consistently led to estimates that the density of the universe is about one tenth the critical density necessary eventually to halt its expansion.

But if galaxies clump together more closely than does dark matter, these determinations would have measured only a fraction of the dark matter and would have seriously underestimated the density of the universe. Texture theory suggests that the remaining dark matter, representing most of the mass of the universe, is more smoothly distributed. Our simulations show huge sheets of galaxies surrounding regions almost devoid of galaxies, similar in size and form to the "great wall" and voids found in the large galaxy survey conducted by Margaret J. Geller and John P. Huchra of the Harvard-Smithsonian Center for Astrophysics. We see no structures larger than about 300 million light-years, a prediction that will be tested by several teams now mapping the very large scale distribution of galaxies and quasars.

Finally, as textures unwind, they should leave a distinctive signature on the microwave background sky. Photons encountering a shrinking texture on their journey to the earth may either lose energy by climbing out of a texture as it shrinks or gain energy by falling into a texture after it unwinds. Textures should lead to a distinctive pattern of about 10 "hot" and "cold" spots on the microwave sky, each roughly 10 degrees of arc across and each having a temperature that deviates from the mean by one part in 10,000. Such fluctuations should be detectable in maps of the microwave sky now being made by *COBE*, by balloon-borne detectors and by teams at the South Pole.

We anxiously await their results.

FURTHER READING

THE EARLY UNIVERSE. Edward Kolb and Michael Turner. Addison-Wesley Publishing, 1990.

THE CASE FOR THE RELATIVISTIC HOT BIG BANG COSMOLOGY. P.J.E. Peebles, D. N. Schramm, E. L. Turner and R. G. Kron in *Nature*, Vol. 352, No. 6338, pages 769–776; August 29, 1991.

GLOBAL TEXTURE AS THE ORIGIN OF COSMIC STRUCTURE. Neil Turok in *Proceedings of Nobel Symposium No. 79: The Birth and Early Evolution of Our Universe*. Edited by J. S. Nilsson, B. Gustafsson and B.-S. Skagerstam. World Scientific/Physica Scripta, 1991.

A HYDRODYNAMIC APPROACH TO COSMOLOGY: TEXTURE-SEEDED CDM AND HDM COSMOGONIES. R. Y. Cen, J. P. Ostriker, D. N. Spergel and N. Turok in *Astrophysical Journal*, Vol. 383, No. 1, pages 1–18; December 10, 1991.

MARCEL JOJOLA liked the Saturn SL1 so much, he had his customized for work.

Like a lot of California towns, Bear Valley has budget concerns. And like a lot of Californians, Chief Jojola is a fan of imported cars. He has two Subarus at home, two at the Bear Valley police station, and when they needed a new squad car this year, he just planned on getting another one.

Then someone suggested that perhaps Bear Valley should buy American. Tax dollars, and all. So Marcel, whose most recent experience was all import, wasn't quite sure what to do.

*Base M.S.R.P. including retailer preparation. Tax, license, transportation and options additional. © 1991 Saturn Corporation.

Until the day he ran across a Saturn ad touting the SL1 and its $8995* price tag. Taking along his special police-issue driving gloves, Marcel went for a "law enforcement-style" test drive. (We were pretty relieved to get back to the showroom.)

Anyway, to cut to the chase, the Saturn gave Marcel the performance he was looking for in a patrol car. It's American, which took care of Bear Valley patriotism. And as for the budget folks, they were so happy with the value they were getting with a Saturn, they let Marcel buy two.

SATURN.

A DIFFERENT KIND *of* COMPANY. A DIFFERENT KIND *of* CAR.

If you'd like to know more about Saturn, and our new sedans and coupe, please call us at 1-800-522-5000.

Why Aromatic Compounds Are Stable

Molecules containing benzenelike rings survive intact over geologic time and even persist in the harsh environment of nebulae. Researchers can now mathematically describe the reasons for the exceptional stability

by Jun-ichi Aihara

In 1825 a young laboratory director at the Royal Institution in London was asked to analyze the liquid residue that formed during the production of lamp gas. Considered one of the best chemists of the time, the 33-year-old researcher, Michael Faraday, made a discovery that would mark the beginning of a new branch of chemistry. He isolated from the residue a new hydrocarbon—a molecule consisting of only hydrogen and carbon atoms—and called it bicarburet of hydrogen. We now call the substance benzene, a compound that consists of six hydrogen atoms attached to six carbon atoms. It is the prototype of a class of molecules known as the aromatic compounds.

Such hydrocarbons share a valuable property: remarkable stability. They have a strong tendency to hold their shapes and are much less reactive than many other kinds of molecules. The stability is one of the main reasons that aromatic molecules are so useful. In gasoline their presence enhances the octane rating, and the compounds are the primary components in such household products as paint thinner and mothballs. Aromatic compounds are also essential as reagents and solvents in modern industry, helping to produce synthetic fibers, resins and dyes. Despite this widespread use, the fundamental reasons for the stability have eluded rigorous analysis. Such an understanding has now been developed by me and other workers.

The challenge we faced arose from the fact that straightforward quantum mechanical calculations do not provide a sufficient basis for predicting or accurately calculating the stability of aromatic molecules. Attempts to do so often relied on several assumptions that, while reasonable, had to be invoked. They did not emerge as integral parts of the theory. The ideas also failed to elucidate other common features of aromatics, such as their unusual diamagnetic properties. More important, experimentalists were able to synthesize several kinds of aromatic compounds whose properties could not be explained by the early schemes.

To fill in these gaps, I and other workers have developed the theory of topological resonance energy. This idea combines the subject of topology—the mathematical study of geometric shapes—and the quantum principles of electron orbitals and energy levels. It has made it possible to express stability in numerical and geometric terms. This approach has proved valid for almost all kinds of aromatic compounds, including the fullerenes, the three-dimensional carbon molecules currently under intense scientific scrutiny.

Our theory rests on foundations laid down during the past century, as researchers began to understand how elements reacted and bonded with one another. The stability of aromatic molecules implies that the compounds do not want to react, that nature must be satisfied with the structure. One of the earliest recognized characteristics of aromatic molecules was the arrangement of their carbon atoms in the form of a closed hexagonal ring. The German chemist August Kekulé was the first to recognize that configuration in 1865. According to historical accounts, he realized what the shape of benzene must be after he dreamed of a snake swallowing its own tail.

This ringlike structure is the distinguishing characteristic of aromatic compounds—not, as one might have suspected, their odor. In fact, many "aromatic" compounds have no smell. The term derives from some fragrant compounds that were discovered to contain benzene rings. Chemists subsequently began to classify any similar compound as members of the aromatic family, and the nomenclature became entrenched. To complicate matters further, aromatic molecules now include organic compounds in which some of the carbon atoms have been replaced by nitrogen, sulfur or oxygen atoms.

More than just a source of stability, the benzene ring seems to be a shape nature prefers to form. The ring-shaped hydrocarbons are now known to be quite prevalent. For instance, they are found in such mundane substances as coal and soot. Indeed, many chemical reactions tend to lead ultimately to the formation of the benzene ring. To illustrate, consider 5-methylene-1,3-cyclohexadiene (C_7H_8), a fairly simple compound with a rather complex name (the numbers in the name refer to specific carbon atoms that participate in bind-

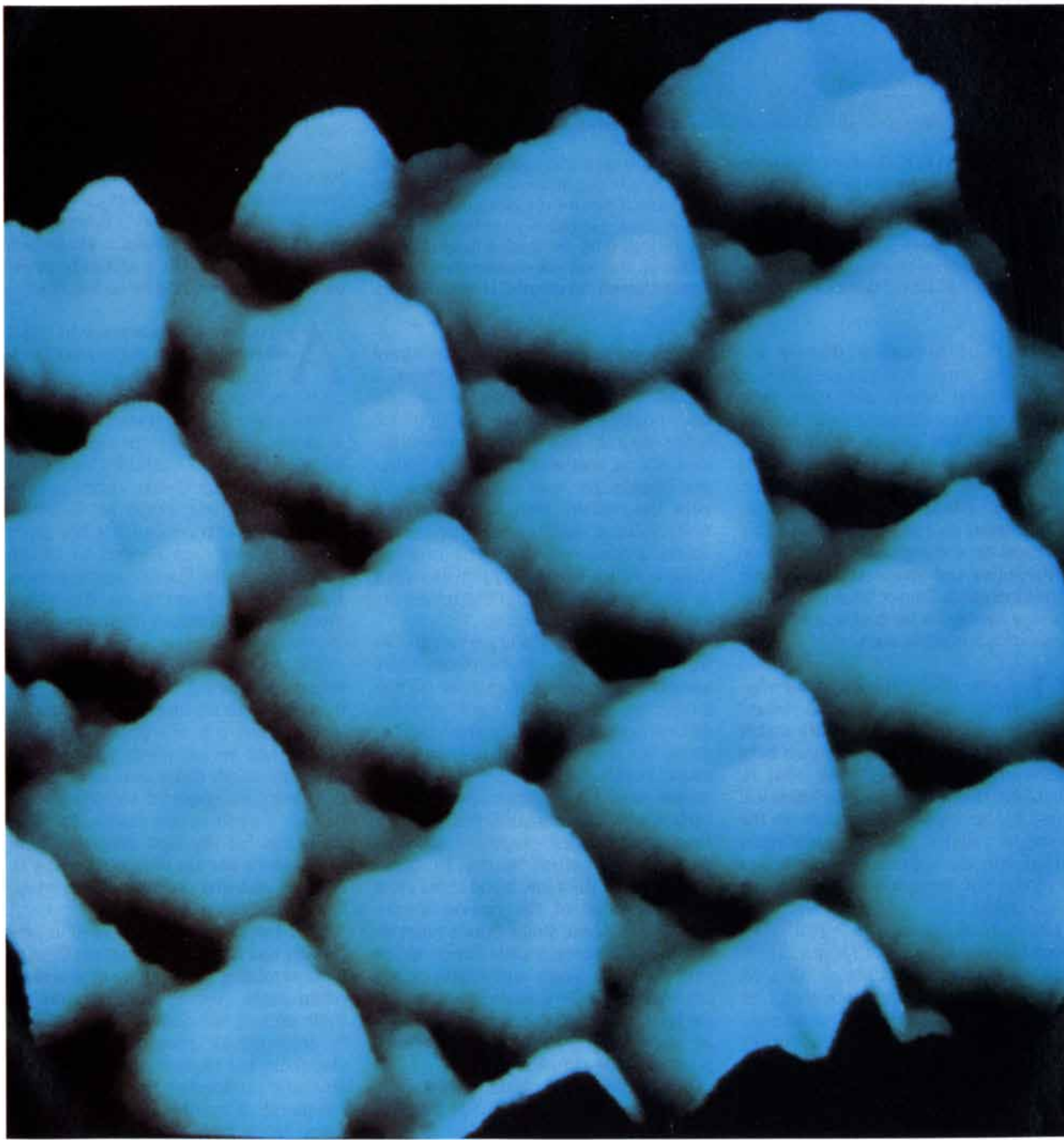
JUN-ICHI AIHARA is professor of chemistry at Shizuoka University in Japan. He received his Ph.D. from the University of Tokyo in 1970 and was a member of the chemistry department of Hokkaido University in Sapporo until 1981. His main research interests include the electronic structures of organic molecules, mechanisms of interstellar dust formation and the origin of life on the earth. This article has been adapted from a story that appeared in *Saiensu*, the Japanese edition of *Scientific American*.

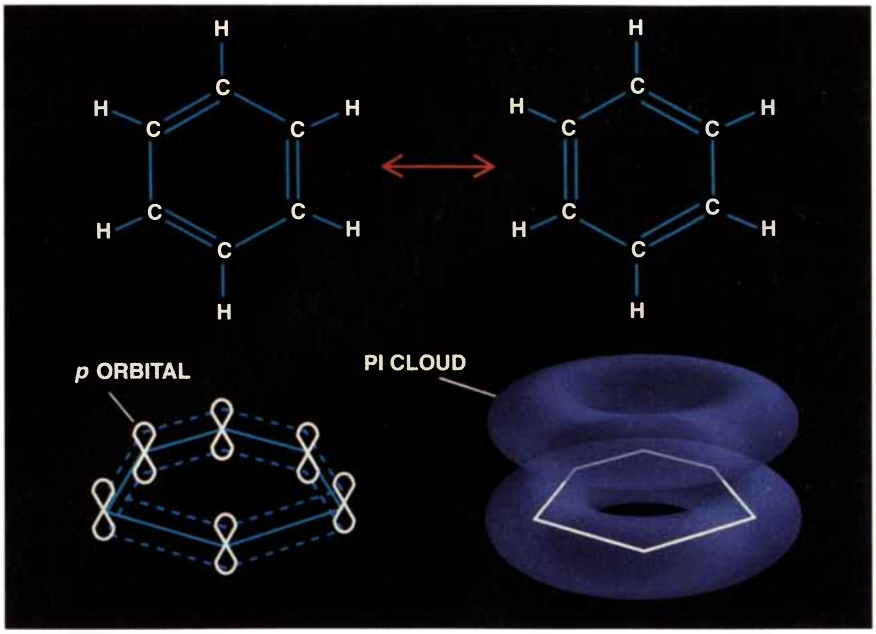
BENZENE, the prototype of the aromatic compounds, reveals its ring shape in this scanning tunneling microscope image. Each molecule appears as a raised bump that has a slight central depression. The fuzziness results from the motion of certain electrons forming the double bonds, called pi electrons.

ing). This molecule belongs to another chemical family, the alkenes. Although its general shape resembles benzene, 5-methylene-1,3-cyclohexadiene is highly reactive. Once a little acid is added, it will rearrange its shape and transform into the aromatic compound toluene (C_7H_8). Toluene is a benzene ring that has a CH_3 substituent attached to one of the ring carbons.

Ordinary combustion produces a wide variety of aromatic compounds. Ideally, when hydrocarbon fuels are burned, they turn into carbon dioxide and water. In reality, complete combustion is rare, and soot and smoke result. Such exhaust gases contain a number of multiring, or polycyclic, aromatic compounds, some of which are highly toxic. Researchers can identify various compounds in the exhaust through chromatography, a process that separates the components of a substance. John C. Fetzer and Wilt R. Biggs of Chevron Research and Technology in Richmond, Calif., and Kiyokatsu Jinno of the Toyohashi University of Technology found that soot from the exhaust of diesel engines contains not only small aromatic compounds such as benzene, naphthalene ($C_{10}H_8$) and phenanthrene ($C_{14}H_{10}$) but also large molecules such as coronene ($C_{24}H_{12}$) and ovalene ($C_{32}H_{14}$). Indeed, soot is thought to be an aggregate of large molecules that have many benzene rings.

The benzene shape forms as easily in space as it does on the earth. Carbonaceous chondrite meteorites, one of the oldest kind of bodies in the solar system, are known to contain many kinds of polycyclic aromatic compounds. These compounds may have formed when hydrogen chemically reacted with carbon monoxide on an appropriate





MOLECULAR STRUCTURE of benzene shows the alternating single and double bonds. Because there are two possible ways the bonds can alternate (*top*), benzene is referred to as a resonance hybrid. Such a resonance picture is associated with the fact that the atomic orbitals in which pi electrons reside (so-called *p* orbitals) overlap with their neighbors. The region of overlap is indicated by the dotted lines (*bottom left*). The pi electrons are hence delocalized—that is, they do not remain static. They travel along common regions, forming doughnut-shaped "pi clouds" (*bottom right*).

surface. It is also possible that the aromatic compounds in meteorites formed in interstellar space and were then trapped by the coalescing solar system. Stars in the late stages of their lives synthesize heavy elements, including carbon, and eject them into space. There the carbon may react with nearby hydrogen. In 1985 Louis J. Allamandola and his colleagues at the National Aeronautics and Space Administration Ames Research Center found the spectra of nebulae to be highly similar to those of automobile soot.

But why does nature find such hydrocarbon rings to be so stable? In general, molecules are stable because all their constituent atoms have sufficient "valence" electrons—that is, electrons that participate in chemical bonding. Such electrons reside in the outer, or valence, "shell," which must fill up with the appropriate number of electrons for the bound atoms to become stable.

One way to do so is to "steal" electrons from other atoms. Chlorine, which has seven valence electrons, takes an electron from a donor atom to fill its valence shell. The addition of an electron negatively charges the chlorine, which will then bind to a positively charged ion. (If the electron donor is sodium, the two will form ordinary table salt.) Such bonds are called ionic.

Carbon, however, prefers to share electrons rather than to steal them. A carbon atom has four valence electrons and requires another four to fill its valence shell. In one of the simplest hydrocarbons, methane (CH_4), each hydrogen shares its sole valence electron with the carbon, thus filling the carbon's valence shell. In turn, the carbon shares a valence electron with each hydrogen, which needs only two electrons to fill its shell. The resulting bonds are known as covalent bonds.

Aromatic compounds do not follow this rule precisely. In benzene, there are only six hydrogen and six carbon atoms, so the carbon must lack the electrons to fill its valence shell. Yet because benzene is extremely stable, one must conclude that all the carbon atoms have filled shells. Therefore, a carbon atom presumably shares more than two valence electrons with its neighboring carbon atom. In other words, more than one bond must exist between some of the carbon atoms. Specifically, three single bonds must alternate with three double bonds in the ring structure. Single and double bonds are written as C – C and C = C, respectively. (Cyclic molecules that have such an alternating pattern are sometimes referred to as annulenes.)

The bonds that form the double bonds are called sigma and pi. The sigma bond is a single bond and is quite strong. In contrast, the pi bond is weak: the electrons in the pi bond are energetically unstable compared with the sigma electrons.

Although chemists refer to specific electrons in bonds, they do so only for accounting purposes. Actually, quantum mechanics allows the pi electrons to travel freely between adjacent carbon atoms. One can therefore regard benzene as having pi bonds of about equal strength. Indeed, benzene is a hybrid of the two possible configurations of alternating double and single bonds and is thus referred to as a resonance hybrid.

Yet the delocalization of pi electrons could not be the full explanation for the unique stability of aromatic compounds. In cyclobutadiene (C_4H_4), a four-carbon ring that contains two double bonds, the pi electrons can roam around the structure. Yet unlike benzene, cyclobutadiene is notoriously reactive and difficult to prepare. Chemists have succeeded in making it only at temperatures below about 20 kelvins, and the compound quickly decomposes on warming. Why would a six-carbon ring be so stable yet a similarly structured four-carbon ring be so reactive?

A partial explanation came in 1931, when the German chemist Erich Hückel devised a rule to determine the stability of the pi electrons within molecules. He treated molecules as atoms. The pi electrons would fill up distinct "molecular orbitals," much as electrons fill up atomic orbitals. Like atoms, molecules would have "magic numbers" that correspond to the number of pi electrons needed to fill up molecular shells. These magic numbers are 2, 6, 10, 14 and so on. Any other number of pi electrons results in unstable compounds.

The rule can be stated in general terms. Single-ring molecules and ions that contain a multiple of four (written as $4n$, where n is any positive integer) pi electrons are extremely unstable. On the other hand, those that have one of the magic numbers of electrons (that is, $4n + 2$ of them) are stable. The rule explained for the first time why benzene, which has six pi electrons ($n = 1$), is stable and why cyclobutadiene, which has four pi electrons, is not.

The Hückel molecular orbital approach was one of the most successful in determining the stability of aromatic compounds. Yet it applied only to molecules of one ring. After World War II, tremendous progress in organic chemistry led to the synthesis of large aromatic molecules, many of which consisted of combinations of several rings. Hückel's law did not seem able to determine the stability of these new, polycyclic molecules.

Further progress had to wait until 1964, when Michael J. S. Dewar, now at the University of Florida at Gainesville, and others developed their concept of resonance energy. While studying the quantum states of pi electrons in molecules, the investigators discovered that the individual bonds of the molecules in a ring are different from those of molecules organized in a chain. In particular, they realized that in chain molecules the same type of bonds have approximately the same binding energies. For example, in a hydrocarbon chain with alternating double and single bonds, the binding energy of C=C is 5.5378 electron volts (eV) and that for C–C is 4.3499 eV. The carbon binds to the hydrogen with about 4.4375 eV.

This fact makes it easy to calculate the amount of energy needed to break a chain molecule into its constituent atoms. This energy is called the atomization energy or the heat of atomization. One simply adds up the binding energies to obtain the atomization energy of the entire chain. For example, butadiene has two double bonds, one single bond and six carbon-hydrogen bonds. The total binding energy is therefore 42.051 eV. The atomization energy of butadiene, calculated through sophisticated molecular orbital theory, is 42.054 eV, in close agreement with the sum of the binding energy values.

Ring structures resist this straightforward approach. For benzene, the sum of the binding energies is 56.288 eV (three double bonds, three single bonds and six carbon-hydrogen bonds). This value should equal the atomization energy. Yet calculations indicate the atomization energy to be 57.157 eV—which is 0.869 eV too large. This difference translates into a stability equal to about 20 kilocalories. Because the carbon-hydrogen bonds are thought to have essentially the same energy whether they are in rings or chains, the 20-kilocalorie difference must lie in the structure of the carbon-carbon bonds—that is, in the ring. Dewar and his colleagues refer to this stabilization energy as the resonance energy. The descriptive term "resonance" refers to the fact that the pi electrons are delocalized and energetically stable. (Many chemistry textbooks list the resonance energy of benzene as 36 kilocalories; however, that value is based on a different definition of resonance energy, one that does not solely involve aromaticity.)

Benzene is not alone. Other ring molecules have atomization energies different from those achieved by summing the binding energies. They may have positive or negative resonance energies. Dewar defined aromatic compounds as those that have positive resonance energies. Today this definition is becoming standard.

The resonance energy per pi electron is actually a better indicator of stability than the resonance energy itself. A large energy per pi electron signifies that the pi bonds, which are ordinarily weak, are being strengthened. Benzene has one of the highest values. Ring molecules such as cyclobutadiene have negative resonance energies per pi electron. These compounds are difficult to synthesize and decompose easily. The resonance energy thus explains why aromatic compounds form so easily and why their ring structure is stable.

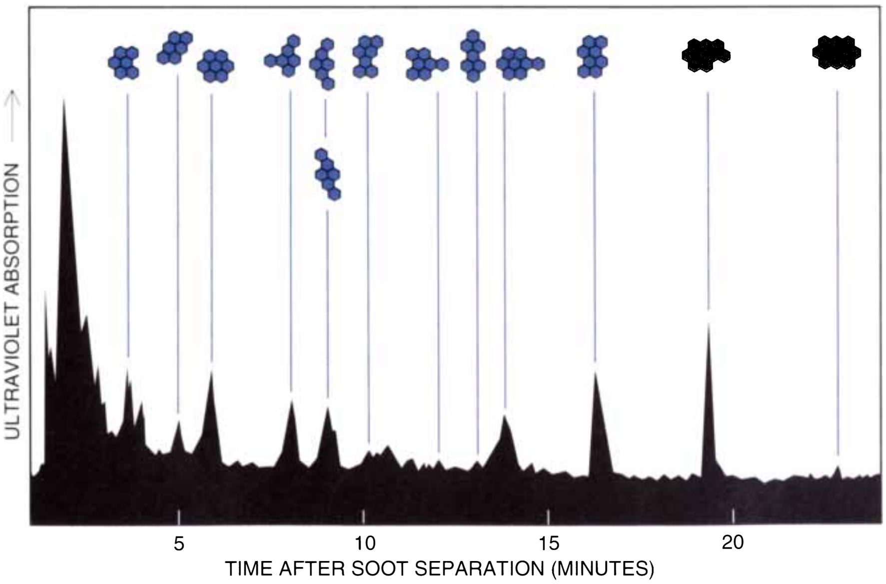
Although Dewar's concept of resonance energy deepened the understanding of aromatic compound stability, it left an unanswered question: Must similar bonds in chain molecules always have the same binding energy? Research so far offers no solution. Dewar himself thought that assigning equal binding energies to the bonds was done for the sake of convenience. He could not explain why ring structures affect stability. Although Dewar's concept is helpful, it relies on empirical relations and thus alone does not determine the true nature of the stability of aromatic compounds.

More important, Dewar's definition does not apply to all ring molecules—specifically, to electrically charged ones. In ions such as the cyclopentadienide ion ($C_5H_5^-$) and the tropylium ion ($C_7H_7^+$), six pi electrons circulate in the ring, just as they do in neutral benzene, but Dewar's definition fails to explain why such compounds are stable.

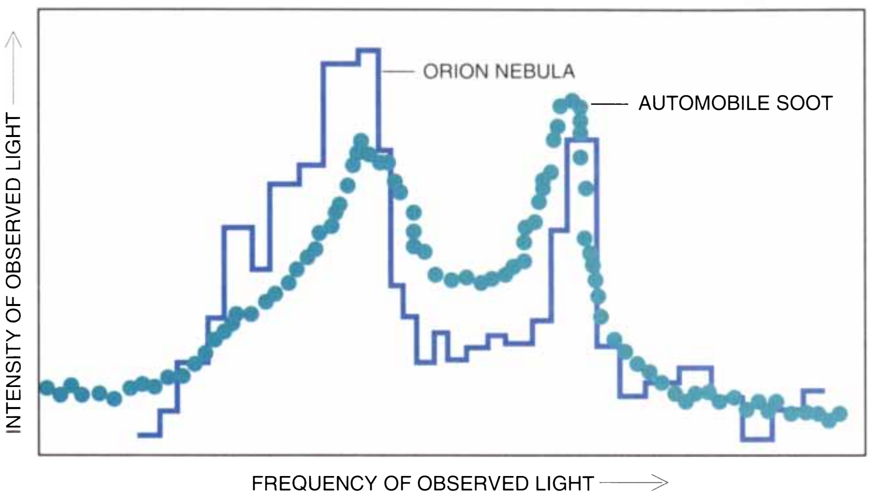
In 1975 I and, independently, Ivan Gutman, Milorad Milun and Nenad Trinajstić of the Rugjer Bošković Institute in Zagreb, Croatia, formulated a method that overcame most of the difficulties Dewar's method could not avoid. We applied topological theory from mathematics to Hückel's molecular orbital theory. Unlike Dewar's method, our description does not require one to calculate the energy in the individual bonds. Instead our theory reinterprets the entire molecular structure in mathematical terms and ultimately enables one to derive the resonance (stabilization) energy for a compound. Essentially, the procedure takes ring structures and mathematically reconfigures them as (hypothetical) chain molecules.

As one can imagine, the calculations involved are quite technical. Fortunately, it is not crucial to understand all the details to gain some insight into the method. The first step is to define a specific equation, called the characteristic polynomial, for the compound. This polynomial can be derived using Hückel's molecular orbital theory. The roots of the polynomial represent the allowed energy levels of the pi electrons.

But these energy levels do not in themselves provide the resonance energy. After all, some distinction needs to be made between the energy from the bonds and the energy from the ring structure itself. To find the difference, one generates a hypothetical structure



ULTRAVIOLET-LIGHT SPECTRUM of diesel engine exhaust reveals that the soot consists of many polycyclic aromatic compounds (*represented as hexagons*). The hydrocarbons, isolated by an organic solvent, show up as strong absorption peaks. The initial peaks represent relatively small aromatic molecules.



ORION NEBULA (*photograph*) probably contains various aromatic compounds. The infrared spectrum of its central region (*diagram*) closely resembles the spectrum of light scattered by ordinary automobile soot (called the Raman spectrum), which is known to contain many aromatic molecules. Researchers now think the compounds are distributed throughout the universe.

made by assuming the pi electrons cannot circulate in a ring—that is, they behave as if they were in a chain. We refer to this structure as a quasichain molecule. One can thus obtain what I call the reference polynomial (Trinajstić and his colleagues use the term "acyclic polynomial").

The solution of the reference polynomial provides the energy levels for the pi electrons of the same atoms organized as a quasichain molecule. The difference between the actual energy levels (the ring) and those in the hypothetical structure (the quasichain) is the "extra" stabilization energy associated with ring structures. This extra energy is a clear indication of an energy contribution from a cyclic source. The resonance energy is expressed in terms of the absolute value of β (written as $|\beta|$). The quantity, which is based on Hückel's molecular orbital theory and has a negative value, represents the extent to which valence shells of adjacent atoms interact with one another.

The topological resonance energy method shows numerically why aromatic compounds are stable. The pi electrons travel through the ring structures in such a manner as to stabilize the entire molecule. Although pi electrons also exist in many chain molecules, they cannot travel in a ring path. The resonance energy is zero.

In this way, our definition of resonance energy is the only unified description available to theoretical chemists. It thus encompasses previous attempts to explain aromaticity. Like Dewar's approach, a positive stabilization energy indicates that the molecule is stabilized by pi electrons moving along a ring path, which means that the molecule is aromatic. (The topological resonance energy reproduces the Dewar resonance energy fairly well, provided β is formally set equal to -3.562 eV.) But our theory avoids Dewar's inescapable assumption that similar bonds in chain molecules have equal binding energies. Perhaps more important, the method applies to ions. It enables one to obtain the resonance energies of charged rings at a level of accuracy identical to those calculated for neutral ones.

The topological resonance energy approach also corroborates the Hückel rule. Hiroshi Ichikawa of Hoshi College of Pharmacy in Tokyo and I have shown that single-ring molecules and ions that have a magic number ($4n+2$) of pi electrons always have positive stabilization energies and are thus aromatic compounds. All other neutral and charged monocyclic molecules have negative energies and are unstable, as Hückel's rule states.

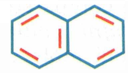
Haruo Hosoya and Kikuko Hosoi of Ochanomizu University in Tokyo and Gutman, now at Kragujevac University in Serbia, used an approach similar to the topological resonance energy theory to show that the Hückel rule applies even to large, multiring compounds. This extension of the rule was previously thought impossible. They postulated that neutral molecules are energetically stable if pi electrons can travel along many hexagonal paths. But if quadrilateral or heptagonal paths exist, the molecules tend to become unstable. This reasoning helps to explain why polycyclic aromatic hydrocarbons have large resonance energies. Because such molecules consist of many connected benzene rings, the pi electrons have several hexagonal paths to follow, which makes the structure highly stable. I proved mathematically that the Hosoya-Hosoi-Gutman rule holds fairly well for most neutral molecules.

Our theory has begun to demonstrate its power and utility. In 1984 Alain Léger and Jean-Loup Puget of the University of Paris VII postulated that there must be a mixture of large aromatic compounds in nebulae. Three years later I determined the kind of aromatic compound that probably exists in nebulae. The compounds could not possibly be small molecules. Topological resonance energy calculations and data from infrared astronomical observations indicated that the intense ultraviolet light from nearby stars would excite and then break down small aromatic molecules. But I showed that large, compact polycyclic aromatic hydrocarbons could survive in nebulae. The calculations indicate that such mol-

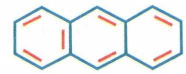
The Stability of Some Carbon Ring Molecules

The topological resonance energies per pi electron (in parentheses), given in terms of the absolute value of ß, is a measure of stability. Stable aromatic compounds have positive energies; negative values represent antiaromatic species. In the structural diagrams, chemists draw only the carbon-carbon bonds of the ring and ignore the carbon-hydrogen bonds. Thus, benzene is simply a hexagon. The red lines in each diagram are the pi bonds.

BENZENE
(0.0454)
VERY STABLE

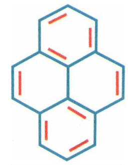


NAPHTHALENE
(0.0389)
STABLE

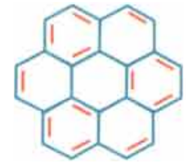


ANTHRACENE
(0.0339)
STABLE

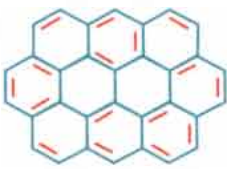
PHENANTHRENE
(0.0390)
STABLE



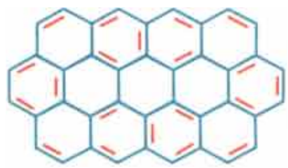
PYRENE
(0.0374)
STABLE



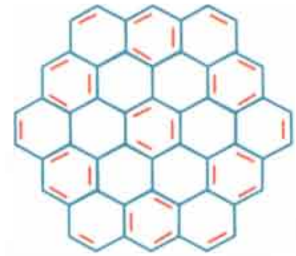
CORONENE
(0.0395)
VERY STABLE



OVALENE
(0.0382)
STABLE



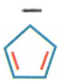
CIRCUMANTHRACENE
(0.0369)
VERY STABLE; MAY EXIST IN NEBULAE



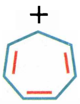
DODECABENZOCORONENE
(0.0388)
NOT YET SYNTHESIZED, BUT THOUGHT TO BE STABLE AND TO EXIST IN NEBULAE

BUCKMINSTERFULLERENE (C_{60})
(0.0274)
LARGE, STABLE, HOLLOW MOLECULE THAT IS THE THIRD FORM OF PURE CARBON; HAS VARIED PROPERTIES WHEN COMBINED WITH OTHER ELEMENTS

CYCLOBUTADIENE
(–0.3066)
EXTREMELY UNSTABLE; CANNOT BE ISOLATED EVEN IF SYNTHESIZED

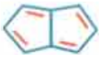


CYCLOPENTADIENIDE ION
(0.0528)
REACTS WITH OXYGEN BUT IS AMONG THE MORE STABLE NEGATIVE IONS

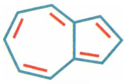


TROPYLIUM ION
(0.0376)
AMONG THE MORE STABLE POSITIVE IONS

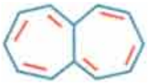
CYCLOOCTATETRAENE
(–0.0744)
VERY REACTIVE; NONPLANAR STRUCTURE PREVENTS PI ELECTRONS FROM TRAVELING IN A RING



PENTALENE
(–0.0269)
EXTREMELY REACTIVE; CANNOT BE ISOLATED EVEN IF SYNTHESIZED



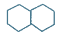
AZULENE
(0.0151)
FAIRLY STABLE



HEPTALENE
(–0.0118)
NONPLANAR; REACTS WITH OXYGEN AND POLYMERIZES WHEN HEATED

***s*-INDACENE**
(0.0046)
THERMALLY UNSTABLE; OXIDIZES EASILY

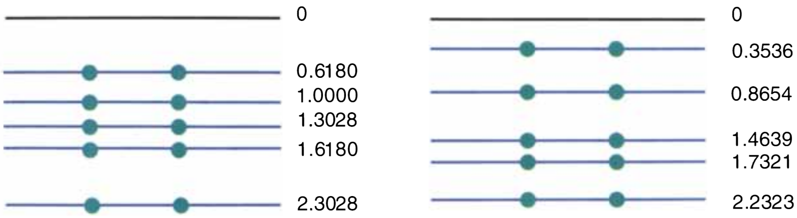
The Topological Resonance Energy of Naphthalene


NAPHTHALENE $C_{10}H_8$

One can determine the stability of this molecule by first solving its characteristic polynomial P(X) (*left column*). The roots of this equation determine the energies of naphthalene's 10 pi electrons (*dots*), which fill up each energy level two at a time. (Five other, higher levels remain vacant and are not shown.) The electron energies (in terms of ß) are compared with those determined by the reference polynomial R(X) (*right column*), which mathematically interprets naphthalene as a "quasichain" molecule. The difference in energies is the topological resonance energy, equal to 0.3886|ß|.

P(X) = $X^{10} - 11X^8 + 41X^6 - 65X^4 + 43X^2 - 9 = 0$

R(X) = $X^{10} - 11X^8 + 41X^6 - 61X^4 + 31X^2 - 3 = 0$



SUM OF ENERGIES = 13.6832

SUM OF ENERGIES = 13.2946

TOPOLOGICAL RESONANCE ENERGY = 13.6832 – 13.2946 = 0.3886

ecules have high resonance energies in excited states and thus can remain stable in the hostile environment. Current astronomical observations are consistent with the predictions.

One of the most likely aromatic compounds in nebulae is dodecabenzocoronene. This molecule, which chemists have yet to synthesize, consists of 19 benzenelike rings arranged as a hexagon. The molecule circumanthracene is also a possibility. I calculated the stability of these compounds in 1987, and this past year Richard D. Broene and François N. Diederich of the University of California at Los Angeles have succeeded in synthesizing circumanthracene. As predicted, the compound is very stable.

The topological resonance energy theory also applies to fullerenes, which are large, hollow, three-dimensional carbon structures. The most easily formed is carbon 60 (C_{60}), produced by laser evaporation of graphite. The shape of this molecule resembles a soccer ball. The versatility and varied properties of such molecules have drawn intense research interest since they were produced in quantity in 1990 [see "Fullerenes," by Robert F. Curl and Richard E. Smalley; SCIENTIFIC AMERICAN, October 1991].

The structure of fullerenes presents problems for the previous aromaticity theories. Hückel's $4n + 2$ rule does not apply to C_{60}, because the molecule has a three-dimensional, multiring structure. Dewar's method also fails because the bond energies for such a system are unknown. In 1988 Hosoya and I predicted that C_{60}, which contains 20 benzenelike hexagons and 12 pentagons, would be moderately aromatic. We calculated the resonance energy per pi electron to be 0.0274$|\beta|$. (The value for the highly aromatic benzene is 0.0454 $|\beta|$.) Negative ions of C_{60} are as aromatic as the neutral molecule. The triply charged C_{60}^{3-}, which becomes superconducting when cooled to very low temperatures, has a resonance energy per pi electron of 0.0241$|\beta|$. Positive ions, however, seem to be much less aromatic.

The topological resonance energy approach is also the only theory to explain in terms of aromaticity a property called diamagnetism. Molecules that are diamagnetic are repelled by magnetic fields. The property results from motions of the electrons in the molecule. The external magnetic field acting on the moving electrons modifies the quantum states of all the electrons. The modification produces a magnetic moment oriented against the applied field.

Although all organic compounds possess this property to some extent, aromatic molecules exhibit very substantial diamagnetism. The repulsion becomes more pronounced as the magnetic field increases. In 1981 I was able to relate the diamagnetic properties of the ring motion to the topological resonance energy. The relation is not simple. In essence, if the pi electrons move along a ring path that stabilizes the molecule, they enhance the diamagnetism; ring paths that destabilize the structure lead to a decrease. The increase in diamagnetism is a manifestation of the molecule's strong tendency to remain stable; it tries to neutralize the effect of the external magnetic field.

There are two potential drawbacks in using topological resonance energy as an index of stability. First, the characteristic polynomial of the theory relies on Hückel's molecular orbital theory, which is not free of certain approximations and thus may contain some flaws. Nevertheless, it is the only framework available that applies to the analytic study of aromaticity.

Second, topological resonance energy actually represents energetic (thermodynamic) stability. It is possible that some compounds may be energetically unstable but chemically inert. That is, the molecule might have a negative resonance energy yet will not readily undergo chemical reactions for some reason. In such cases, the topological resonance energy would not represent the observed stability of the molecule.

Fortunately, most energetically stable molecules are also chemically stable. The general predictive ability of the topological resonance energy method remains valid for all but a few compounds. It seems safe to say, after more than 100 years of theoretical attack, that an explanation for the stability of aromatic compounds has finally been achieved. Investigators should find it a useful and predictive tool with which to explore the chemical world.

FURTHER READING

A NEW DEFINITION OF DEWAR-TYPE RESONANCE ENERGIES. J. Aihara in *Journal of the American Chemical Society,* Vol. 98, No. 10, pages 2750–2758; May 12, 1976.

GRAPH THEORY AND MOLECULAR ORBITALS, Part 19: NONPARAMETRIC RESONANCE ENERGIES OF ARBITRARY CONJUGATED SYSTEMS. I. Gutman, M. Milun and N. Trinajstić in *Journal of the American Chemical Society,* Vol. 99, No. 6, pages 1692–1704; March 16, 1977.

UNIFIED THEORY OF AROMATICITY AND LONDON DIAMAGNETISM. J. Aihara in *Journal of the American Chemical Society,* Vol. 103, No. 19, pages 5704–5706; September 23, 1981.

AROMATICITY AND DIATROPICITY. J. Aihara in *Pure and Applied Chemistry,* Vol. 54, No. 5, pages 1115–1128; May 1982.

CHEMICAL GRAPH THEORY, Vol. 2. N. Trinajstić. CRC Press, 1983.

SPHERICAL AROMATICITY OF BUCKMINSTERFULLERENE. J. Aihara and H. Hosoya in *Bulletin of the Chemical Society of Japan,* Vol. 61, No. 7, pages 2657–2659; July 1988.

SCIENTIFIC AMERICAN

CORRESPONDENCE

Offprints of more than 1,000 selected articles from earlier issues of this magazine, listed in an annual catalogue, are available at $1.25 each. Correspondence, orders and requests for the catalogue should be addressed to W. H. Freeman and Company, 4419 West 1980 South, Salt Lake City, Utah 84104. Offprints adopted for classroom use may be ordered direct or through a college bookstore. Sets of 10 or more Offprints are collated by the publisher and are delivered as sets to bookstores.

Photocopying rights are hereby granted by Scientific American, Inc., to libraries and others registered with the Copyright Clearance Center (CCC) to photocopy articles in this issue of SCIENTIFIC AMERICAN for the flat fee of $1.50 per copy of each article or any part thereof. Such clearance does not extend to the photocopying of articles for promotion or other commercial purposes. Correspondence and payment should be addressed to Copyright Clearance Center, Inc., 27 Congress Street, Salem, Mass. 01970. Specify CCC Reference Number ISSN 0036-8733/92. $1.50 + 0.00.

Editorial correspondence should be addressed to The Editors, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017. Manuscripts are submitted at the authors' risk and will not be returned unless accompanied by postage.

Advertising correspondence should be addressed to Advertising Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017.

Address subscription correspondence to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 3187, Harlan, IA. 51537. Telephone inquiries: 1-800-333-1199, U.S. only; other 515-247-7631/32. The date of the last issue on subscriptions appears on each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (if convenient, a mailing label of a recent issue) as well as the new one.

Name ____________________

New Address

Street ____________________

City ____________________

State and ZIP ____________________

Old Address

Street ____________________

City ____________________

State and ZIP ____________________

Spider Webs and Silks

Spiders' silk and web designs are governed by the same rules and constraints that apply to human materials science and structural engineering. So spiders may be able to teach us a thing or two

by Fritz Vollrath

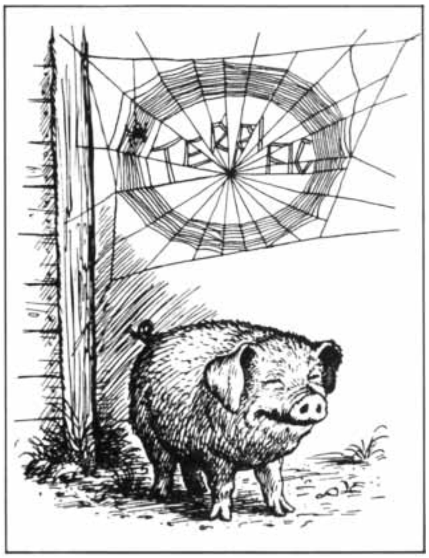
Spiders are perhaps better known for their magnificent feats of structural engineering than for the manufacture of choice materials. Unlike the thread of the silk moth larvae (*Bombyx mori*), spider silk has not yet been harnessed for mass production. Webs have been used as dressings for wounds and even as nets for fishing, but the silk itself has found employ only as cross hairs in optical instruments.

Yet the silk of a spider's web has many fabulous qualities. It is much stronger than the silk of a silkworm. Moreover, it is variable: spiders can produce different types of silk for different functions. This range of qualitatively different silks has led to the evolution of a stunning array of web designs.

By comparing the properties of spiders' silk with those of its well-studied counterpart, silkworm silk, we can begin to understand how spiders became equipped to construct their remarkable webs. And by studying the functional mechanics of certain webs, we can see how the stiffness of most spider silk limits the range of web designs and their prey-catching ability. Like all building materials, silk ultimately constrains architectural endeavors. Any structure, be it web or bridge, is only as effective as the materials from which it is made. That is, unless the builder has a trick up its sleeve.

The dexterous heroine of E. B. White's tale, *Charlotte's Web,* offers an example of a spider that has just that. With brilliant tinkering, the garden cross spider (*Araneus diadematus*) has turned an inherent weakness of silk—its softness when wet—into an invaluable benefit. As we will see, coating specific threads with water made the spider's web cost-effective. It also allowed the ancestors of *A. diadematus* to expand into a novel ecological niche, that of open spaces.

FRITZ VOLLRATH is assistant professor at the Institute of Zoology at the University of Basel in Switzerland and research associate in the department of zoology at the University of Oxford. After studying the ecology of social and kleptoparasitic spiders at the Smithsonian Tropical Research Institute in Panama, Vollrath turned his attention to the evolution of spider webs, particularly orb webs, to study how they are constructed and how they work.



The garden cross spider's invention provides an example of an important evolutionary phenomenon: constraints. Evolutionary change is brought about by adaptive shifts in behavior and morphology. These gradual changes eventually come up against physiological or physical limits. Subverting such a constraint opens the way to new opportunities. The analogy with human technological progress is obvious.

Spider and silkworm are not at all closely related, yet their silks are quite similar. Both use keratin, a protein that is also found in hair, horn and feathers. In hair and horn, this protein is called alpha-keratin, and it appears as braided, helical strands of amino acid chains. Ancestral arachnids may have used alpha-keratin as a cement to cover their eggs; ancestral insects may have used it to attach their pupal cases to vegetation.

Both silkworm and spider have evolved a pump-and-valve pressure system that enables them to make threads from these proteins. Instead of exuding a slowly drying gel of alpha-keratin chains, glandular ducts thickened the substance into a highly viscous state: a nematic liquid crystal, in which the molecules are organized in parallel lines. Strong shearing forces applied to the emergent thread by an extrusion nozzle cause many of the alpha chains to form a stable, tertiary structure, called a beta-pleated sheet. As we observe in silkworm silk, such sheets can fold back onto one another in accordionlike fashion, stacking up to form a crystal.

These protein crystals are in turn embedded in a rubberlike matrix composed of amino acid chains that are not linked into beta-pleated sheets. Instead these helical strands are tangled up in a state of high entropy. It is precisely this randomness that lends silk, like rubber, exceptional elasticity. Stretching the thread pulls the protein strands out of disarray—which they resist—whereas releasing the thread allows them to contract back into blissful disorder.

Silk's elasticity is balanced by its strength. Because it is a composite material, like glass fibers embedded in a resin, silk is strong. Its crystals and matrix resist breaking. A stretched thread usually snaps because a crack on the surface cuts into it like a wedge. Forces acting along the fiber concentrate at the crack and cause it to rip with increasing speed ever deeper into the material. Such cracks, however, can travel only if they do not encounter resistance. The crystals in the rubber matrix provide

ARANEUS DIADEMATUS, the spider who starred in E. B. White's _Charlotte's Web_ [_see illustration on this page_], has overcome the constraints of wet silk to fashion a highly efficient and structurally advanced orb web, like the one shown at the right.

obstacles that divert and weaken the rending force.

The raw, liquid silk is produced in large abdominal glands and emitted through spigots. In the silkworm a pump in the larval head releases a double filament. In spiders the filament emerges from veritable arachnid gun towers: batteries of spigots arrayed atop three sets of paired spinnerets. (The spinnerets evolved from pairs of legs that may have held the animal's eggs close to its body.) Because the silk emanates from paired sources, each fiber is composed of two strands.

Although researchers have not yet uncovered every detail of the valves and the silk production mechanisms, we must assume that some combination of the forces applied by the presses and of the molecular composition of the liquid alpha-keratin acts to produce certain kinds of silk. Indeed, amino acids vary in the silkworm and the spider and between different silks of the same spider.

The spider, however, seems to be uniquely able to exert some control over its silk production system. Some spiders can modify a thread while they spin it by altering its diameter, strength and elasticity. It is likely that they do this by controlling the valves rather than the chemical composition of the silk. If radically different silk is required, they just switch to another gland. The female garden cross spider, for instance, can produce at least seven silks.

In the near future, people may be able to use these variable spider silks for more than optical devices. The genes for such silks have recently been synthesized, expressed in microorganisms and patented.

The evolution of myriad spider silks is reflected in the dazzling abundance of web types. All designs derive from a simple silk mesh used by ancestral spiders to line earth burrows more than 380 million years ago. Presumably these spiders, like the living fossil *Liphistius,* may have used threads radiating from their burrows to serve as signal lines and trip wires for unwary travelers. (Just what these travelers were is open to imagination because insects were only beginning to emerge in the Devonian.) For efficient signal transmission, these threads may have been raised from the ground, strung between tiny silken stilts. The liphistiid form of trip-wire web may be ancestral, but in some of its modern expressions it certainly is not primitive: the suspension of its few radiating threads can be technically shrewd.

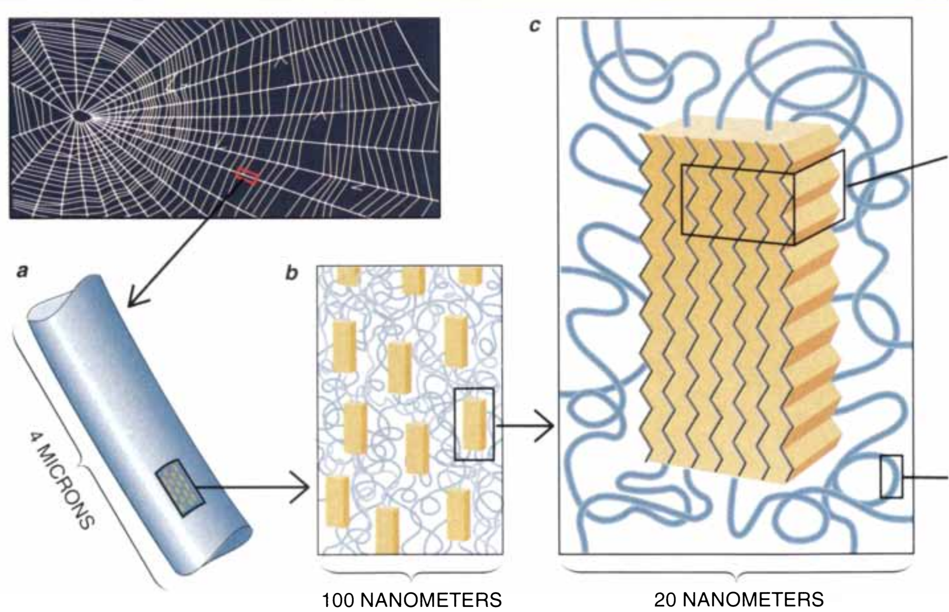
Over time, trip wires gave rise to traps that made more extensive use of silk. The design of the funnel-web spider (*Agelena labyrinthica*), for example, consists of a fine, strong sheet, held taut by lateral as well as by lower and upper guy, or guide, ropes. The spider runs about on top of the sheet to collect prey that have fallen or jumped on board. Vibrations alert the spider to the presence and location of the prey.

In addition, *A. labyrinthica*'s web has a mechanism that amplifies prey movements. Some threads are always stretched out of place and snap into a new position with a twang when dislodged. This adaptation may well have been necessary to ensure continued success in capturing dinner: in the arms race between web spiders and their prey, the prey is forever improving ways to avoid entrapment. For instance, butterfly scales readily stick to adhesive web threads, yet they also easily come off the butterfly—allowing it to escape while leaving behind a swath of shed scales and inoperative silk.

Another basic design is the aerial knockdown web: the three-dimensional complex of threads one often finds in bushes or hedges. These structures are notoriously difficult for the spider to control, a fact that is readily apparent when one compares the number of kleptoparasitic guests in these webs with the number in other web types. In three-dimensional structures, such intruders, often other spiders, can be very common. They move about with relative impunity to glean small prey or even to steal big insects from under the proprietor's mighty fangs.

To ensure a regular meal, the builders of most three-dimensional webs set aside clearly defined areas for capturing prey, regions that they can monitor with relative ease. In one such web, that of the black widow (*Latrodectus mactans*), highly modified lower guy threads form the capture area—a dense forest of sticky mesh. At their ends, the taut fibers are coated with a viscous glue. As soon as they adhere to a crawling insect, their attachment to the ground breaks, and the victim is lifted up by the recoiling thread.

Other three-dimensional webs, like that of the filmy dome spider (*Neriene radiata*), incorporate nonsticky sheets to serve as the capture area. This spider travels about underneath its domed sheet to attack prey that have fallen onto it, stopped in midair by the tangle of upper guy threads.



The Structure of Spider Silk

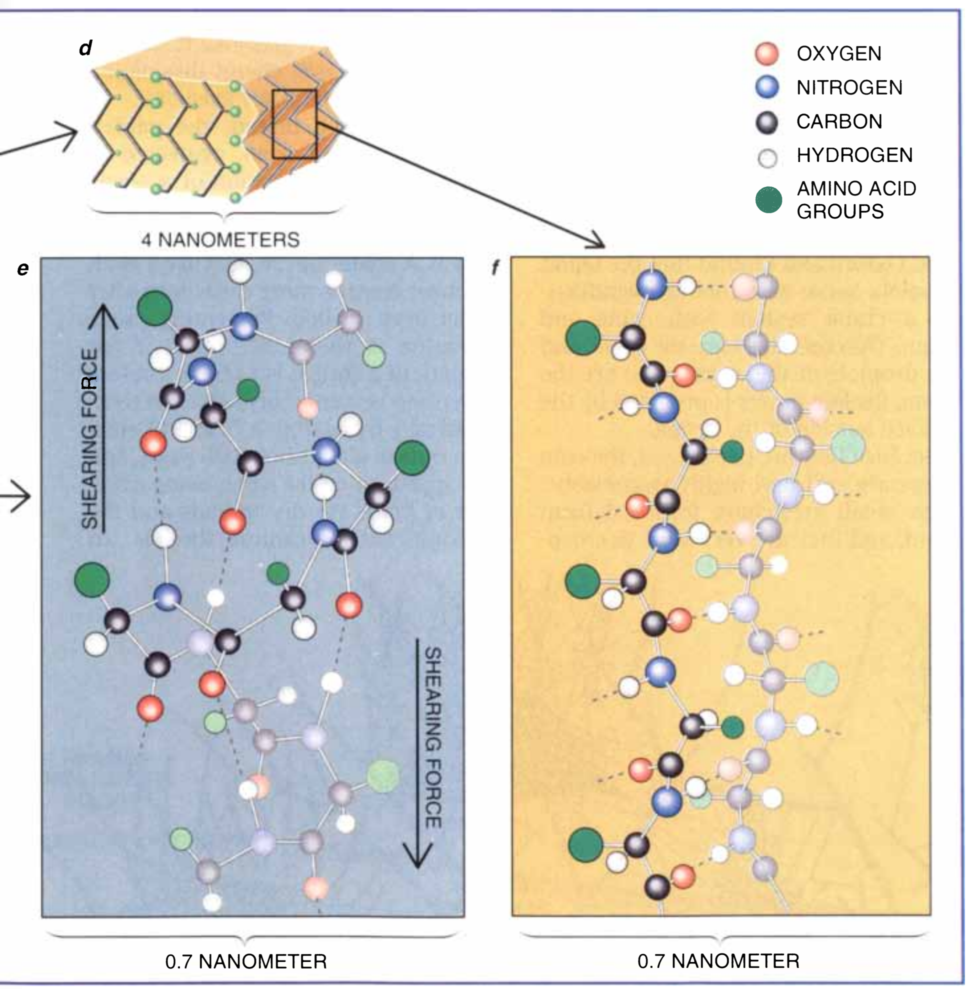
Enlargement of spider thread (*a*) shows it to be a composite material made of strands of disordered amino acid chains and ordered crystals (*b* and *c*). Each crystal is made up of different-size amino acid groups pressed into an accordionlike formation, called beta-pleated sheets (*d*). The surrounding strands are called alpha helixes; their contracted disarray provides silk with its elasticity. When silk is emitted, shearing forces like those shown at the right (*e*) are applied to some alpha helixes. Consequently, their hydrogen bonds break and they become beta-pleated sheets (*f*), as the similarity of the highlighted molecular strands shows. Most of what is known about the molecular structure of silk comes from the silkworm. In this illustration, it is assumed that spider silk resembles that of the silkworm.

Although each of the three types of webs described above has a different modus operandi, they all share the same basic principle. An insect is stopped in its path, and its presence is advertised by vibrations or changes in the tension of the structural threads. The business of stopping the insect, that is, of dissipating its kinetic energy, is accomplished by dislodging or breaking strands. An analogy is that of a hapless balloonist landing in a tree: the strength and elasticity of the branches, their tensions and interconnections determine how far our balloonist penetrates the canopy before his or her fall is arrested.

The energy-dissipating thicket of threads in three-dimensional webs is missing in orb webs, such as that of the garden cross spider. These two-dimensional structures work in a very different fashion—and it is in these webs that the coevolution of remarkable structures and building materials can be seen most vividly. Because there is no dense tangle to stop the insect, a few broken strands would allow it to fly straight through the web. Thus, a two-dimensional web requires more refined materials and architecture than those used in three-dimensional webs.

Probably the best known of all two-dimensional designs is the polar mesh typified in the orb web of the garden cross spider. The web is a familiar one. Firm, dry threads radiate from the center like the spokes of a bicycle wheel. Laid down in a tight spiral fixed to these radii is one long thread of sticky silk. The functions of this, and any, web are diverse—including service as an early-warning system against predators, as a burglar alarm and, for a courting male, as harp and dance floor. But its overriding function is to trap.

To be an effective trap, the web must be able to stop the hurtling insect—comparable, from the web's perspective, to a guided missile—and retain it long enough for the spider's inspection and bite. These functions are aided by the basic architecture of the web. But the two-dimensional orb web cannot ensnare prey by dissipating the victim's energy through breaking strands, as does the three-dimensional web. Instead the main work of dissipating and absorbing the insect's kinetic energy is done by the inherent properties of the silk, in particular the capture thread.



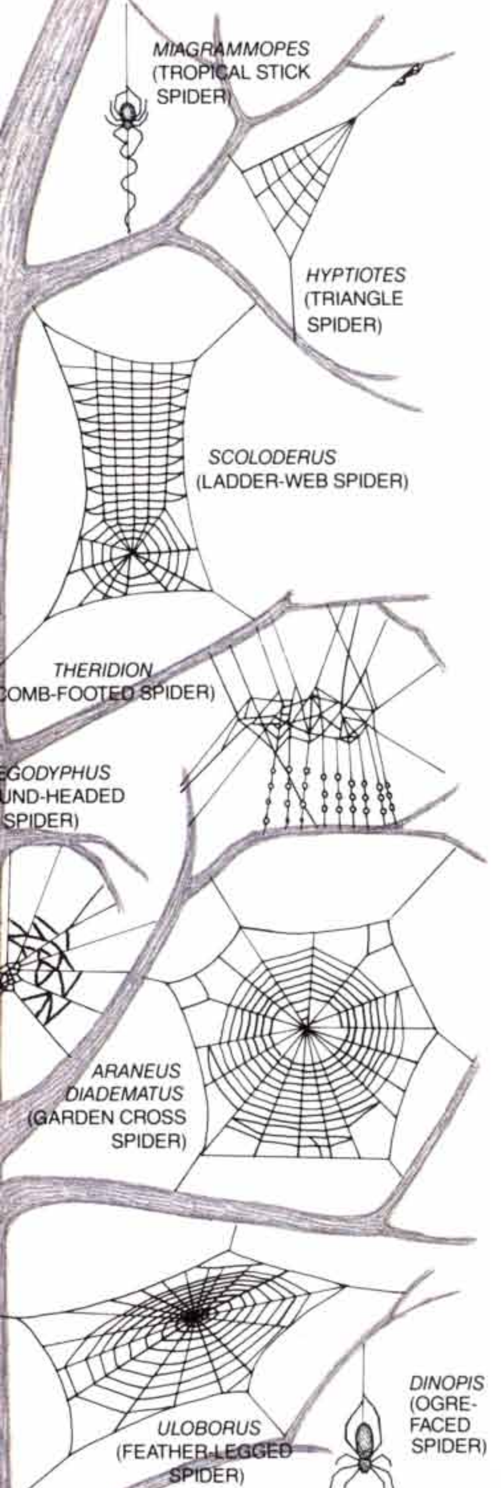
This fiber is strong at great extensions, for it must not break, and soft at small extensions, for it must not provide the trapped insect with purchase. Its silk must stretch without rapid elastic recoil—otherwise the prey would be flung back, trampoline fashion, whence it came. The silk must also be elastic enough to contend with the continual stretching and relaxing of a network buffeted by wind and distorted by flexible supports, such as blades of grass. Indeed, because the threads of the capture area are highly sticky, any substantial sagging or flapping about could quickly cause them to adhere to one another, an event that would create gaps in the even mesh of the trap and provide loopholes for potential prey.

To be effective, therefore, the orb web needs strong capture threads that show great energy-absorbing deformation yet spring back rapidly without sagging: a tall order, since the combination of these properties is normally absent from silk. Although they are strong, typical spider threads, such as those in the web's radii, generally break when they are extended beyond 25 percent. Such threads have great stiffness at smaller extensions, but when relaxed they sag for quite some time before resuming their original length.

The capture silk of *A. diadematus,* however, shows an altogether extraordinary behavior not found in any other type of silk. Threads of the capture spiral can be rapidly contracted to a small fraction of their original length without sagging. The same thread can also be stretched to over four times its original length without snapping. Thus, the capture silk effectively has an extensibility of several thousand percent without any obvious sagging and with a fast, yet perfectly buffered, recoil. Although it would seem that this unlikely silk is intrinsically different from other silks, that is not the case.

Instead *A. diadematus* has devised an ingenious trick. Its capture silk is modified into something that transcends a composite material and that is more appropriately called a mechanism. The strands of the garden spider's capture spiral consist of a pair of core fibers surrounded by a coat of highly viscous liquid. The liquid originates in special glands that lie next to those producing the core fibers. It is applied to the outgoing core fibers as a continuous film, which—after the uptake of atmospheric water—quickly separates into little droplets.

Chemical analysis done by Edward K. Tillinghast and his colleagues at the University of New Hampshire and by Robert J. P. Williams, Wayne Fairbroth-



er and me at the University of Oxford showed that this liquid contains a number of interesting chemicals. Stickiness is provided by microscopic glycoprotein doughnuts that straddle the core fibers like beads on a string. The glycoproteins are bathed in a solution of five major compounds closely related to amino acids that serve as the spider's neurotransmitters. These amino acids have been pressed into novel service because of their hygroscopic, or water-attracting, chemical ability.

This arrangement of drawing water to the capture threads is no accident. Silk is typically stiff when dry. It resists stretching and if stretched shrinks back to its former length over a matter of hours. But when water seeps into the silken core fibers, they become easily deformed and highly viscoelastic. Under most circumstances, such softness and slow recovery are undesirable, because threads should not stretch to many times their length—just imagine a web in which all threads have this property and then imagine what havoc would be wreaked by the weight of a traversing spider. Consequently, all ancestral and most modern silks are dry and may even be shielded from water by a thin oily coat.

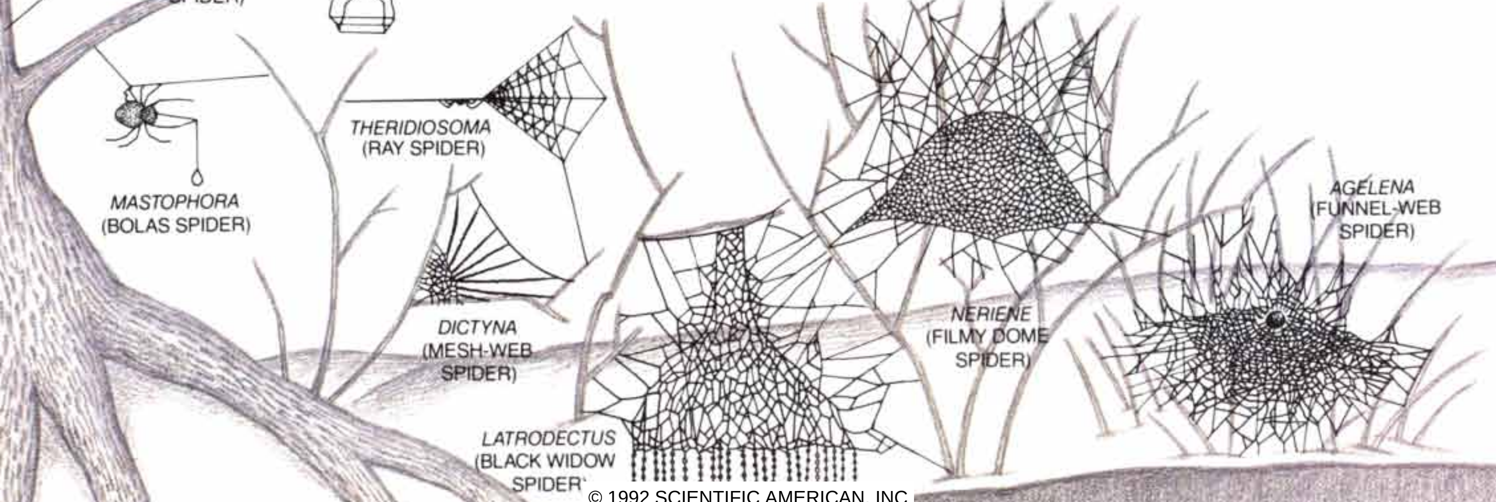
But in the capture spiral, these undesirable characteristics are useful because the threads can absorb much more kinetic energy when wet. Although we do not know how the water does its plasticizing on a molecular level, we do know how it modifies the capture thread on a macroscopic level. Donald T. Edmonds of the Clarendon Laboratory at Oxford and I found that the liquid droplets serve as miniature windlasses—a crank system with cable and drum. The core fibers are the cable, and the droplets of the watery coat are the drum. Reeling power is provided by the surface tension of the drops.

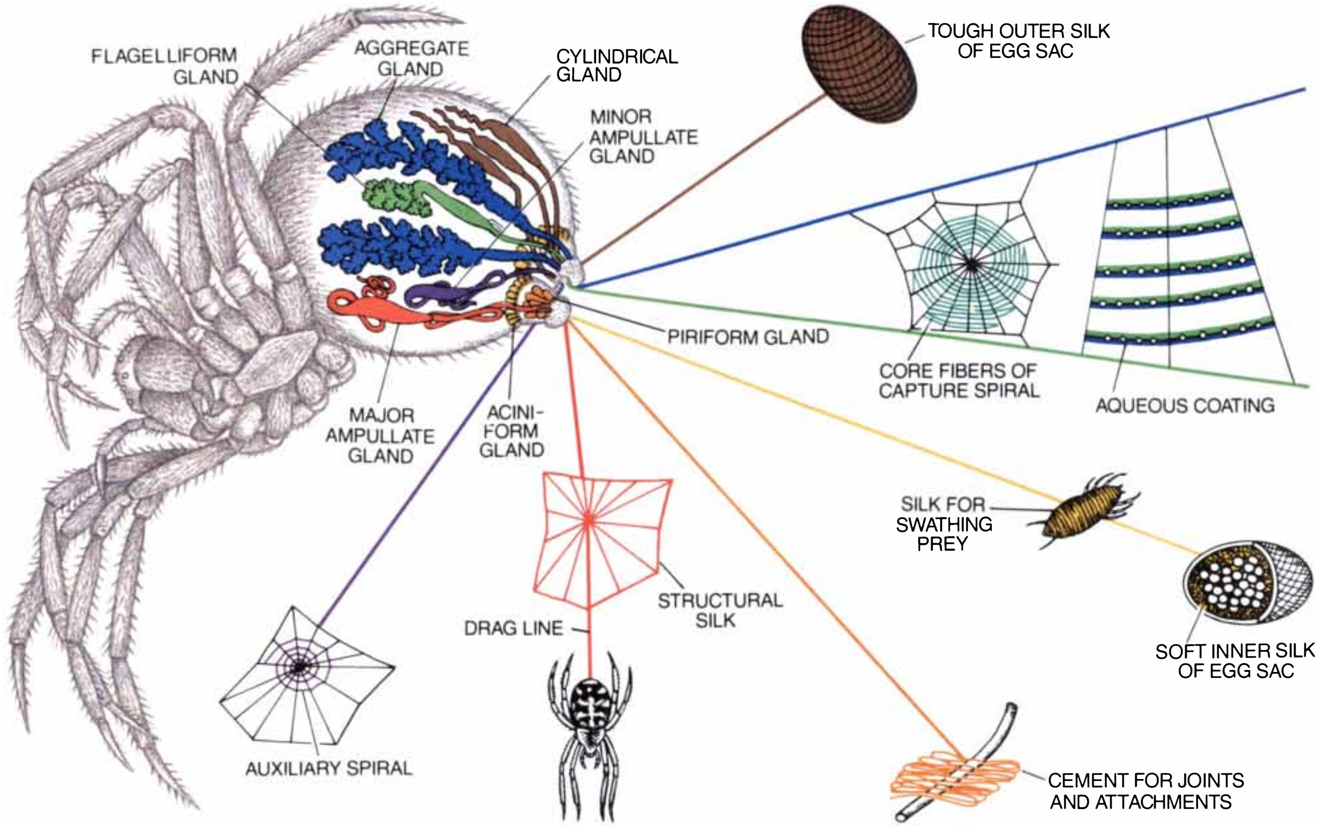
Because they are plasticized, the core fibers are soft and highly viscoelastic. Even small stretching forces deform them, and they are very slow in creeping back. The surface tension of the droplets, however, is strong enough to reel in the floppy core fibers. Contracting and stretching the sticky thread in rapid succession wind and unwind the core fibers inside the droplets. Thus, the entire system of core fibers and coating is always under tension, keeping the sticky thread taut. Energy applied by buffeting winds or blundering insects is not absorbed by the silk itself but by the entire system.

The core fibers do their share of the work as well. Plasticized and therefore essentially like reinforced rubber, they benefit directly from the fact that entropic elasticity is temperature dependent. Because the kinetic energy of the prey is largely converted into heat, the thread warms up. The heating increases entropy, and consequently the core fibers grow stronger. The absorbed energy of the hapless prey actually strengthens the capturing thread and does so only because of the spider's clever trick of applying aqueous coating.

While being stretched, the plasticized core fibers show very little resistance at first—for the initial 300 percent or so, that is. Then they suddenly develop increasing resistance. At this point in their extension, they behave just like dry threads. When fully stretched, the polymer chains enter a glasslike transition in which viscoelastic energy dissipation takes over from entropic elasticity.

By coating the thread, the garden cross spider has also protected it against the first formation of a crack. Even the smallest surface blemish is extremely dangerous for a material under tension. A spider thread, just like a cloth or rubber band, is more easily torn after it has been nicked. Preventing crack formation is therefore crucial if the strength of a thread is to be increased. Prevention is commonly achieved technologically by coating a "thread," such as an optical glass fiber, with paint. Spiders appear to do the same, using a fine layer of lipids for dry threads and the aqueous coat for capture threads. To



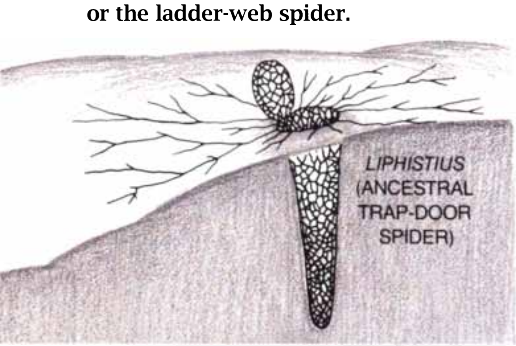


DIFFERENT SILKS for different functions can be produced by the same spider. *A. diadematus* can switch between silks with varied amino acid compositions. The spider uses abdominal glands and spigots to produce seven kinds of silk.

guard against bacteria and fungi that could feast on the nutritious proteins of the coating, spiders incorporate fungicides and bactericides into the droplets. (This antiseptic quality of the spider web may account for its renown as a folk remedy for dressing wounds.)

The evolutionary invention of a web that is relatively cheap—both in construction time and building material—enabled orb spiders like *Araneus* to invade sites dangerously exposed to the elements. The orb is built into a two-dimensional frame that is supported by anchor threads of any length. This feature allows these spiders to take to the open air, to leave behind the dense tangle of supporting vegetation required by spiders with three-dimensional webs. Thus, orb spiders effectively fly, although they have no wings. And by inhabiting the air space of forests and glades, they can prey on flying insects.

WEB DESIGNS exhibit great diversity. Spiders' common names can reflect their web's shape—such as the filmy dome or the ladder-web spider.

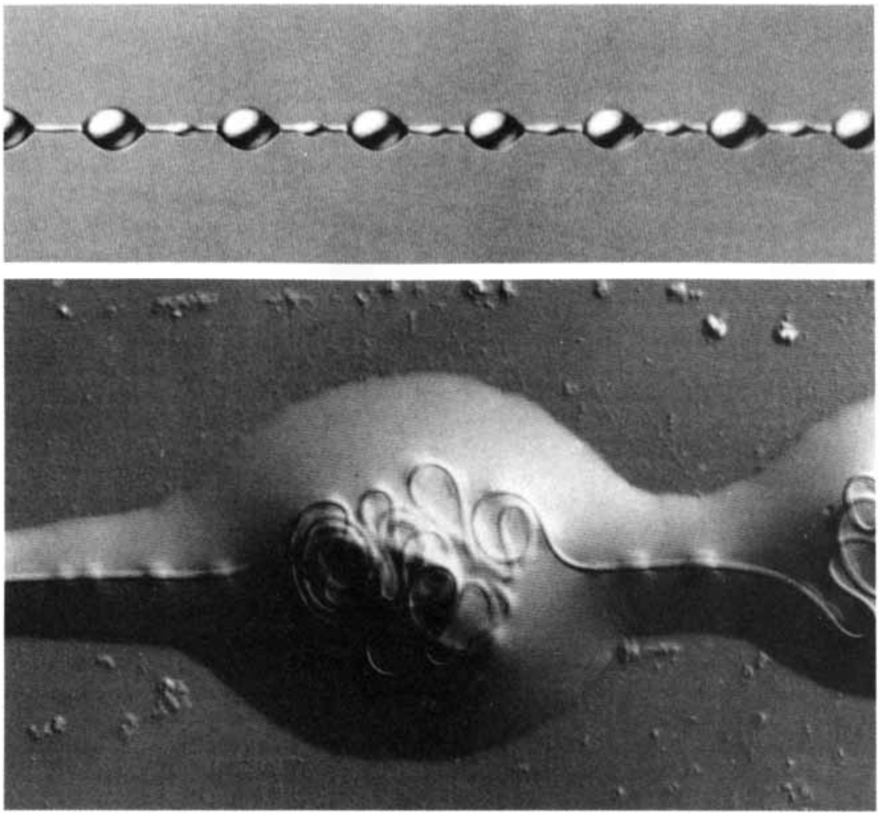


The adaptive radiation of two-dimensional webs that use plasticized silk and aqueous glue has resulted in a large variety of web architectures and ways of capturing prey. Orb spiders have been so successful in this ecological niche that, except for robber-flies, damsel flies and dragonflies, few insect groups provide significant competitors. Cheaply built, ephemeral webs that can resist huge distortions are also useful in other exposed areas like meadows, and orb webs abound there as well.

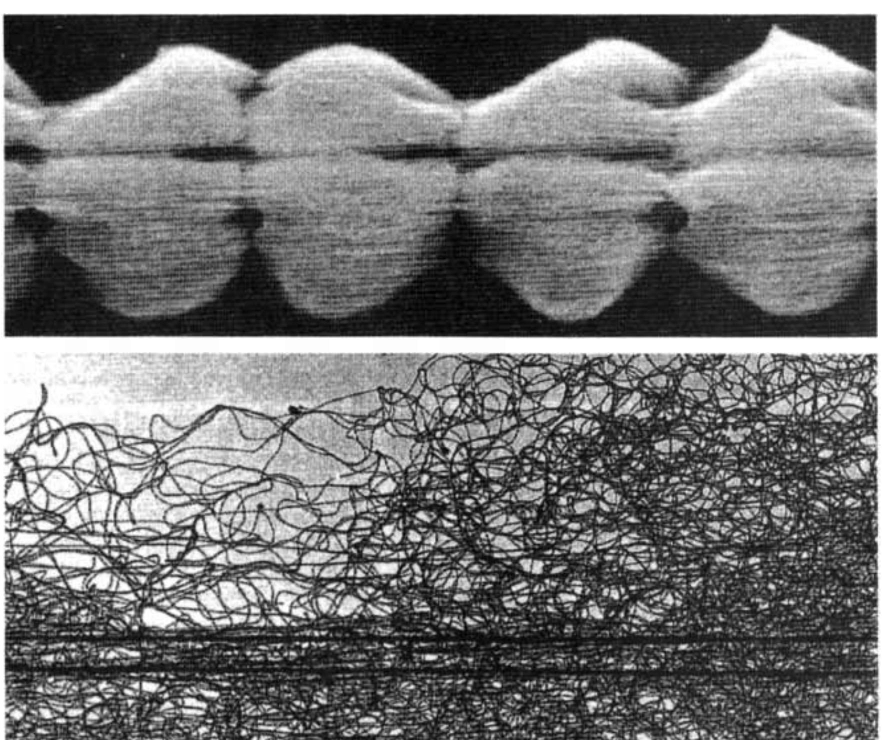
Even the orb webs of a small group of very different spiders, while similar in overall design, cannot compete with *Araneus*'s prey-catching prowess, because they do not incorporate its plasticized silk and revolutionary windlass system. These cribellate spiders, named after their platelike extra spinneret, or cribellum, have devised other strategies to improve the strength and stickiness of their capture threads. Such spiders, including *Uloborus,* use dry silk throughout their two-dimensional webs, but this silk is highly specialized in its own right. From hundreds of the finest imaginable spigots grouped together on the cribellum, they produce an uncoated, multistranded, or hackled, band of fine threads. Each thread has a diameter of less than 0.05 micron, as opposed to a diameter of about 1.0 micron for ecribellate capture silk. (Ecribellate spiders are those that, like *Araneus,* lack the ancestral cribellum plate and whose threads are coated.)

The cribellate spider draws the threads through a fine comb located on the shin of its hind leg. This combing process either prestretches and stiffens the threads or charges them electrostatically. In any event, hackling causes the fibers to puff out and assume the function of a spring. In the center of the spring lie core strands that are much thicker and stronger. Because hackled fibers are completely dry, this web's stickiness appears to be achieved by electrostatic forces or the entanglement of the prey's appendages and hairs.

The impact of an insect apparently stretches the core as well as the hackled threads until one after another breaks, thus absorbing energy. Hundreds of stiff threads work together, and so they are quite effective. By breaking strands to neutralize energy, the cribellate web works in principle just like the dry, three-dimensional tangle webs. Therefore, these threads are worse at absorbing energy than are the water-coated threads of ecribellate spiders. This



CAPTURE THREAD of an ecribellate spider, such as *A. diadematus,* is covered with an aqueous coat (*top*), shown magnified 100 times. Surface tension within each drop causes the core fibers to bunch up, creating a windlass system, shown in its contracted state (*bottom*), magnified 300 times.



CRIBELLATE SPIDER THREAD is combed into hundreds of strands—shown magnified 200 times (*top*) and 1,000 times (*bottom*)—that puff out as a result of hackling. The uncoated, cribellate threads are as sticky as coated, ecribellate threads, but combing requires more energy and time than coating does.

shortcoming is partly counterbalanced, however, by a comparable if not greater stickiness and an extended working life for the web. Despite these compensating factors, cribellate orb weavers are outnumbered by their ecribellate cousins by a factor of one to 100, both in terms of species and individuals.

This seeming discrimination has its roots in the same considerations that govern all construction projects: economics. The capture thread of the cribellates is much more expensive to produce. Centimeter for centimeter it requires more raw material in the form of protein, and it carries higher assembly costs, since combing the hackled strands takes time as well as energy. Indeed, a cribellate spider takes four times longer than an ecribellate spider to produce the same length of capture spiral. So, although both types of thread are similarly effective, the ecribellate thread is more efficient because it gives comparable value at lower cost.

By tinkering, ecribellate orb weavers like *A. diadematus* have ingeniously turned a fine but still limiting material—basic, dry silk—into a fabulous absorber of energy. The trick lies in the hygroscopic chemicals that attract the water that plasticizes the silk and powers the windlass that keeps the capture thread taut.

The spider's invention of this astonishing mechanism added efficiency to the architecture of the orb web and allowed ecribellate orb weavers to assume the important role that they occupy today in most ecological food chains. What else would one expect from many millions of years spent on research and development? After all, if the recent interpretation of an early Cretaceous fossil as an orb weaver is correct, then orb spinners have had more than 180 million years to get it right.

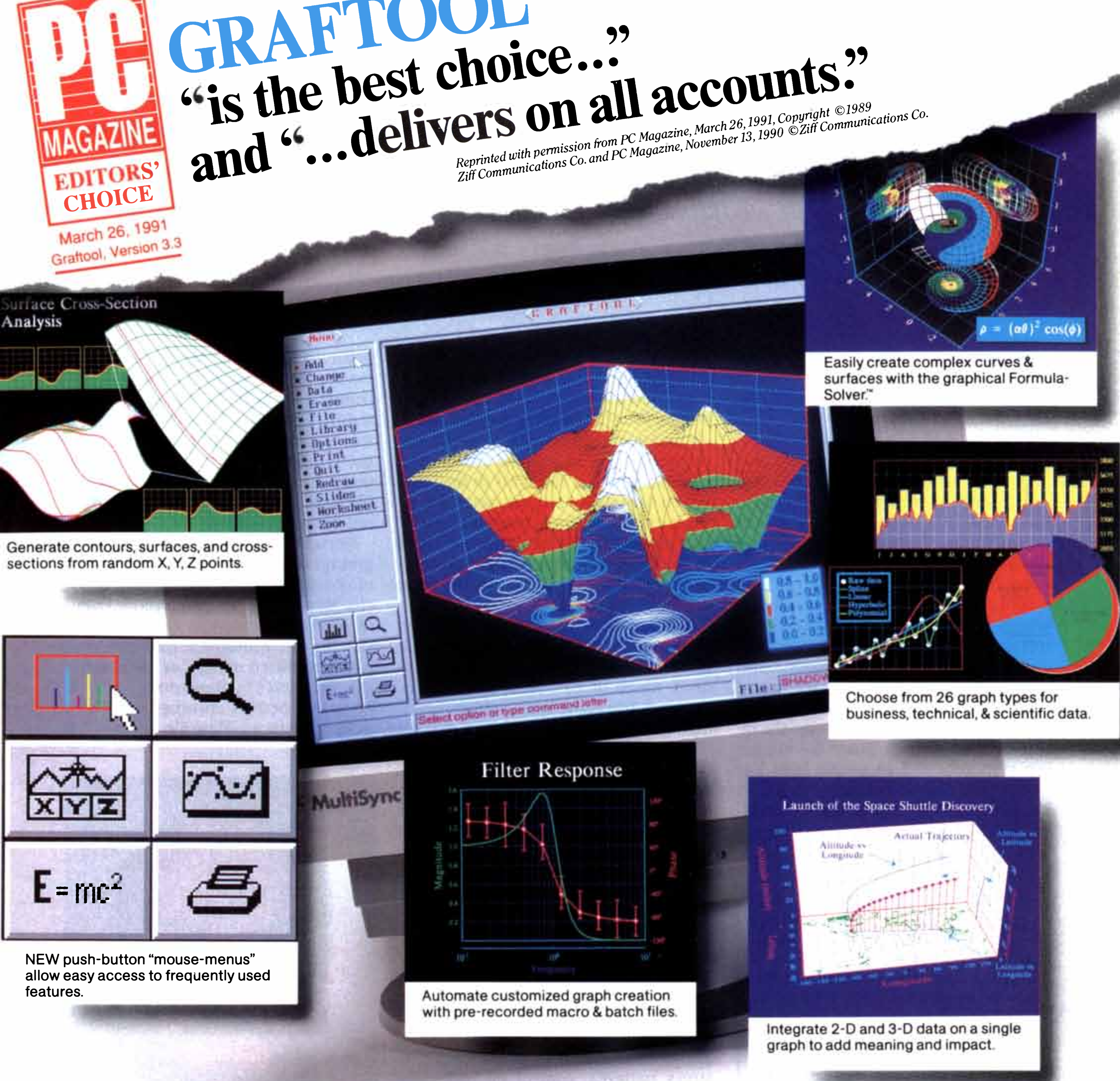
FURTHER READING

SPIDERS: WEBS, BEHAVIOR AND EVOLUTION. Edited by William A. Shear. Stanford University Press, 1986.

THE STRUCTURE AND PROPERTIES OF SPIDER SILK. John M. Gosline, M. Edwin Demont and Mark W. Denny in *Endeavour,* New Series, Vol. 10, pages 37–43; January 1986.

MODULATION OF THE MECHANICAL PROPERTIES OF SPIDER SILK BY COATING WITH WATER. Fritz Vollrath and Donald T. Edmonds in *Nature,* Vol. 340, No. 6231, pages 305–307; July 27, 1989.

GLYCOPROTEIN GLUE INSIDE THE SPIDER WEB'S AQUEOUS COAT. F. Vollrath and E. Tillinghast in *Naturwissenschaften,* Vol. 78, pages 557–559; December 1991.



Presentation graphics and analysis for scientific users.

Your technical data requires more muscle and sophistication than basic business graphics and plotting packages can provide. It's simply a matter of using the right tool for the job. In addition to publication-quality graphics, you need powerful analysis tools and capabilities such as Graftool's ■ Intelligent Data Cursor™ to read out data points on curves & surfaces ■ Linear & non-linear curve-fitting ■ Unlimited zoom & rotation ■ Multiple axes in linear, log, or probability scales ■ Advanced data handling, allowing over 268,000,000 data points ■ Powerful scientific spreadsheet which can directly read your Lotus or ASCII files ■ Unrestricted placement of graphs & text ■ Direct compatibility with Microsoft Word & WordPerfect.

With Graftool, all this power and flexibility purr quietly under the hood, while pop-up menus and push-buttons bring an ease of use previously unheard of in scientific software. Just "point & process" with increased productivity and greater understanding.

GRAFTOOL — the right tool for your technical solutions.

GRAFTOOL $495.

- ■ Interactive demo available
- ■ Academic discounts.

3-D VISIONS

Call Toll Free for Details

1-800-729-4723

2780 Skypark Drive, Torrance, CA 90505

Infrared Video Cameras

Similar to home video cameras, these devices can "see" thermal radiation. This emerging technology offers a host of applications, including night vision, visualization of heat flow and medical analysis

by Jerry Silverman, Jonathan M. Mooney and Freeman D. Shepherd

Warmth, warmth, more warmth! for we are dying of cold and not of darkness. It is not the night that kills, but the frost.

The association of light with heat and of darkness with cold is a familiar metaphor, as the above quote from the Spanish writer Miguel de Unamuno shows. It turns out that the association has a basis in fact. A warm object emits more radiation than a cool one does. Moreover, it radiates largely visible light, whereas a cool object gives off largely infrared rays.

The difference in the amount and type of radiation emitted by warm and cool objects is the foundation of an important emerging technology: the infrared video camera. Since the human eye responds well to visible light but poorly to infrared radiation, much of the information about an object encoded in its infrared radiation is lost. An infrared video camera is able to capture this information and convert it into images that can be seen by a human. One mature infrared video technology is based on platinum silicide. A platinum-silicide camera is similar in many respects to a home video camera, except that it responds to the long-wavelength radiation characteristic of the infrared, ranging from one to 5.7 microns (millionths of a meter), instead of the short-wavelength radiation characteristic of visible light, ranging from 0.4 to 0.7 micron.

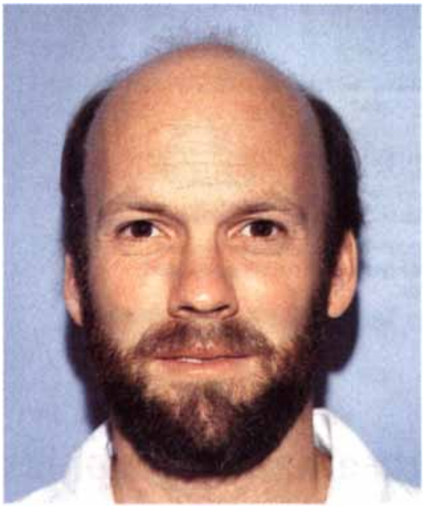
Infrared video cameras have obvious applications in night vision, surveillance, targeting and identification. Indeed, much of the initial impetus for the work has come from military funding. It appears likely, however, that infrared video cameras will open a host of other applications. Some civilian applications will undoubtedly spring from the military ones, such as foul-weather and night-landing aids for commercial aircraft and nighttime surveillance devices. Infrared video cameras may also help control laser machining and visualize heat flow in fluids.

But perhaps the greatest asset the new cameras offer is their ability to complement what a visible image shows. Older infrared technologies do this in the electrical, construction and petrochemical industries. They have also been used to help detect leaks and stress patterns, control oil pollution, fight fires and conduct meteorologic studies, land surveys and medical analyses. Although older infrared imaging techniques are useful, platinum-silicide cameras offer a low-cost, video-quality alternative. They have even been used to peer into distant regions of the universe otherwise obscured by interstellar dust.

Discovery of infrared radiation dates to the early 19th century, when the English astronomer Sir William Herschel raised questions about the basic similarity of light and heat. Herschel demonstrated the existence of infrared radiation by using a prism to split sunlight into its spectral bands. As he moved a thermometer through the bands, he found that the temperature rose from the blue end of the spectrum to the red. Furthermore, the temperature continued to rise beyond the red, where there was no visible light. The invisible light beyond the red became known as the infrared.

Herschel's discovery was put on firm ground by one of the founding fathers of quantum theory, the German physicist Max Planck. He proposed a theory of blackbody radiation that characterizes the electromagnetic radiation emitted by idealized objects, or blackbodies, as a function of temperature. (A blackbody is a perfect absorber of all incident radiation.) Planck's theory pre-

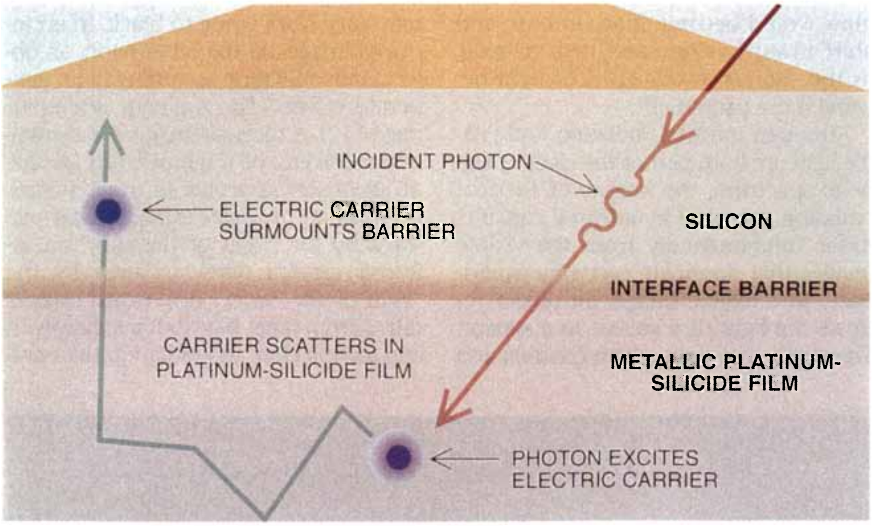
JERRY SILVERMAN, JONATHAN M. MOONEY and FREEMAN D. SHEPHERD are at the U.S. Air Force Rome Laboratory at Hanscom Air Force Base in Massachusetts. Silverman, who is in the infrared surveillance branch of the laboratory, earned his bachelor's degree from Brooklyn College in 1956 and his Ph.D. (in physical chemistry) from the Massachusetts Institute of Technology four years later. He focuses his research on the device physics and image processing of Schottky-barrier infrared cameras. Mooney received his bachelor's degree in physics from the University of California, San Diego, in 1981 and his Ph.D. in optical sciences from the University of Arizona in 1986. He works in the laboratory's focal plane array branch, where he studies the effects of mosaic detector arrays on the quality of infrared images. Shepherd is the Air Force's senior scientist for infrared sensors and arrays. He received bachelor's and master's degrees in electrical engineering from M.I.T. in 1959 and a Ph.D. from Northeastern University in 1965.



FACES take on an eerie aspect when viewed in the infrared. One of us (Mooney) is seen here in visible light (*above*) and infrared (*right*). Familiar characteristics, such as eye and hair color, are absent from the infrared image, whereas the thermal features are pronounced. The nose and ears are usually cooler than the rest of the face, and the eye sockets are warmer. The appearance of the nostrils and mouth will vary, depending on whether the subject is breathing in or out.

dicts that the radiation emitted by a blackbody at room temperature peaks at wavelengths near 10 microns and is negligible at wavelengths shorter than two microns. To relate this to everyday experience, imagine turning on an electric stove. One feels radiative heat before seeing any change in the color of the heating coil. As the temperature rises, the coil begins to glow red, corresponding to the short-wavelength portion of the emitted radiation. If one could continue to increase the heat, the glow would become more intense and shift toward yellow and then to blue, as the shorter wavelengths of light became more prominent.

Although infrared radiation and visible light are both part of the electromagnetic spectrum, the images of thermal emission captured by infrared cameras differ fundamentally from the visible images that dominate everyday experience. In a visible image, an object reflects the light of a source to a sensor. The image may have a high contrast and may vary from white to black. In an infrared image, on the other hand, an object emits radiation according to its temperature [see "The Amateur Scientist," page 112]. A thermal image is essentially a recording of a temperature profile of an observed scene. In other words, whereas an ordinary camera takes pictures by the "light of the sun," an infrared camera takes pictures by the "light of the earth." Real-world objects differ from ideal blackbodies, however, and always emit less radiation than pre-



PLATINUM-SILICIDE INFRARED DETECTOR is seen edge-on in this representation. It consists of a thin film of platinum silicide deposited on a silicon substrate. The absorption of incident infrared photons, or quanta of radiation, in the platinum-silicide film excites electric carriers. Some fraction of these carriers escapes into the silicon, forming a measurable current. The optimal thickness for a film of platinum silicide is approximately three nanometers.

dicted. In addition, one often finds the simultaneous presence of emitted and reflected radiation. Both of these effects tend to complicate the relation between the object brightness seen in a thermal image and the true object temperature.

Infrared imagery confounds expectations and differs most dramatically from the familiar world of visible experience at night or in indoor scenes where sunlight is absent. Faces, for example, remain recognizable but project an eeriness reminiscent of the ghost of Banquo: "Thou hast no speculation in those eyes which thou dost glare with." In most situations, however, the objects in infrared scenes have approximately the same temperature and radiate with nearly equal intensity. Any thermal image taken of the scene is dominated by the background radiation component corresponding to the average temperature. Consequently, it is often difficult to distinguish objects, which is to say that most thermal scenes have a very low contrast. Lack of uniformity in the detector-to-detector response characteristics obscures low-contrast infrared imagery even more than a similar problem does in the visible part of the spectrum. The desire to highlight the contrast as much as possible has been a major goal in infrared camera technology.

Early infrared cameras simply responded to the heat of infrared radiation. The Evaporograph, developed in 1929, imaged infrared radiation onto an oil-coated membrane. The oil evaporated at a rate that depended locally on the amount of incident infrared radiation. The thermal image was seen as a pattern of color variations arising from optical interference in the thin oil film, an effect that is responsible for the colors seen in oil films on water. Later, other heat-sensing cameras were developed that had electric readouts, but the devices had limited application since they were fragile and slow.

Faster response and greater efficiency have been obtained by detecting infrared photons, or quanta of radiation, directly. Detectors based on this principle are called quantum detectors, in contrast to so-called thermal detectors, such as the Evaporograph. During World War II, German forces used quantum infrared sensors to track ships in the English Channel and tanks on the eastern front. Both Germany and the U.K. began development of airborne infrared sensors for detecting bombers, but the war ended before either side could produce a useful working system, and the U.K. terminated its program in favor of airborne radar, which has better weather penetration capabilities. Early U.S. infrared imagers included the Sniperscope and the Snooperscope. Both used quantum detectors called 1P25 tubes, which were developed in the 1930s by RCA. Infrared radiation incident on a cathode, or negatively charged electrode, in the tube caused electrons to be ejected into a vacuum. The 1P25 was unable to detect photons emitted by room-temperature objects, however, because of the unavailability of metal cathodes sensitive to low-energy photons.

For a 20-year period after World War II, many attempts were made to extend television camera technology into the infrared. Limited success was obtained in the near infrared (one to three microns), but attempts in the spectral range beyond three microns failed completely as a direct result of the low contrast of infrared scenes. Designers then turned to so-called scanning systems,

INFRARED IMAGES of the clock tower of the Boston Customs House and the surrounding area are shown for a sunny day (*left*) and a clear night (*right*). The dominant visible sources, especially the lights inside the buildings, are absent, which helps to account for the remarkable overall similarity of the two images. In contrast to day-versus-night photography done in visible light, the f-stop and shutter

in which infrared images are created with a limited number of detectors in combination with moving mirrors. Scanning systems work quite well, but their complexity, high cost and extensive maintenance requirements limit their range of application.

In 1973 Andrew C. Yang of what is now called the Air Force Rome Laboratory and one of us (Shepherd) proposed an infrared sensor technology based on a detection process similar to that used in the 1P25 tube but with one crucial difference: instead of being emitted into a vacuum, the ejected electrons would be emitted into silicon. The process is called internal photoemission. Underlying the proposal was the observation that certain metals such as platinum react with silicon to form a metal-silicide cathode that has both the correct spectral response and spatial uniformity required for high-performance thermal imaging.

PLATINUM-SILICIDE INFRARED CAMERA is shown imaging the technician who wired the device. The thermal imprint left by touching the hand to the face is seen in the infrared image displayed on the monitor. Next to the monitor is the electronic control box used to focus and adjust the image. The silvered bulk of the camera is the dewar that cools the array of 39,040 detectors (not shown). Both the array and the readout circuitry fit on a single chip about the size of a dime.

Infrared cameras based on internal photoemission were first demonstrated in the mid-1970s. The technology has evolved steadily since then. Today's arrays have better resolution than standard television cameras and can image differences in temperature smaller than 0.02 degree Celsius. Platinum-silicide cameras are commercially available from manufacturers in the U.S. and abroad.

Lying at the heart of each platinum-silicide camera is an array of individual detector cells made from a thin film of platinum deposited in vacuum on a silicon substrate. Heating the wafer and film to a high temperature converts the platinum to platinum silicide. The resulting interface between the metallic, or conducting, platinum silicide and the semiconducting silicon is called a Schottky barrier diode (named after the German physicist Walter Schottky). The "barrier" is an uphill potential at the interface that in the dark and at low temperatures allows few electric carriers to move from the metal into the semiconductor. Infrared photons, if absorbed in the metal, can excite some of the carriers to an energy above the barrier. The process of photoemission of carriers over the Schottky barrier is analogous to kicking balls over the ridge of a hill. The fraction of carriers that escapes constitutes a measurable current; the overall rate of success of the process depends on many factors, including the thickness and structure of the film, the wavelength of the radiation and the height of the barrier.

speed for both of the infrared images on these two pages were the same (f/2.0, 1/30 second). These infrared images and the others in this article, unless otherwise specified, were taken by William Ewing with an infrared camera designed in the authors' laboratory at Hanscom Air Force Base. It has a 160-by-244 array of detectors made at the David Sarnoff Laboratories.

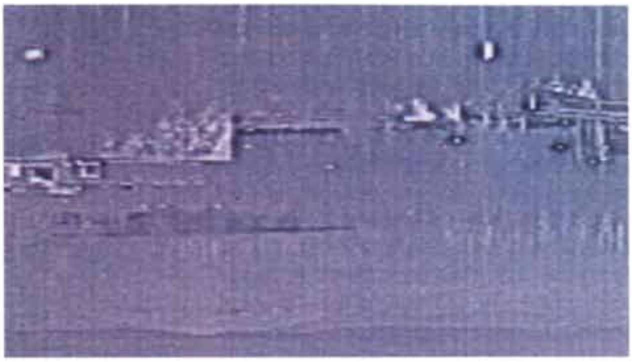
In an actual camera, a large number of Schottky diodes, each approximately 25 microns on a side, are arrayed in a two-dimensional grid on a silicon chip. The chip includes an electronic network from which the current from each diode may be read. To display a video image, one converts the electrical signal from each diode into an illumination level on a monitor. The diode array and readout network are made using standard silicon processing techniques, which results in low-cost, high-quality products.

The advantages offered by platinum silicide are tempered somewhat by two drawbacks. First, the device must be cooled to the temperature of liquid nitrogen (–196 degrees C) to operate. Second, the diode converts infrared radiation into electric current at a rather low efficiency, less than 1 percent. The requirement of low-temperature operation was at one time thought to be a serious problem; recent advances in cryo-

Working with Infrared Images

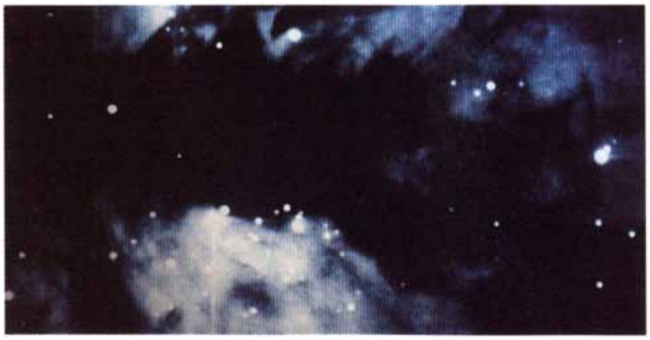
The conversion of raw signals from infrared detectors into display images can be accomplished in numerous ways. Four of them were used to make the aircraft pictures on the opposite page. The first picture was made with histogram equalization, which is superior at accentuating the detail of the background. In this approach, display space is assigned in proportion to the number of pixels (individual detectors) having a given signal level. In the second picture, the technique is histogram projection, which brings out small objects against the background. Display space is based only on the presence or absence of pixels at a given level. The display techniques in the last two pictures accentuate the thermal effects of the structural supports of the aircraft by sacrificing sensitivity to large-scale details for an increased sensitivity to local variations. The algorithm used for the third picture maps the raw data cyclically; the one used for the final image filters out low spatial frequencies (large-scale details).

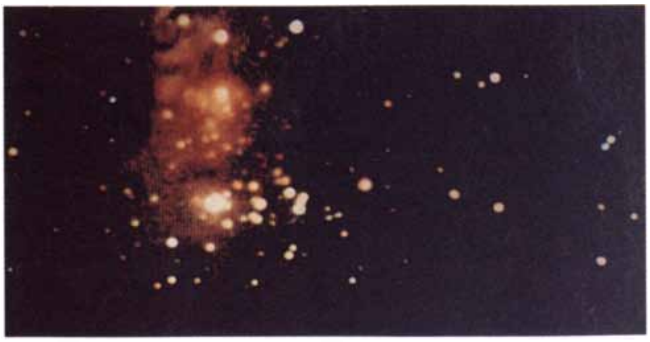
The two photographs below show housing at Hanscom Air Force Base in Massachusetts on a winter night. In an exposure made at $^1/_{30}$ second (*left*), temporal noise dominates. The noise results largely from the randomness inherent in the arrival and detection of infrared photons. An eight-second exposure of the scene (*right*), processed by the technique used for the final aircraft picture, reduces the effect of the temporal noise but reveals the underlying residual spatial noise. Spatial noise results from nonuniform responses by individual detectors to the same level of incoming radiation. The residual spatial noise reflects an imperfect correction made by subtracting the detector array's long-term response to a uniform scene from the uncorrected image.



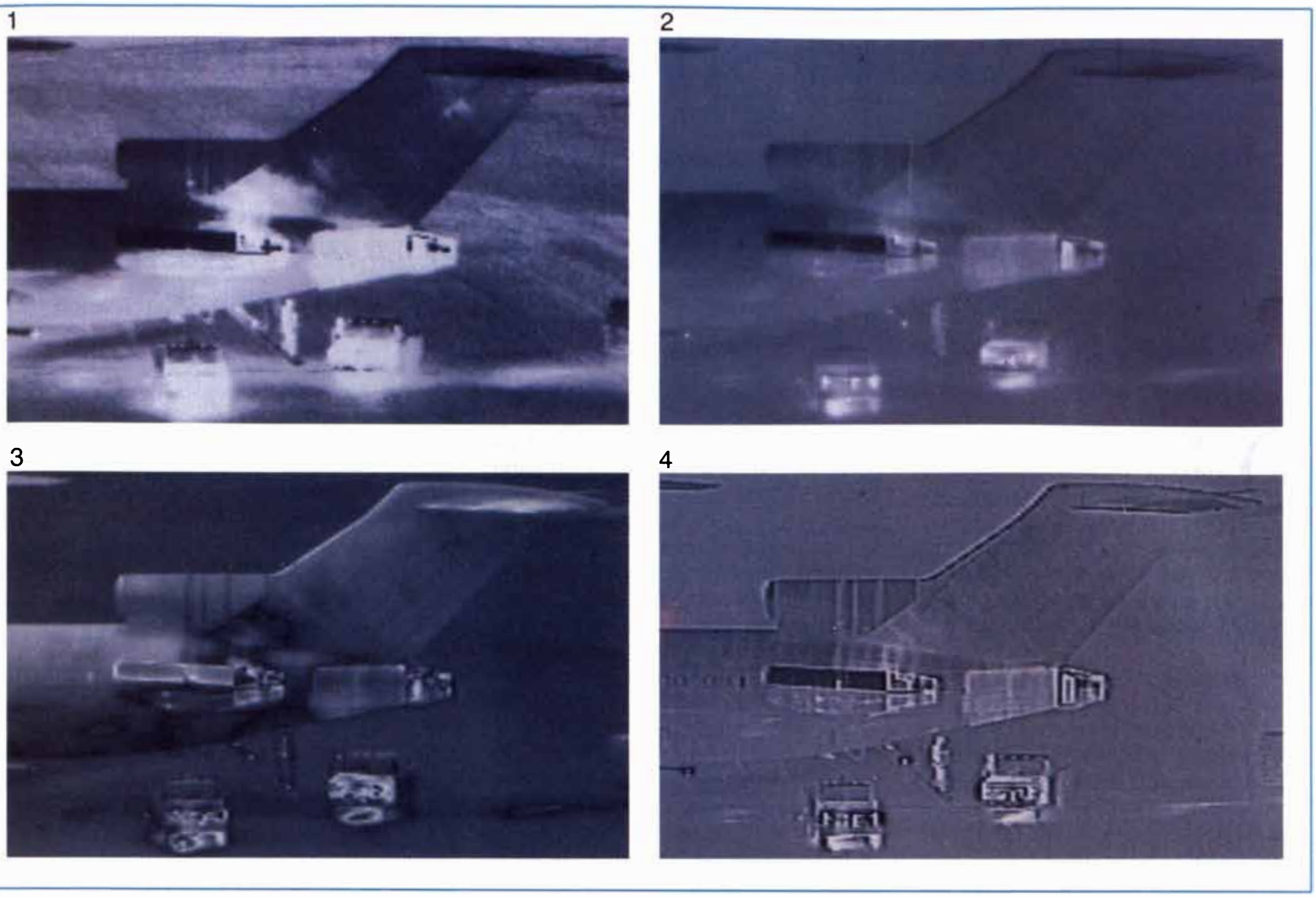
genic refrigerators have reduced the impact of this issue, however. The inefficiency of the device, which at first led many workers to dismiss Schottky diodes as infrared detectors, is partially overcome by the fact that each diode "stares" at the same small segment of the scene. Nevertheless, the use of thousands of detectors does place a premium on the uniformity of response from detector to detector on the array. A comparison of platinum silicide with alternative modern technologies reveals a roughly inverse relation between efficiency and uniformity. As a result, the match of technology to application is a subject of debate.

One of the main challenges in designing infrared video cameras lies in converting the electrical signals from the diodes into illumination levels. For visible images, a direct linear mapping from raw signal to display scale usually produces an image that matches the expectation of the viewer. In displaying infrared images, one encounters several problems. The observer's expectations are based on experience with visible imagery, which does not always translate well into infrared images. And the signal in a visible image can vary by as much as a factor of 1,000 from the darkest to the brightest extreme, whereas the signal from a typical infrared scene varies by about a factor of only two. The low contrast of infrared images magnifies





INTERSTELLAR DUST obscuring NGC 2024 (*left*), a region of star formation located adjacent to Orion's belt and the Horsehead Nebula, is penetrated by infrared radiation (*right*) to yield a quite different view of the region. The colors in the infrared image are produced by displaying the images at three different wavelengths (1.2, 1.6 and 2.2 microns) as blue, green and red components of a computer-generated image. The infrared image, which is presented through the courtesy of Ian Gatley, Albert M. Fowler and K. Michael Merrill, was taken at the Kitt Peak National Observatory's 1.3-meter telescope using a 256-by-256 platinum-silicide array made at the Hughes Aircraft Company.



the impact of noise sources that would not otherwise be significant. There are several ways to convert raw signals into display images.

Another important challenge in building infrared cameras is the minimization of noise, or unwanted fluctuations. (For electrical engineers, noise joins death and taxes in a triad of inevitabilities.) The noise comes in two distinct varieties: temporal and spatial. Temporal noise imprints the image with a grainy "salt and pepper" appearance. Optimally, it arises from the randomness inherent in the arrival and detection of infrared photons. It turns out that the magnitude of such noise increases as the square root of the measurement time, whereas the magnitude of the signal increases linearly. Consequently, an increase in the measurement time improves the signal-to-noise ratio in proportion to the square root of the measurement time, thereby reducing the temporal noise.

Even with the reduction of temporal noise, one is still left with the second problem: spatial noise. It arises from nonuniform response of detectors to uniform incoming infrared rays. Spatial noise thus represents the "fingerprint" of individual detector differences in the display. Several computer-based techniques can reduce the effect, although none as yet is completely satisfactory. In the simplest approach, the pattern observed when the camera is subjected to uniform illumination is subtracted from subsequent imagery.

The effect of spatial noise on the performance of an infrared camera is analogous to driving in fog with a dirty windshield. Array nonuniformities correspond to the dirt on the windshield; low-contrast infrared images correspond to objects seen through the fog. Just as the dirty windshield degrades the driver's ability to discern details in the fog, so detector nonuniformities degrade the performance of infrared cameras. A great advantage of the Schottky arrays is that the intrinsic nonuniformity pattern is small enough and stable enough to permit real-time correction.

Future work on infrared video cameras is likely to center on improving current technology by extending the camera's response to even longer infrared wavelengths. The ideal approach to that end depends on one's ability to fabricate arrays with high efficiency and high inherent uniformity. The performance of these ideal arrays would be fixed not by art but by nature, namely, the random arrival rate of infrared photons. Although such arrays are now as common as the unicorn, today's dreams might well turn into tomorrow's reality.

FURTHER READING

SILICIDE INFRARED CAMERA TECHNOLOGY. F. D. Shepherd, Jr., in *Proceedings of SPIE: Advanced Infrared Sensor Technology*, Vol. 395, pages 114–118; April 18–19, 1983.

THE THEORY OF HOT-ELECTRON PHOTOEMISSION IN SCHOTTKY-BARRIER IR DETECTORS. J. M. Mooney and J. Silverman in *IEEE Transactions on Electron Devices*, Vol. ED-32, No. 1, pages 33–39; January 1985.

PTSI INTERNAL PHOTOEMISSION: THEORY AND EXPERIMENT. J. M. Mooney, J. Silverman and M. M. Weeks in *Proceedings of SPIE: Infrared Sensors and Sensor Fusion*, Vol. 782, pages 99–107; May 19–21, 1987.

SILICIDE INFRARED STARING SENSORS. F. D. Shepherd in *Proceedings of SPIE: Infrared Detectors and Arrays*, Vol. 930, pages 2–10; April 6–7, 1988.

DISPLAY OF WIDE DYNAMIC RANGE INFRARED IMAGES FROM PTSI SCHOTTKY BARRIER CAMERAS. J. Silverman, J. M. Mooney and V. E. Vickers in *Optical Engineering*, Vol. 29, No. 2, pages 97–104; February 1990.

The Gundestrup Cauldron

An enigmatic bowl found in Denmark 100 years ago has been variously attributed to the Celts, Germans and Thracians. Recent analysis challenges such judgments and the theories on which they are based

by Timothy Taylor

On May 28, 1891, peat cutters working near the hamlet of Gundestrup in Northern Jutland spied a curious metal object. When they rubbed away the peat, they found a great silver bowl containing plates, some concave and some convex, that had once fitted back-to-back to form the sides of a magnificent cauldron. The plates were vividly decorated with scenes of war and sacrifice: bearded deities wrestled ferocious beasts, a bare-breasted goddess stood flanked by elephants, and a commanding figure in stag's antlers brandished a ram-headed snake in one hand and a twisted neck collar in the other.

The cauldron was placed in the National Museum in Copenhagen, and the Danish government paid out a handsome reward to the finders, who quarreled bitterly over it. But their dispute was soon overshadowed by an academic argument that has continued ever since. When and how was the cauldron made, by whom and at whose behest? How did it come to be in the bog? What do its pictures mean? Finally, how does it fit with our picture of prehistoric society?

I shall try to answer these questions while reviewing developments in scholarly interpretation of the cauldron. The century of controversy has reflected both the advent of new technical approaches and changes in how archaeologists reconstruct prehistory. In many ways, the Gundestrup cauldron is a touchstone of archaeological method.

TIMOTHY TAYLOR is a lecturer in the department of archaeological sciences, University of Bradford, and assistant editor of the journal *Antiquity*. He read archaeology and anthropology at St. John's College, University of Cambridge, graduating in 1982, and then carried out research in Bulgaria and at Oxford. In 1988 he returned to Cambridge as a junior research fellow at King's College. He moved to Bradford in 1990.

Solutions to some of the problems posed by the cauldron are now at hand. There is broad consensus for dating its manufacture to the later Iron Age—the second and first centuries B.C. Recent scanning electron microscopy by Erling Benner Larsen of the School of Conservation of the Royal Academy of Fine Arts in Copenhagen has shown that the cauldron was the work of five virtuoso silversmiths. My own analysis of the pictorial representations shows that although the cauldron displays features usually considered to be Celtic and although it was found in a part of Europe considered Germanic at this date, the silversmiths probably lived in Thrace, in southeastern Europe.

Deeper issues are also coming into focus. Analysis of the Gundestrup cauldron challenges received wisdom on life in Iron Age Europe; attempting to understand the artifact in its context consequently helps to refine the theories archaeologists use to reconstruct prehistoric society. Ethnicity, gender and mythology may all have been more complex than previously supposed. Firm cultural boundaries may not have existed, humanity and its gods may have been viewed as having more than simply male and female genders, and religious beliefs may have been flexible and multifaceted.

Paleobotanical analysis carried out at the time of the cauldron's discovery determined that it had been left on dry ground. Wetter conditions, beginning around the time of Christ, had caused peat to grow around the vessel, covering it in a layer three feet thick. Thus, far from being a votive bog offering (as many have assumed), the bowl seems to have been hidden in the rough grass beyond the immediate circuit of nearest settlement. Somebody clearly meant to come back for it.

Although the cauldron was not found in association with more easily datable objects, such as shards of pottery, it does carry representations of helmets, spurs and other objects that did not come into use much before 150 B.C. By the time of its abandonment, however, the cauldron was already old, damaged and missing one of its eight original outer plates. Furthermore, as Larsen has shown, the circular base plate was originally a horse's bridle decoration; only

SILVER TREASURE from a Danish bog appears in one of several plausible orderings of the bowl and the plates that surmount it (*right*). The plates were found disassembled (*far left*), making the reconstruction of their pictorial narrative difficult. The bowl was beaten from a single ingot, a technically demanding feat. Later repairs show far less skillful workmanship: a top view (*left*) shows how a hole in the base was plugged by soldering on a phalera, or silver bridle decoration.

later was it soldered to the bottom of the bowl to repair a hole. The cauldron was therefore probably lost at some time in the first century B.C., between 50 and 150 years after it had been made. It was repaired, and possibly dismantled and reassembled, more than once during the intervening period.

The original craftsmanship greatly excelled the repair work. The bowl, 27 inches (70 centimeters) across, was beaten from a single ingot—a challenging task. The plates were worked in the repoussé manner: the shapes were first beaten from a sheet with frequent annealing, and then the raised surfaces, supported from behind, were decorated with pattern punches and other tools.

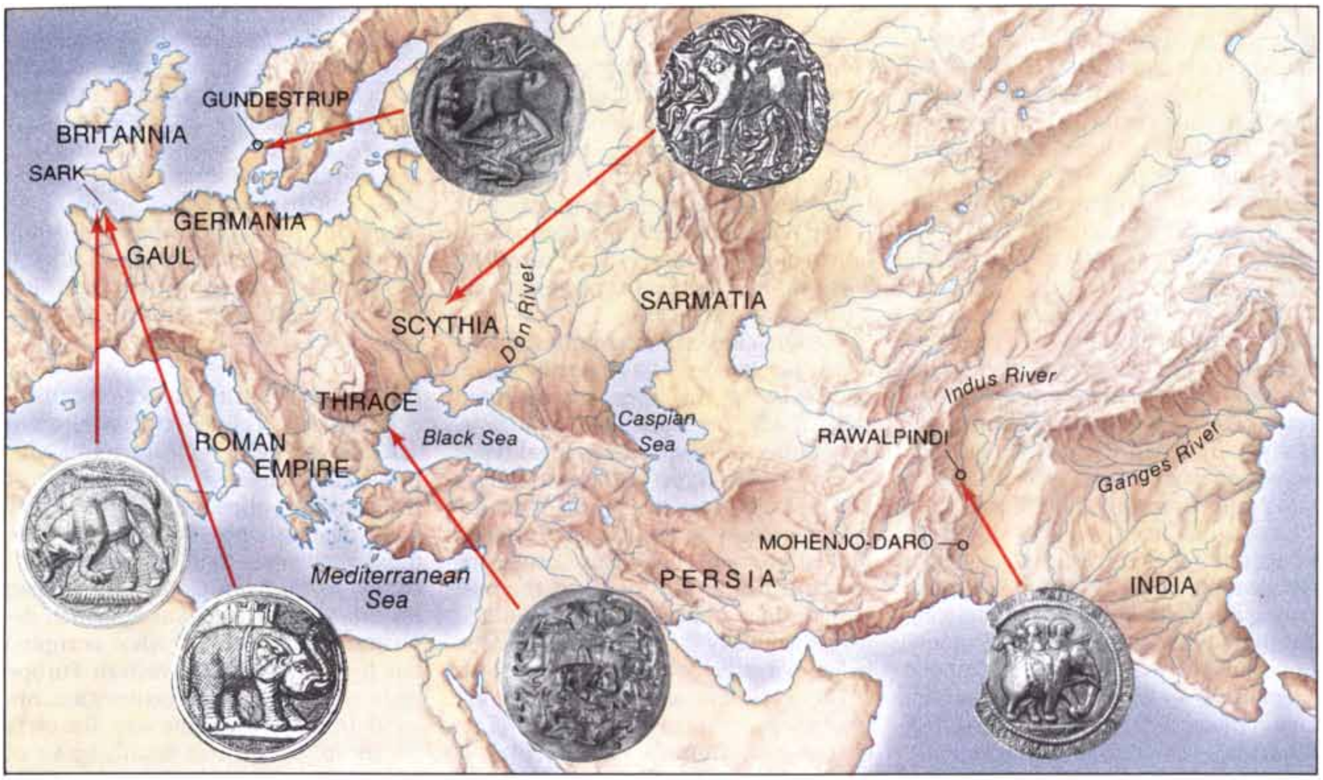
Larsen has used the tool marks to identify the makers. He made silicone rubber casts of the cauldron's surface, plating the casts with gold and imaging the gold under a scanning electron microscope. The pictures gave diagnostic "signatures" of five sets of tools, which Larsen plausibly ascribed to five different artists. His analysis independently corroborates stylistic interpretations.

Numbering the silversmiths is easier than placing them. Their nonliterate society is known only from partial accounts by classical authors and from the problematic reconstructions of archaeologists. I say problematic because such reconstructions are based on an oversimplification of culture that stems from the German prehistorian Gustav Kossinna. In the 1920s he defined an archaeological culture as a recurring association of artifact types found exclusively in one region.

Kossinna linked such assemblages to a "people" possessing a specific spirit, or *Völksgeist*. This approach mirrored that of the pre-Darwinian biologists, who understood species as ideal types or essences. And just as the essentialists had ranked species in a "chain of being," with humanity at the top, so Kossinna ranked peoples. The Nordic German, or "Aryan," people was believed the best by Kossinna and others who identified with it—notably the Nazis, who used the belief to justify their policy of conquest.

The weakness of Kossinna's theory, or ideology, is apparent. By equating archaeological cultures with peoples, it implies an impossibly perfect coincidence of anthropological variables. In reality these variables rarely coincide neatly. A nation, although governed by a dominant group, typically contains several religions, languages and physical types; moreover, art and technology often cut across national boundaries. Yet the notion of "pure peoples" persists, albeit in a form cleansed of its more extreme racist overtones. Many archaeologists remain cultural integralists, tacitly identifying "cultures" and "culture groups" with "tribes" and "peoples," respectively.

The integralists have often considered the Gundestrup cauldron to be of Celtic culture and Gallic provenance. Indeed, photographs of the cauldron appear in almost every book dealing with the Celts. Yet one must remember that the word "Celt" comes to us from Greek and Roman historians, who used it as a blanket description for people who occupied much of central and western Europe from at least the sixth century B.C. onward. In much the same way, the early Europeans in America usually spoke of "Red Indians" and only occasionally noted more specific formations; so did the ancient historians speak generally



EURASIAN SILVERWORK suggests an Iron Age tradition that transcended cultural boundaries, such as those generally assigned to Germans, Celts, Thracians, Scythians and Indians. These phalerae show striking commonalities: butting bulls, for example, appear on the phalera from Gundestrup (*top row, left*) and on another from Sark Island (*bottom row, left*).

of "Celts" and sporadically of "tribes."

Most modern Celticists are integralists: they generally represent the Celts as a unitary people of great antiquity, distinctive for their language, physical type, religion, institutions and material culture. These workers regard some of the body armor and equipment worn by the smaller figures on the cauldron—such as twisted neck rings (or torcs), carnyx war trumpets, animal-crested helmets and long shields with circular bosses—as diagnostic of Celtic culture.

These scholars also read the Celtic pantheon into the pictures on the cauldron, identifying, for instance, the large figure wearing stag's antlers as "the horned one," or Cernunnos. A figure holding a wheel has been more tentatively identified as the Celtic wheel god, Taranis. Beyond this there is little consensus, and the artifact's rich iconography has fueled many flights of mythological fancy. Some Celticists have explained the curious elephants, for example, as referring to Hannibal's crossing of the Alps.

The fundamental weakness of the standard Celtic argument is the absence in Gaul (now France)—and northwestern Europe generally—of any tradition of repoussé work in sheet silver. To address this problem, Ferdinand Drexel of the University of Frankfurt argued in 1915 that the cauldron had originated in southeast Europe, the nearest source of the requisite silverworking skills. This technological constraint was the mainstay of Drexel's contention that the cauldron was Thracian in origin.

Drexel's argument got me interested in the cauldron when I was studying Thracian art (itself an integralist term for cultural assemblages of southeastern Europe), a field I had entered at the suggestion of John Alexander, my undergraduate tutor at the University of Cambridge. While visiting museums in Romania and Bulgaria, I met Anders Bergquist, a Swedish graduate student who was also fascinated by the cauldron, and we began to collaborate. We built on Drexel's work by refining the dating and extending the range of cultural materials to which we compared the cauldron, considering not only technique but also formal and iconographic attributes. Notable among these were the depictions of shoelaces on the cauldron [*see illustration at left on page 88*] and on two other pieces of well-dated Thracian silverwork.

Such similarities led us to conclude that the cauldron had been made either in Transylvania or the adjacent lower Danube basin. Both regions were in a state of flux in the late Iron Age. Historical sources talk of "Celts" raiding southeastern Europe just before the cauldron was made, sacking Greek Delphi in 279 B.C. and establishing the kingdom of Tylis in what is today Bulgaria. Other sources speak of raids by German war parties after the cauldron was made. We therefore reasoned that southeastern Europe contained a multiethnic society and that the Thracians, the Celts and the Germans had each played a role in the cauldron's story.

We conjectured that an intrusive Celtic tribe called the Scordisci had commissioned the cauldron from native Thracian silversmiths and that Germanic looters had then carried it off. The looters we tentatively identified as the Cimbri, freebooters who classical authors say came south from the lower Elbe region to attack the Scordisci in 118 B.C. After their defeat by the Romans five years later, some of the Cimbri retreated north to settle in Himmerland (thought to be etymologically related to "Cimbri"), the part of Jutland where the cauldron was dug out. Still, we did not depend on the Cimbri for our explanation of the cauldron's transportation. Cultural connections known purely through archaeology can be demonstrated between the Black Sea and the Baltic at this time.

Bergquist and I at first refined existing arguments without challenging the fundamentals of the reigning integralist model, although we allowed more than one "pure" people on stage at a time. Neither did we then provide any detailed explanation of the pictures on the cauldron beyond the customary reference to the horned figure as Cernunnos. The elephants had never posed a problem to scholars of the Thracian School, as the animals were in use in Asia Minor at this time and pictures of them appeared on coins.

The next advance began with a false step. We suggested that an analysis of the lead isotopes in the cauldron's silver plates might identify the mine from which the ore had been extracted. It turns out, however, that such identification is impractical because coinages from different mines were routinely remelted to make metal vessels. But this possibility opened up a new line of analysis: if the cauldron was made from melted coins, its weight might be a clear multiple of a standard coin weight.

I put this idea to the test in Copenhagen in May 1991, while attending the centennial conference of the cauldron's discovery. I made accurate measurements of the weight of individual pieces, compensated for what was missing and submitted my postulated original weights to Michael Vickers, a specialist in metrology at the Ashmolean Museum of the University of Oxford. Reconstructing an original total weight of 9,445 grams and a weight of 4,255 grams for the hemispheric bowl alone, Vickers was able to find exact dividers based on the Persian *siglos,* a coin whose top weight is 5.67 grams. (Multiplying this value by 1,666 and 750 produces 9,445 and 4,255, respectively.)

This finding is consistent with a Thracian origin of the cauldron. Persian precious metalwork had a major formative influence on the development of silversmithing in Thrace in the fifth and fourth centuries B.C., and Persian weights continued in use long after.

The cauldron was not the only piece of silverwork to be carried out of Thrace. The bridle decoration, or phalera, that was used to repair the great bowl is one of many that found their way to the west, in most cases probably borne on the horses of Thracian mercenaries serving in the Roman army. Of special importance are the phalerae found on the Channel Island of Sark in 1718, now lost but preserved in detailed drawings. The "butting bull" phalera [*see illustration on opposite page*] resembles the Gundestrup base plate—as well as several phalerae from the lower Don basin, then controlled by a group called the Sarmatians. The "elephant and castle" phalera is similar to two from a North Pakistani site near Rawalpindi.

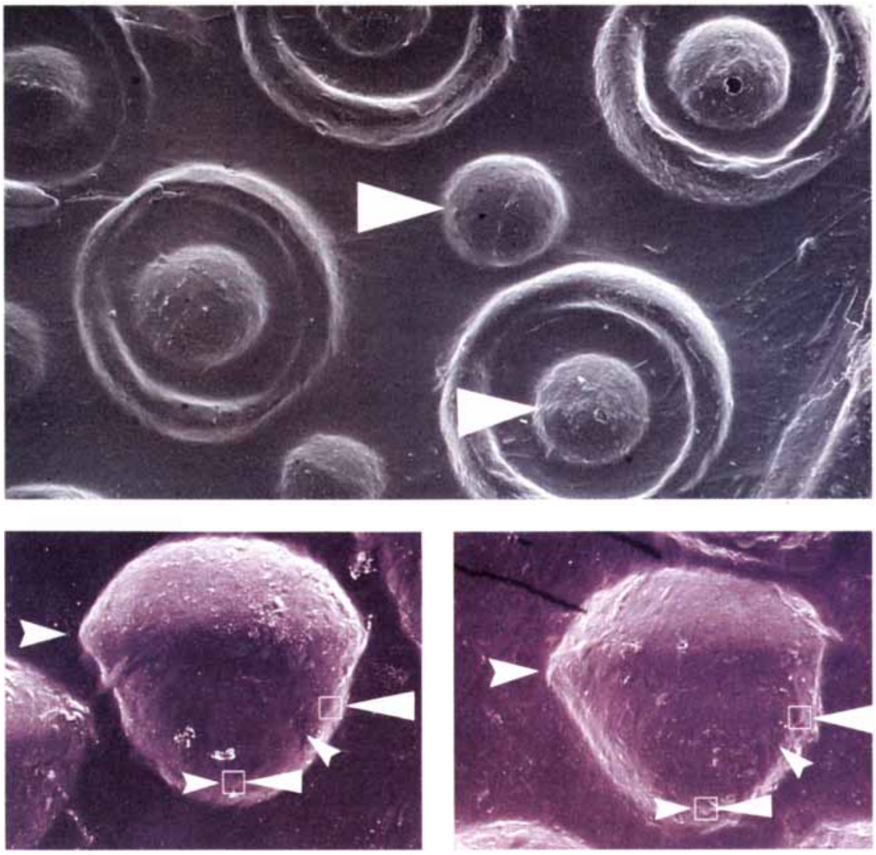
A third phalera from Rawalpindi depicts a woman with her hair in tresses and a bird perched above each shoulder. The woman can be identified as Hariti, the protectress of children, clear versions of whom appear on Sarmatian metalwork from the Don basin. That region had also been heavily influenced by northern India, a connection noticed in the 1930s by the great Russian historian Mikhail Rostovtseff (later a professor at Princeton University).

I had already noted a figure with exactly these attributes—a similar goddess with paired birds and a child at her breast—on a phalera from northeast Bulgaria as well as on the Gundestrup cauldron. When I saw the link to Sarmatia and, ultimately, to India, I was stunned. This connection promised to explain the presence of elephants on one of the cauldron's other plates. The obvious associations of a piece of silverwork depicting elephants had been staring everyone in the face for a century! Following up the Hariti connection, I saw that the Gundestrup scene showing a pair of elephants flanking a central female figure clearly depicted the lustration, or ritual bathing, of Lakshmi, the Hindu goddess of good fortune.

India had in fact been mooted as the cauldron's place of origin soon after its discovery. In 1895 Japetus Steenstrup of the University of Copenhagen proposed an origin in medieval northern India, Tibet or Mongolia. But Steenstrup's intuition outran his capacity for analysis: he presented a mass of vague and unsystematic comparisons without explanation. In retrospect, one can see that those who dismissed his conclusions threw the baby out with the bathwater.

What can account for the similarities among artifacts from such far-flung sites, as well as for the systematic variations found in them? The similarities must derive from networks of artisans that crossed ethnic and political divides. Silversmiths, for instance, had contacts stretching 4,000 miles, from northern India to the Balkans. That network can be inferred from surviving artifacts, scanty though they are.

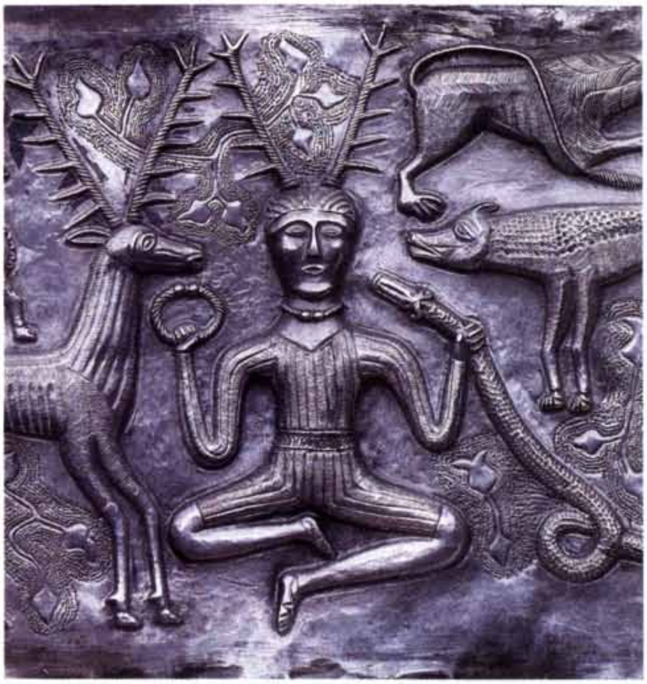
Variations in iconography reflect the inaccuracies that accumulated as artisans copied the work of a master or a rival. As in a game of "gossip" ("Chinese whispers"), the message changes each time it is relayed. A good example is



TWIN PUNCH MARKS on the cauldron were struck by one tool, from similar angles and with similar force, as shown optically (*top*) and in an electron micrograph (*bottom*). Microanalysis (*arrows*) of such marks led Erling Benner Larsen of the Royal Academy of Fine Arts in Copenhagen to attribute the bowl to five silversmiths.

Was Culture Transmitted from India to Denmark?

A shared pictorial and technical tradition stretched from India to Thrace, where the cauldron was made, and thence to Denmark. Yogic rituals, for example, can be inferred from the poses of an antler-bearing man on the cauldron (*below, left*) and of an ox-headed figure on a seal impress from the Indian city of Mohenjo-Daro (*right*). The cauldron's Thracian provenance is attested in such details as the man's zigzag shoelaces. Three other Indian links appear under Gundestrup figures on the opposite page, from left to right: ritual baths of goddesses with elephants (the Indian goddess is Lakshmi); wheel gods (the Indian is Vishnu); and goddesses with braided hair and paired birds (the Indian is Hariti). Although some scholars label the Gundestrup deities Celtic, they were also pancultural.



provided by the bizarre pointed shoulders of the Gundestrup elephants, which are clearly copied not from the animal itself but from a copy of a copy of one. Even more telling, the ears of the Gundestrup elephants sit oddly high on their foreheads, evidently because they derive, at many removes, from decorative hats depicted in Indian models. Nevertheless, the structure of the highly stereotyped scene remains intact.

The existence of related representations across a wide area and over long periods need not imply that the original meaning was preserved. An earlier manifestation of Hariti, protectress of children, was Hariti, the ogress who ate children. The meaning of the figure we have identified as a type of Hariti on the cauldron may also have undergone several transformations.

Nerys Patterson of the School of Celtic Languages and Literatures at Harvard University has pointed out to me the striking similarity between Hariti and Rhiannon—a figure in the Mabinogion stories, written down in Welsh (a Celtic language) in the 12th century A.D. Rhiannon also makes the switch from predator to protectress and is accompanied by a pair of birds as well. It is unclear whether this similarity results from coincidence, a survival of the Hariti type as first transmitted 1,000 years earlier on the cauldron and related silverwork or a later and independent transmission.

Implicit in this model of cultural transmission through transnational networks is the notion of artisans as belonging to an itinerant class—perhaps one of many in the ancient world. We can divine some aspects of their social relations with their clients from the pictures on the cauldron and on other artifacts. The beardless figure on the side of the wheel opposite the bearded ruler may, for example, be a ritual specialist [*see illustration on opposite page*]. Indeed, he may belong to the same group, guild or caste as the five silversmiths, for metalworking was itself an important ritual occupation.

A useful analogy here is provided by the modern speakers of Romany called Gypsies, who have often performed services deemed ritually dangerous by the settled societies in which they move. These services include making music at funerals, telling fortunes and working metal. Just as the Gypsies have generally gone unmentioned in written histories, so our archaeologically identified groups may easily have escaped the attention of classical authors. Even though they lived in southeastern Europe, the Gundestrup silversmiths may not have considered themselves Thracian or anything else we could now name.

They might instead have resembled the Enarees of Scythia, now southern Russia, as they were described in the fifth century B.C. by the Greek writers Herodotus and Hippocrates. Biologically male but dressed as women, the Enarees interpreted omens and settled disputes for the Scythian aristocracy. Such specialists are attested across Eurasia in the Iron Age, not just the shamans of Scythia and the yogis of India, but the seers of Thrace, the druids of Gaul and, a few centuries later, the bards of Ireland. In Ireland the biologically male bard who praised the king in song was described as female, in opposition to the ruler's maleness. Ethnically, both bards and silversmiths were set apart from the settled population, holding a status akin to an Indian caste.

The most famous image on the cauldron, the stag-antlered Cernunnos, fits into this picture quite remarkably. In 1959 Heinz Mode of Martin Luther University in Halle noted the striking resemblance between the cauldron figure and one on a seal stone from the Indus valley site of Mohenjo-Daro, dated to the second millennium B.C. Both figures sit in apparently yogic poses, but Mode, believing the cauldron to be essentially Celtic, argued that poses from so early a date could not really be yogic and were instead cross-cultural.

But Thomas McEvilley of Rice University has recently pointed out that although the Mohenjo-Daro figure has male genitalia, it is dressed in the female fashion of the day. He also notes that its yogic posture resembles one still practiced in rural India by low-caste sorcerers—yogis who make their living selling charms, making tattoos, handling snakes and working spells. Moreover, the posture is intended to channel sexual energy. McEvilley's observations throw a very different light on the similarity between the Mohenjo-Daro and Gundestrup figures.

Within the pictorial grammar of the cauldron, on which males are shown bearded and breastless and females beardless and breasted, the horned figure is of ambiguous gender. Further, it has one heel pressed against the perineum—in a pose related to that of the Mohenjo-Daro example and probably adopted with the same energy-channeling aim in mind. The figure is partly levitated, balancing on one toe, and it has a round torc, a ram-headed snake and antlers. These attributes make sense in terms of a shamanic extension of power over three usually separate domains: female, male and animal.

The Gundestrup artists were widely separated in space and time from those of Mohenjo-Daro and worked in different media and with different technologies. Yet they produced similar images because they operated under similar ritual constraints. Both constituted castes that moved outside settled societies, and both provided a range of ritual services to those requiring them. Those rituals were evidently based on a magical tradition common in Eurasia, surviving today in tantric yoga and Siberian shamanism.

Such castes appear, therefore, to have played a crucial role in knitting together the culture of the ancient world. This insight alone may prove more important, in the end, than recovering the original meaning of the pictorial narrative on the cauldron. The problem itself, like certain mathematical conundrums, may prove more valuable than its solution.

FURTHER READING

AN ARCHAEOLOGY OF YOGA. Thomas McEvilley in *Res,* Vol. 1, pages 44–77; Spring 1981.

SEM-IDENTIFICATION AND DOCUMENTATION OF TOOL MARKS AND SURFACE TEXTURES ON THE GUNDESTRUP CAULDRON. Erling Benner Larsen in *Recent Advances in the Conservation and Analysis of Artifacts.* Compiled by James Black. University of London Institute of Archaeology, Summer Schools Press, 1987.

THE ORIGIN OF THE GUNDESTRUP CAULDRON. Anders Bergquist and Timothy Taylor in *Antiquity,* Vol. 61, pages 10–24; March 1987.

GUNDESTRUPKEDELEN: BAGGRUND OG BILLEDVERDEN. Fleming Kaul. Nyt Nordisk Forlag Arnold Busck, 1991.

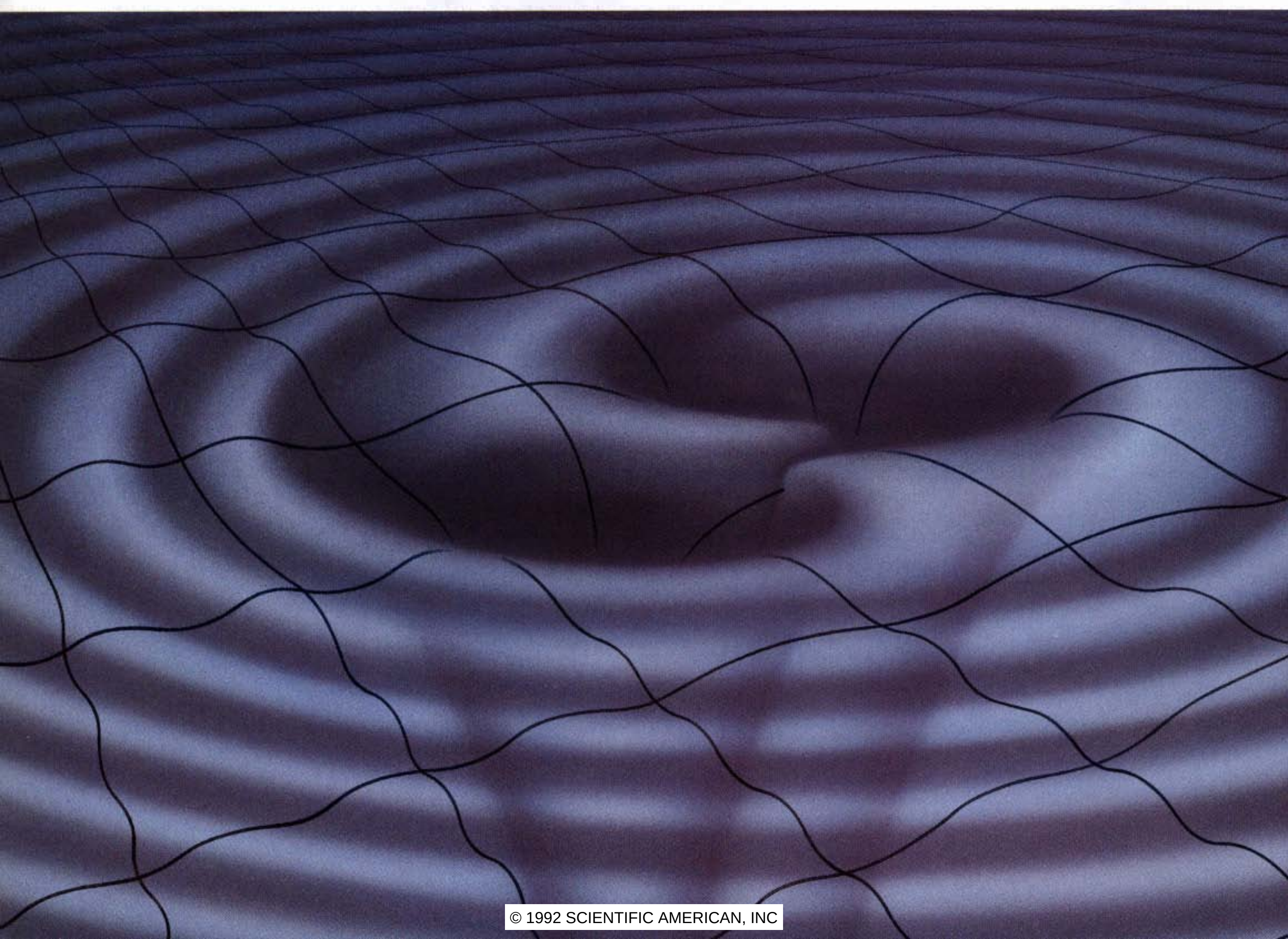
CATCHING THE WAVE

by Russell Ruthen, *staff writer*

Space is not just the nothing between the earth and the stars. Nor is it simply the void between the electron and the atomic nucleus. It is a ubiquitous medium more resilient than rubber, more rigid than steel.

Nearly 75 years ago Albert Einstein realized that space takes its shape from the mass it contains. Our sun is too meager to warp space much. But a black hole is the ultimate space bender. A small black hole holds three times more mass than our sun in a sphere a million billion times smaller in volume. Were two black holes to collide, they would curve, knot and twist space in ways that theorists have only begun to imagine. These cataclysmic collisions would provide the ultimate test for Einstein's most sophisticated ideas, his theory of general relativity.

After decades of research—and many partial attempts—physicists are ready to construct an unusual telescope that might finally allow them to observe such space-warping events. But unlike any other telescope, their instrument is not sensitive to electromagnetic radiation: not light, radio waves, gamma rays or any part of the spectrum. It is designed, instead, to detect gravitational waves, small changes in the shape of space.

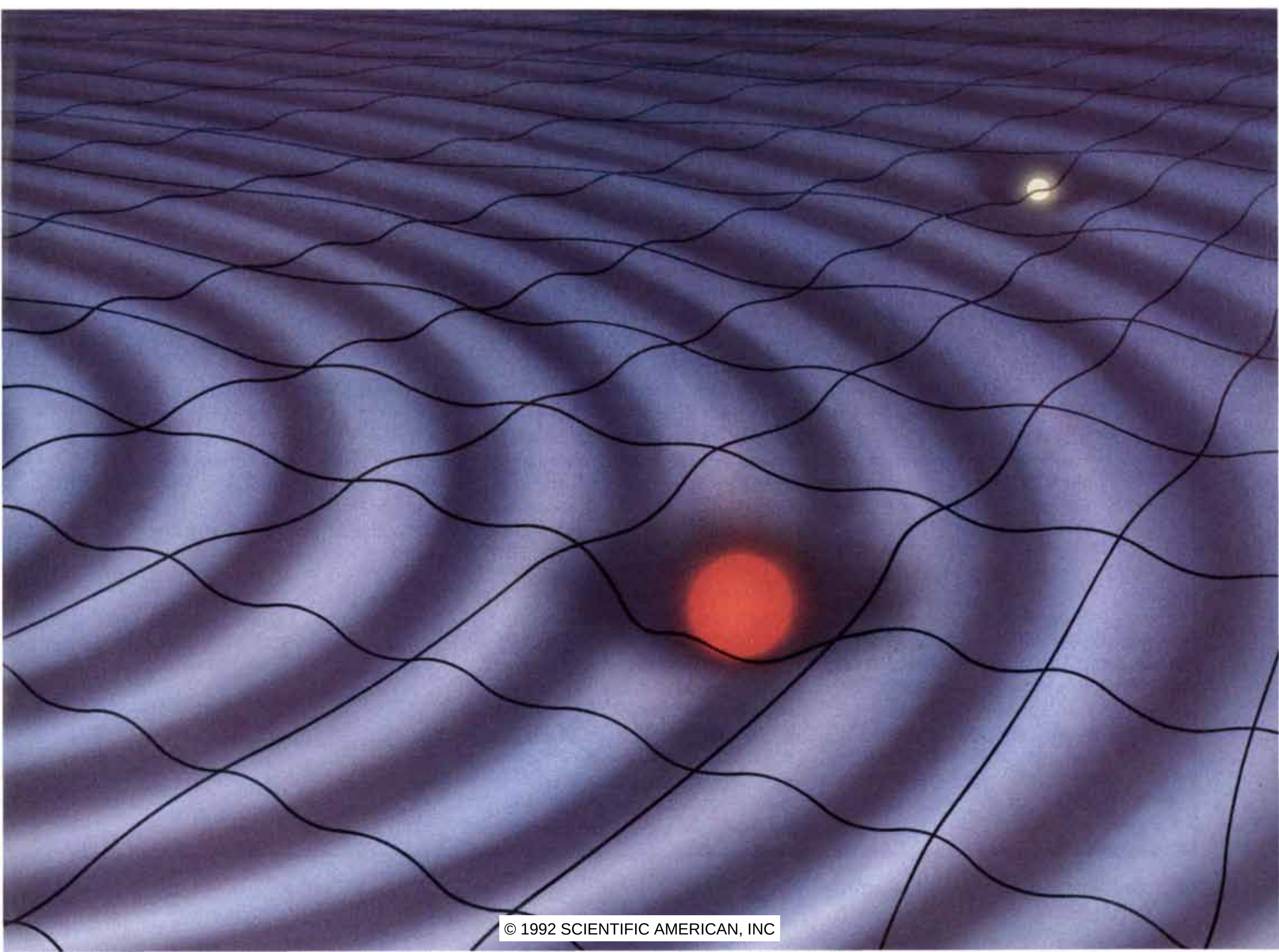


Scientists hope to explore the nature of gravity and the frontiers of the universe by detecting gravitational waves. The U.S. government has promised them $211 million to build two kilometer-size detectors. Will they succeed?

Astronomers have seen the effects of gravitational waves, but no instrument built so far has been sensitive enough to detect the waves directly. According to Einstein's theory, gravitational waves propagate outward from their source like ripples traveling across a pond. As the waves expand into space, they weaken. But gravitational waves, unlike electromagnetic radiation, will not be blocked by stars or cosmic debris or—when they eventually reach it—the earth.

With a sensitive detector, it should therefore be possible to observe gravitational waves as they compress and expand both space and matter by an infinitesimal amount. For example, if the gravitational waves from two colliding black holes in a distant galaxy were to pass through a detector one kilometer long on the earth, the length of the detector would change by less than a billionth of a

TWO BLACK HOLES produce gravitational waves as they spiral around each other. Here space is represented as a strong, elastic sheet, like the fabric of a trampoline. Any object with mass, such as the stars below, will exert a strain on space, which will distort the fabric. Each black hole stretches the fabric to infinity. As the black holes move around each other, some of the distortions travel outward as gravitational waves.



billionth of a meter. That's a distance about 1,000 times smaller than the diameter of an atomic nucleus.

A team of physicists from the California Institute of Technology and the Massachusetts Institute of Technology headed by Rochus E. Vogt hopes to catch the first wave—and more. During the next five years, his group plans to erect two facilities on opposite sides of the U.S. that will house the world's largest interferometers. Each device will generate intense laser beams that will bounce back and forth along two paths four kilometers long and interfere with each other at a point. If a gravitational wave of sufficient strength passes through the device, the distance that the light beams must travel will change slightly, altering the way the light beams interfere.

A bill appropriating the first funds for construction was signed by President George Bush in October 1991. Known as the Laser Interferometer Gravitational-Wave Observatory (LIGO), the project is expected to cost a total of $211 million. If the LIGO team achieves its goals and current predictions are trustworthy, the interferometers should be sensitive enough to observe the gravitational waves emitted from a pair of colliding neutron stars. If so, the project may produce the first direct confirmation of the existence of gravitational waves by the end of the decade. But, most important, it will permit researchers to peer into the universe in a way that differs radically from any previous observation.

By detecting gravitational waves instead of electromagnetic radiation, researchers should be able to detect both bright objects such as exploding stars and dark objects such as black holes. If luck is with them, they may discover unknown celestial bodies or even gravitational waves emitted at the moment of creation. "I believe profoundly," Vogt exclaims, "that LIGO will become most famous not for neutron star binaries and black holes and everything but for the things we can't even think of yet."

Many scientists do not share the optimism of the LIGO team, and the need for the project is hotly debated. Although theorists can calculate the strength of the gravitational waves that would emerge from two orbiting black holes, they do not know how many such systems exist. Neutron star binaries are the only source of gravitational waves for which both the strength and number can be predicted with confidence. No one can be sure, however, that LIGO will be sensitive enough to observe them.

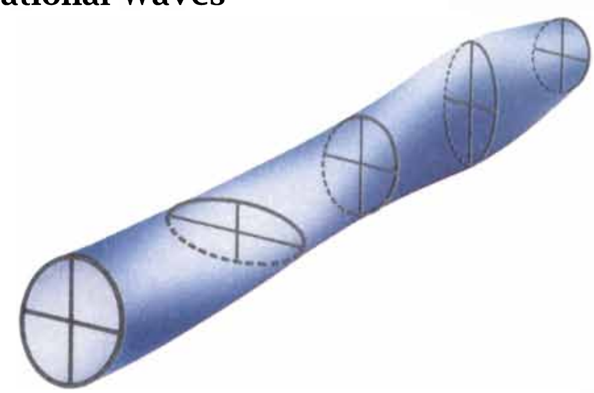
Einstein's Glitch?

Then again, no one was certain that gravitational waves existed until recently. Even Einstein had doubts. In 1916 he proposed in his theory of relativity that gravitational waves, and indeed the force of gravity, are manifestations of the bending of space. For decades, theorists argued passionately over whether gravitational waves were real or were merely a glitch in Einstein's theory. The debate provoked Sir Arthur Eddington, whose observations confirmed general relativity, to comment that gravitational waves "propagate ... at the speed of thought!"

During the 1950s, theorists finally reached a consensus that gravity waves did indeed exist. In 1957 Joseph Weber, a physicist at the University of Maryland, set out to build the first gravita-

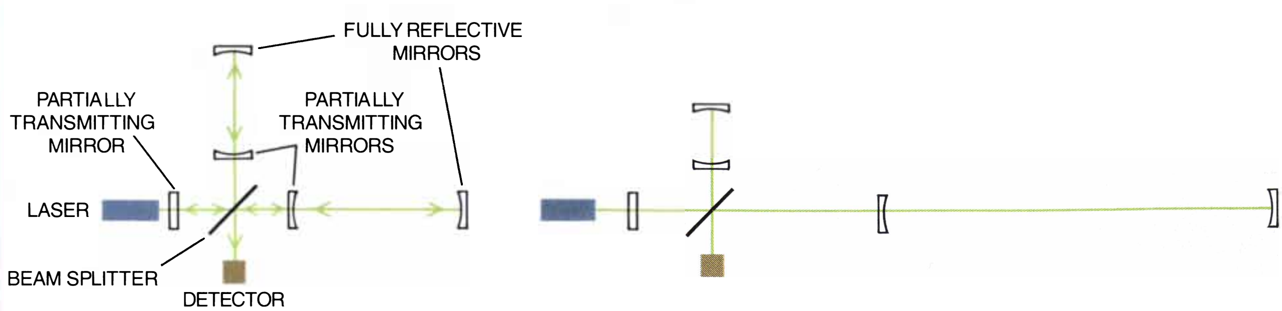
Detecting Gravitational Waves

It should be possible to observe gravitational waves because they shrink and stretch space and matter. Imagine, for example, a gravitational wave that passes through a long cylinder [*see diagram at right*]. As the wave travels from one end of the cylinder to the other, it does not distort the cylinder equally in all directions. The wave can compress matter in one direction while expanding it in a direction perpendicular to the compression.



The LIGO team hopes to detect gravitational waves using a sophisticated interferometer similar to the one depicted below. Laser light passes through the recycling mirror and hits a beam splitter. The light then enters one of two Fabry-Perot cavities, each of which consists of a partially transmitting mirror and a fully reflective mirror. The light bounces many times within the cavities and then leaks out through the partially transmitting mirrors. The cavities and the recycling mirror are included to store the light, boosting the power and sensitivity of the device.

The light leaked from each cavity hits the beam splitter and is recombined. Some of the recombined light is reflected back into the interferometer by the recycling mirror. The rest of the recombined light reaches the detector, which measures the intensity of light. If a strong gravitational wave passes through the detector, the compressions and expansions of the instrument (*bottom right*) will affect the way the leaked light recombines at the beam splitter. As a result, the detector records the passage of the gravitational wave as a change in light intensity.



tional-wave detector. Its key component was a cylindrical bar, weighing several tons, which was suspended from cables in a vacuum chamber and isolated as much as possible from outside vibration. After considering all sources of gravitational radiation known at the time, Weber determined that most of them would emit gravitational waves at a frequency around 1,000 cycles per second. He therefore engineered the size, shape and composition of the bar so that it would resonate like a tuning fork to waves of that frequency.

If a passing gravitational wave caused the bar to expand and then contract, a transducer would convert the tiny motions into electrical signals that could be measured. The strength of the wave would be related to the strain (the change in length divided by the length) induced in the bar. After constructing several prototypes, Weber settled on a two-meter-long bar detector that was capable of measuring strains of one part in 10^{16}, equal to a change in length of about two tenths of a millionth of a billionth of a meter.

During the past 35 years, Weber has reported several events that he claims are "evidence for gravitational waves." His most famous results, in 1969, appeared to show that gravitational waves were radiating from the center of our galaxy. Since Weber's report, many laboratories around the world have built sophisticated bar detectors to confirm his results, but none has revealed statistically significant evidence for gravitational waves.

Nevertheless, Weber's pioneering work inspired others to search for gravitational waves. Soon after Weber began publishing his results, Rainer Weiss, who is now one of the principal investigators for LIGO, began teaching a course in relativity at M.I.T. His students were excited by Weber's experiments, and Weiss searched for a way to explain the results in a simple way. It was then that he seized on the idea of using light beams to detect gravitational waves. (At the time, Weiss was unaware that Weber and several others had also thought about using light to detect gravitational waves.)

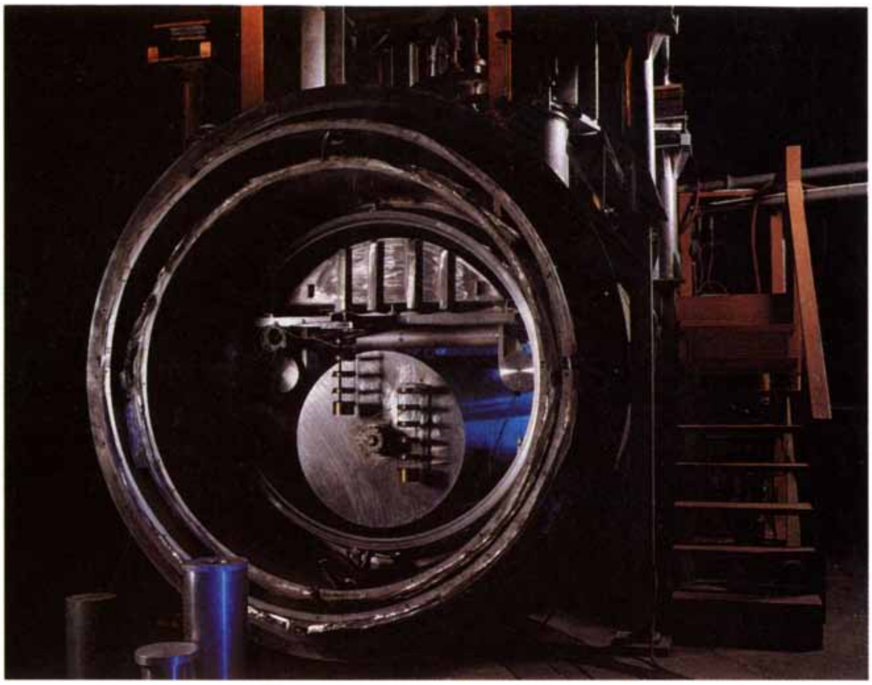
The first gravitational-wave detector that used light beams was built in 1971 by Robert L. Forward and his colleagues at Hughes Research Laboratories. It was based on the interferometer, which American physicist Albert A. Michelson had invented 90 years earlier to disprove the existence of the cosmic ether.

The modern form of Michelson's device consists of a laser, a beam splitter, two mirrors and a photodetector arranged to form a cross. The laser is placed at, say, the west end of the cross; the detector rests at the south end; the mirrors are positioned at the east and north ends; and the beam splitter sits in the middle. The laser beam first passes through the beam splitter, which redirects half the light toward the north mirror and guides half toward the east mirror. Both mirrors reflect the light back along its previous path to the splitter. There the beams are recombined and directed onto the detector.

What the detector measures depends on the distances between the beam splitter and the mirrors. At certain distances, when the light waves emerge from the splitter, headed for the detector, the crests of the light waves that came from the north are in sync with the crests of the light waves from the east. As a result, the light waves reinforce each other, increasing the intensity of the recombined beam that falls on the detector. But if the distances between the beam splitter and each mirror are then changed by half the length of the light waves, the crests of the light waves from the north emerge from the splitter at the same time as the troughs of the light waves from the east. The two cancel each other out, and no light reaches the detector.

For the purpose of detecting gravitational waves, the mirrors are positioned so the light waves cancel each other out. But should a gravitational wave pass through the interferometer, it will change the distances between the components slightly. As a result, some light will hit the detector, which will then record a change in intensity that will be proportional to the strength of the gravitational wave [*see box on opposite page*].

Most experimentalists believed interferometers should ultimately be more sensitive to gravitational waves than bar detectors are. An interferometer reacts at the speed of light, whereas the components of a bar detector respond at a speed equal to that of sound.

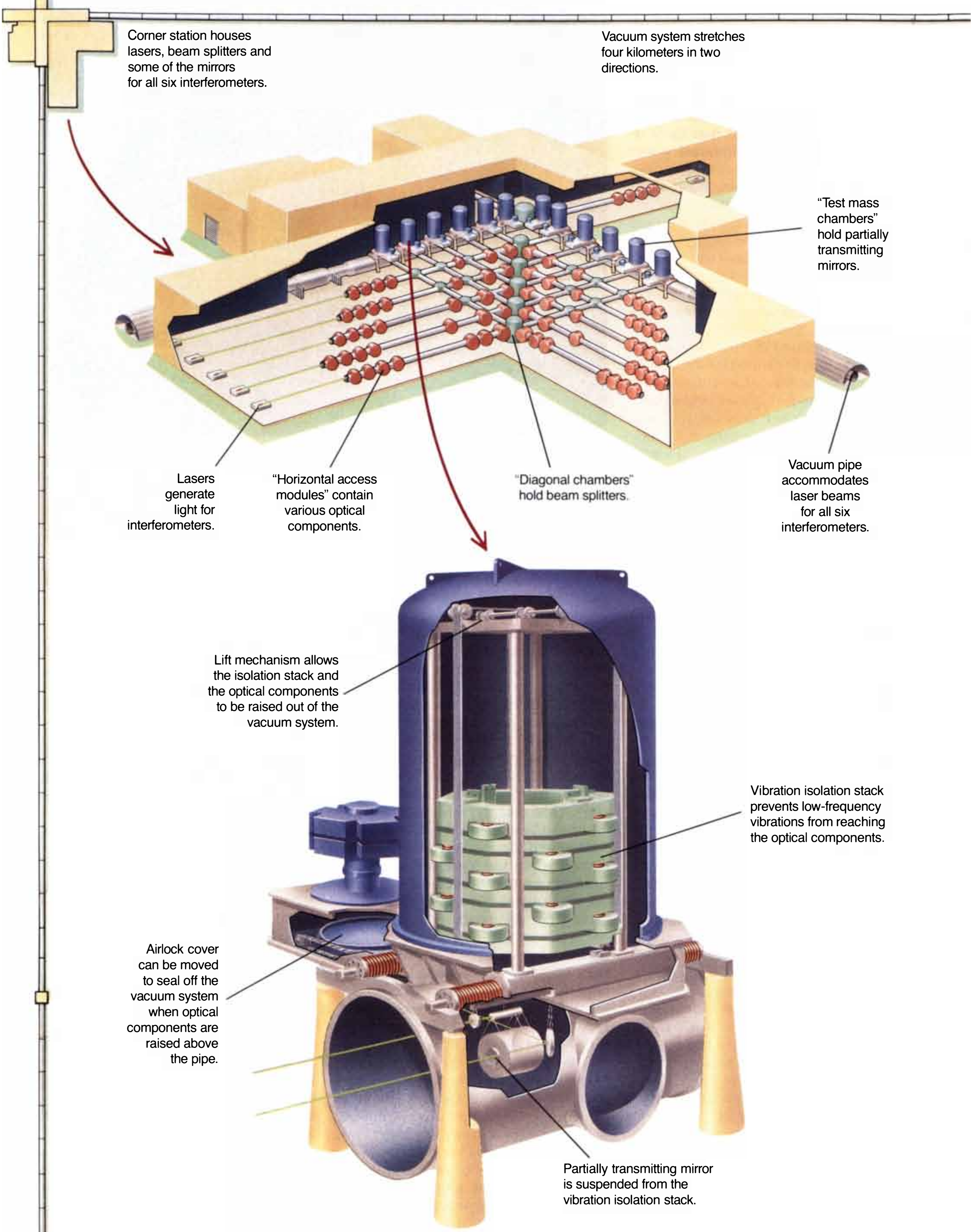


BAR DETECTOR was constructed to sense gravitational waves. If a strong wave were to pass through the detector, it would produce a strain on the bar—an aluminum cylinder suspended in a vacuum chamber. The strains and the resulting vibrations could then be detected. The apparatus, in a laboratory at Stanford University, cools the bar to four kelvins to reduce extraneous vibrations caused by heat.

An Astronomical Detector

Although those first interferometers were about 100 times more sensitive than Weber's bars, none has yet produced evidence of a gravitational wave. Ironically, the only convincing observation so far that gravitational waves exist came not from gravitational-wave detectors but from radio telescopes. In 1974 astronomers Joseph H. Taylor, Jr., and Russell A. Hulse, then at the University of Massachusetts at Amherst, found a "neutron" star known as PSR 1913+16, which has since provided strong, quantitative evidence for gravitational waves. Like other neutron stars, PSR 1913+16 has somewhat more mass than the sun, compressed into a sphere

The LIGO Facility



less than 10 kilometers in diameter. At that density, matter exists most comfortably as neutrons, hence the name.

PSR 1913+16 is a kind of neutron star called a pulsar. It has a very strong magnetic field that rotates with the star. The field accelerates charged particles in the vicinity of the star, generating beams of radiation that emerge from each of its magnetic poles. The beams spin around with the star, shining into space like a lighthouse beacon. By observing them, Taylor and Hulse discovered that PSR 1913+16 is rotating at the rate of about 16.9 times a second with a regularity that rivals that of an atomic clock.

What makes PSR 1913+16 even more unusual is that it orbits once around a companion star every eight hours. It reaches a top speed of 400,000 meters per second, only 750 times slower than the speed of light. In addition, the distance between the pulsar and its companion is 100 times less than the distance between the earth and the sun.

In essence, the pulsar is a precise clock orbiting rapidly through a severely warped region of space—the distortions caused by the great mass of the companion. These conditions are ideal for observing relativistic effects. Taylor explains that according to Einstein's theory of relativity, the orbital period should be "gradually decreasing as the system loses energy in the form of gravitational waves."

Since 1974 Taylor and his colleagues have been observing the decay of the orbital period. Their measurements agree with the predictions with an uncertainty of less than 0.5 percent. "Remarkably, we now have data that give us confirmation of this tiny effect that Einstein suspected would never be observed," Taylor remarks.

Unfortunately, the gravitational waves emitted by PSR 1913+16 are far too weak to be detected by Weber's bar, extant interferometers or even the proposed LIGO. Nevertheless, the discovery of PSR 1913+16 encouraged physicists to continue to think about what objects in the universe might produce sufficient gravitational radiation for earthbound instruments to detect.

One of the leading theorists of that group—and a key adviser to the LIGO team—is Kip S. Thorne. In the 1970s and 1980s he and several other theorists from around the world demonstrated that the universe should contain many different kinds of sources of gravitational waves. But they were frustrated in their attempts to make quantitative predictions. In all cases, they were missing at least one piece of important information that made it impossible to predict either the strength of the source or how many sources existed.

A decade ago, for example, theorists believed the most plausible source of gravitational waves that could be detected on the earth would be a supernova, the explosion of a massive star. They estimated that millions of supernovas occur throughout the universe every year. Those sheer numbers gave them the assurance that many stars would explode in relatively nearby galaxies.

But that alone did not guarantee detectability. Critics noted that astrophysicists simply do not know the detailed dynamics of supernovas and that the strength of gravitational waves produced by a supernova would depend on whether or not the collapse of the star was asymmetric. According to theory, a strong source of gravitational waves requires a massive, compact object that must be nonspherical, like a football, a cylinder or a barbell. Most important, the source must move rapidly in such a way as to accentuate the nonspherical component. For example, a football-shaped star that spins rapidly about its long axis produces no gravitational waves. But if the same star rotates end over end, it will be a strong emitter.

Caught Up in the Wave

Thorne, Bernard Schutz of the University of Cardiff and their colleagues have now identified neutron star binaries as the one type of potential source of gravitational waves whose strength can be unequivocally predicted from fundamental principles of physics and whose number can be estimated from astronomical observations. A neutron star binary is like a massive barbell rotating end over end. Over hundreds of millions of years, the two stars will spiral toward each other until they collide and merge. In the final moments before coalescence, the individual stars will be separated by a distance of about 20 kilometers and be moving at speeds comparable to that of light.

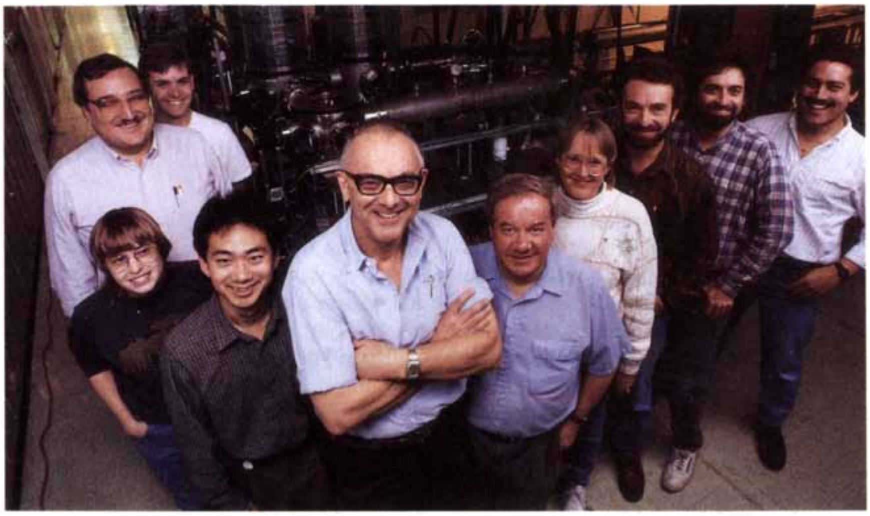
The gravitational waves from a neutron binary in a galaxy 650 million light-years away would produce a strain on the earth as large as four parts in 10^{22}. The frequency of the gravitational waves would equal twice the rate at which the neutron stars spiral around each other. A few minutes before the neutron stars coalesce, they should rotate about five times per second. Then, in their final throes, they should accelerate to more than 500 rotations per second. The frequency of the gravitational wave will therefore increase from 10 to 1,000 cycles per second.

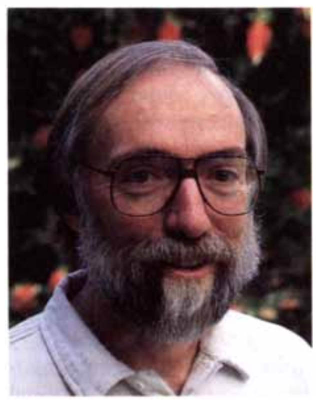
Neutron star binaries are rare in the universe and difficult to detect from their electromagnetic radiation. Astronomers have catalogued more than 400 neutron stars, but they have confirmed only four pairs of orbiting neutron stars in our galaxy. From these four, astrophysicists have figured out approximately how many neutron stars collide in the universe every year. Ramesh Narayan and his colleagues at the Harvard-Smithsonian Center for Astrophysics and E. Sterl Phinney of Caltech have independently estimated that a few pairs of neutron stars are likely to merge every year within about 650 million light-years of the earth.

In theory, then, if scientists construct LIGO so that it measures strains of four parts in 10^{22}, over a year they stand a good chance of detecting the gravitational waves from a few neutron star binaries. In practice, researchers would need to build at least two detectors at separate distant sites so that they could distinguish local disturbances from gravitational waves. They would have to erect another detector at a third site to determine the location of the source in the sky.

If LIGO snares a burst of gravitational waves from a neutron star binary before the turn of the decade, the event will make Thorne the winner of a bet he made with Jeremiah P. Ostriker of Princeton University in 1981. At stake is a case of "good red wine." Thorne contended that gravitational waves would be detected by the year 2000; his eminent colleague agreed that gravitational waves exist, but he believed Thorne had overestimated the strengths of astronomical sources of gravitational waves. Ostriker has not changed his mind. Does Thorne still expect to win? "I think it will be nip and tuck," he concedes.

Indeed, the history of LIGO has been one of fits and starts since Thorne touched off a contentious race with





***Top:* Members of the Caltech team, including LIGO director Rochus E. Vogt (arms folded in the center) and physicist Ronald W. P. Drever (to the right of Vogt). *Bottom left:* The M.I.T. group, including physicist Rainer Weiss (seated at the right). *Bottom right:* Caltech theorist Kip S. Thorne.**

Weiss's group at M.I.T. to be the first to detect gravitational waves with an interferometer. In 1979, a year after Taylor and Hulse presented the first convincing evidence for gravitational waves, Thorne convinced the physics faculty at Caltech to enter the field of gravitational-wave detection. To lead the effort, the Caltech physicists recruited Ronald W. P. Drever, who had worked on bar detectors and interferometers at the University of Glasgow.

In response to Caltech's newfound enthusiasm for detecting gravitational waves with interferometers, the National Science Foundation (NSF) stepped up its funding. It began supporting gravitational-wave research at both Caltech and M.I.T. at the level of about $1 million a year. Drever and Weiss began to compete to build more sensitive and sophisticated interferometers.

The sensitivity of the devices can be enhanced by boosting the power of the lasers and by increasing the distance between the mirrors and the beam splitter. But the sensitivity can be compromised by several noise sources that arise because of small changes in the frequency of the laser light or small vibrations that move the mirrors and other optical components. An increase in laser power and the length of the beam can reduce some kinds of noise sources while exacerbating others.

To increase the sensitivity of the interferometer, Weiss worked on devices in which laser light bounces between two mirrors many times along different paths. This system, known as an optical delay line, effectively increases the length of the interferometer. Meanwhile Drever developed a scheme that utilized "Fabry-Perot cavities." In that system, laser light bounces between two mirrors along the same path to boost the power of the interferometer.

Weiss experimented with an interferometer whose two L-shaped "arms"—the length between the mirrors and the beam splitter—were 1.5 meters long. Drever, meanwhile, built and operated a 40-meter interferometer.

With the Caltech group appearing to be taking the lead, Weiss decided in 1981 to "do something dramatic." To attract the attention of the NSF and draw funding to M.I.T., he began a detailed study of the design and costs of a kilometer-scale interferometer. But he also aroused the interest of Thorne and Drever, who wanted Caltech to remain at the forefront.

The NSF settled matters by merging the M.I.T and Caltech groups in a "shotgun wedding" that marked the official beginning of the LIGO project. To manage the research, the NSF set up a steering committee that consisted of Thorne, Drever and Weiss. "That turned out to be a fatal error," Weiss says. The three physicists had protracted arguments about how a large-scale interferometer should be designed and managed.

In 1986 the NSF sponsored a workshop to review the plans for the LIGO project. During the workshop, 55 scientists and engineers debated the merits of LIGO before a panel of eight physicists, who had expertise in either investigating gravitational waves or managing large science projects. The panel strongly endorsed the scientific goals of the LIGO project, and it asked that the NSF fund the construction of LIGO.

But sensing problems in the management of LIGO, the advisory group also recommended that the NSF appoint a director for the project. A year later Rochus Vogt accepted the position. Vogt, a distinguished professor of physics and a former vice president at Caltech, was displeased with the progress of the M.I.T. and Caltech teams. He tried "to build a bridge between M.I.T. and Caltech" and pushed them through a painful transformation.

Drever, Weiss and their co-workers had been accustomed to doing experimental physics on a modest scale. They enjoyed their independence and their freedom. But Vogt realized that LIGO was not tabletop physics but big science. He needed to focus the team strictly on the task of developing LIGO. "As an R&D project to build the best possible, small L-shaped detector, it was great," he comments. "As an R&D project to prepare building the four-kilometer LIGO, it was a disaster."

In 1987 Vogt decided to delay plans to start construction of LIGO and insisted the team do more research until he was confident it could reach its goals. He also decided to concentrate on development of the Fabry-Perot interferometer and stop work on the delay-line system—a decision that Weiss reluctantly supported.

Noise Control

The LIGO team spent the next two years analyzing prototype components, including a thorough review of all conceivable sources of noise. According to LIGO's deputy director Stanley E. Whitcomb, the greatest concern at low frequencies is seismic noise, which is caused by vibrations that are transmitted from the ground to optical components. At high frequencies, the problem is photon "shot" noise, the result of small fluctuations in the power of the laser beams. The main obstacle at intermediate frequencies is suspension thermal noise, which arises because all structures vibrate slightly in proportion to their temperature.

The researchers first derived models that described each noise source and verified each model in tests on Caltech's 40-meter interferometer. Then they projected their results to predict the effect each noise source would have on the sensitivity of the four-kilometer LIGO. "In this business every sin is unforgivable," Vogt declares. "If you make one mistake, you are wiped out."

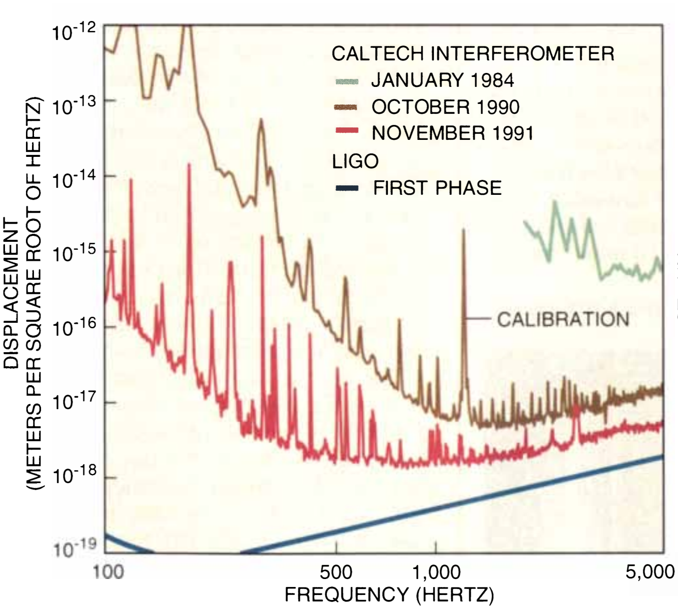
At the end of 1989 Vogt and his team decided they were ready to present the NSF with a construction proposal. As it is now envisioned, LIGO will consist of two facilities at widely separated sites. Each facility will contain an L-shaped vacuum system whose arms will be four kilometers long. The two facilities could eventually accommodate a total of nine interferometers that would operate simultaneously. The idea is to install interferometers gradually over a period of years, starting with the simplest systems and adding more advanced designs. Vogt explains, "The worst thing that we can do is to build a gold-plated Cadillac that has every sophistication built in from the start."

Around 1997, if all goes according to plan, the LIGO team will install one four-kilometer interferometer in each facility and an additional two-kilometer interferometer at one of the sites. With the three detector systems, the team believes it can rule out false signals that might be mistaken for gravitational waves. The team projects that this initial installation will be able to measure a strain of about three parts in 10^{21}, which may be 10 times too insensitive to detect the gravitational waves from neutron star binaries.

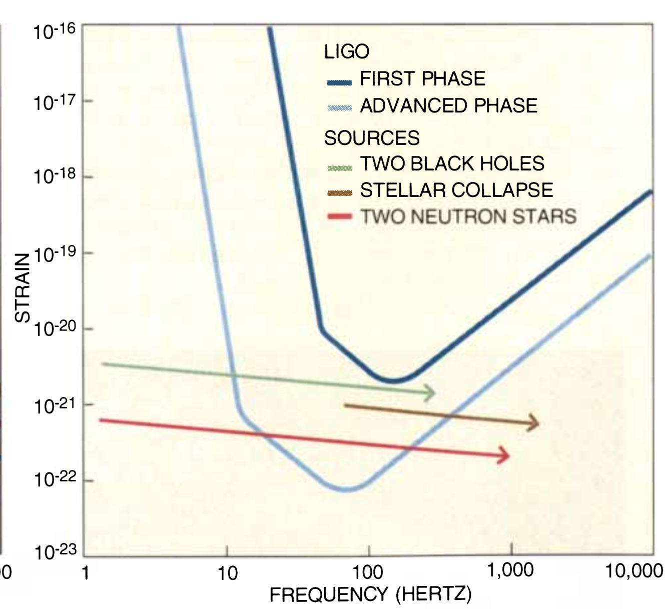
After a search for gravitational waves at that sensitivity, the LIGO designers plan to install several advanced detection systems that are currently being developed. They claim that these advanced systems will boost the sensitivity to the magic number of four parts in 10^{22}—the value needed to have a good shot at detecting neutron star binaries. By the time they install these detectors, they hope that a third kilometer-size detector will be built by one of the groups in Europe, Japan or Australia [*see box on next page*]. The third detector will allow them to determine the location of sources of gravitational radiation in the sky.

LIGO must still face the vagaries of funding and politics, however. In May 1990 the National Science Board approved its construction, but the following November Congress appropriated only $500,000 for continued engineering and design work. In March 1991 a subcommittee of the House Committee on Science, Space and Technology reviewed the project. It listened to the testimony of Vogt, J. Anthony Tyson of AT&T Bell Laboratories and others.

The opposition to LIGO was spearheaded by Tyson, whose research includes gravitational-wave detection. He argued that the early LIGO would probably be 100 times too insensitive to detect neutron star binaries. "In my judgement it is premature to commit to a full-scale LIGO," he said, recommend-



DISPLACEMENT SPECTRUM shows how the sensitivity of the Caltech interferometer has improved in recent years and how it compares with the projected sensitivity of LIGO. The interferometer's ability to detect gravitational waves depends on both the displacement spectrum and the nature of the waves. For a short pulse of gravitational radiation having a particular frequency, the smallest detectable change is approximated by multiplying the displacement at that frequency by the square root of that frequency. The narrow peaks do not reduce the sensitivity of the device significantly.



STRAIN SPECTRUM shows LIGO's projected sensitivity and three plausible sources of gravitational waves. (The projection allows for an adequate signal-to-noise ratio to rule out false events.) The red line indicates the strength and frequency range of the waves that could be measured on the earth if two neutron stars, 650 million light-years away, spiral together. The brown line shows the predictions for a star collapsing asymmetrically 100 million light-years away. The green line represents two coalescing black holes, 10 times larger than the sun, 650 million light-years away.

Entering the Age of Gravitational Astronomy

By observing gravitational waves, scientists hope to investigate some of the most mysterious events in the universe: the explosions of massive stars, the interactions between neutron stars and the collisions between black holes. To record useful information about gravitational-wave sources, astrophysicists expect that they will need at least three kilometer-size facilities located at different sites around the world. The U.S. has appropriated $23.5 million for LIGO, which consists of two four-kilometer facilities.

The approval of LIGO has raised hopes that funding may be forthcoming for similar projects in other countries. At the moment, the U.S. leads the field of gravitational-wave detection, but Germany, Britain, France and Italy are not far behind. Japan has started an ambitious program.

Each gravitational-wave facility would house one or more devices known as interferometers. These devices are sensitive to small movements that would be produced when the components of the interferometer interact with a strong gravitational wave. The sensitivity of the instrument is related to its length. By operating three interferometers simultaneously, scientists can determine the approximate location of gravitational-wave sources in the sky. The approximations should improve as the number of facilities and the distance between each increases.

Researchers in Germany and Britain have agreed to build a three-kilometer interferometer. Although the project has received scientific approval in both countries, the German government has not yet appropriated funds, and British agencies are not prepared to invest heavily in the project.

Karsten Danzmann is the co-director for the effort to build the German-British interferometer, and he also heads a team that is reconstructing a 30-meter interferometer at the Max Planck Institute for Quantum Optics in Garching. "I consider the approval of LIGO a big help for the approval of the European plans," he comments. "It still seems a real possibility to complete the European and American detectors at about the same time and to start the age of gravitational astronomy."

The German and British groups hope to team up with their French and Italian colleagues, who have plans for their own three-kilometer interferometer. In the spring French and Italian officials will decide whether to fund the project. According to Alain Brillet, co-director of the French-Italian collaboration, the project will not be granted "anything like $23 million in 1992, but the go-ahead for LIGO is a strong incentive for accelerating the decision."

Japanese researchers have quickly become strong competitors in the field. They recently built a 10-meter interferometer at the Institute of Space and Astronautical Science (ISAS). In April 1991 Japan began a four-year, $5-million project to construct a 20-meter interferometer at the National Astronomical Observatory and a 100-meter device at the ISAS. Nobuki Kawashima, director of the program at ISAS, says the 100-meter interferometer will be built to avoid the "engineering risks" involved in jumping from a 10-meter device directly to a kilometer-size instrument.

LIGO director Rochus Vogt comments: "The time is right to pluck the apple."

LASER LIGHT bounces between mirrors in a demonstration of the delay-line system for the 30-meter interferometer in Germany.

ing instead that construction of LIGO be postponed until the researchers demonstrated greater control over noise sources using the Caltech interferometer. The House Committee agreed: in June it continued to endorse "the expenditure of funds for research activities or design studies" but advised against authorization of funds for construction of LIGO.

The NSF, Caltech and M.I.T. retaliated by lobbying members of the Senate. They argued that LIGO had been strongly endorsed by physicists who had conducted peer reviews of the project. They also claimed that Tyson had underestimated the sensitivity of LIGO.

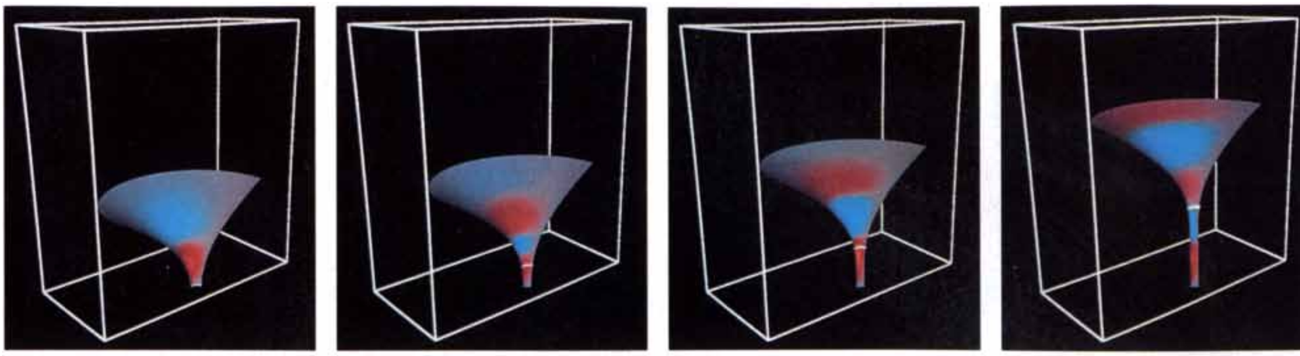
The LIGO advocates argued they were prepared to build the facility. They felt that a smaller version of LIGO would not be cost-effective, would slow the development of the technology and would greatly reduce the possibility of detecting gravitational waves.

For now, the advocates seem to have carried the day. Last September Congress passed a bill appropriating $23.5 million to begin construction of LIGO. But the project has not seen the last of the opposition. Many physicists feel that LIGO—like other "big science" projects, such as the Superconducting Supercollider—should not be a national priority. Astronomers fear that money spent on LIGO will indirectly reduce funds for astronomy, despite arguments by Vogt and others that LIGO does not divert funds from other research, because it is a line item in the NSF budget.

As plans stand, though, the LIGO team will create the final engineering diagrams for the twin facilities during the coming year and select two sites that are widely separated on the continental U.S., ideally ones that are flat, large and quiet both acoustically and seismically. The team has received 19 proposals from 17 states that wish to give one of the facilities a home.

The cost of the sites—indeed, even that of the interferometers—pales before that of the vacuum system. All the components of LIGO must be maintained under vacuum conditions to prevent scattering of the laser beams by gas molecules. Each LIGO site will require steel pipes that enclose a volume of 9,000 cubic meters, making it the world's largest vacuum system. The cost of the system, the related structures and buildings will consume more than 90 percent of the LIGO budget.

"I think LIGO is kind of a long shot, in the sense that it's probing uncharted territory," Joseph Taylor warns. "But that's why you do science—because you don't know all the answers ahead of time."



GRAVITATIONAL WAVE interacts with a black hole in a supercomputer simulation. The black hole curves space, which is represented by the off-white funnel. The red and blue regions represent gravitational waves. The region below the white band indicates the interior of the black hole. In the sequence of images above (*from left to right*), a gravitational wave approaches the black hole and enters, causing the black hole to generate new waves. The images are based on simulations conducted by David Hobill, Larry Smarr and David Bernstein of the University of Illinois.

Ultimately, the proof of LIGO's worth will be in what it reveals about the universe and the nature of gravity. Simply confirming the existence of gravitational waves once and for all would be an accomplishment that could justify the entire enterprise. But that verification would be only the beginning of a series of experiments that would further confirm—or challenge—prevailing theory.

Physicists should, for example, be able to determine such basic properties as the speed of gravitational waves. If the waves travel at the same speed as light, as predicted by theory, a burst of gravitational waves should arrive at the same time as a burst of electromagnetic radiation from the same event.

Beyond the First Wave

Thorne readily admits that LIGO is after bigger game than neutron star binaries: at the top of his wish list are black holes. Although astronomers have very little knowledge about how many black holes might exist, let alone how many black hole binaries might exist, "most pundits expect that if LIGO can look through the whole universe, it will see the coalescence of black hole binaries at a rate of many per year," Thorne suggests.

Two black holes, like two neutron stars, will spiral toward each other and produce a gravitational signal that rapidly increases in amplitude and frequency. A gravitational signal from a binary system, as Schutz has shown, contains information about the eccentricity and inclination of the orbit, the masses of the objects and the absolute distance to the source. And the eventual collision of two black holes will provide an extraordinary opportunity to test the theory of relativity. "For the first time, we may really see what a black hole looks like experimentally," Thorne asserts.

Einstein's theory has a perfect record in predicting how space and time react to masses that are relatively small and moving slowly. But theorists equipped with Einstein's equations have not succeeded in predicting what will happen to space when black holes collide. The reason, Thorne explains, is that Einstein's equations are "horrendously nonlinear," and so the dynamics of the collision may be very sensitive to such parameters as the masses of the black holes and the speed of their rotation and their orbit.

Thorne looks forward to the days when experimentalists record the gravitational waves from such black hole collisions and when theorists learn to simulate the events using supercomputers. "By comparing the observations with the computations, we may get a far, far deeper understanding of what gravity is all about," he says.

It is even remotely possible that researchers will eventually be able to detect the gravitational radiation from the birth of the universe, or what is called the big bang. Actually, theorists calculate that gravitational radiation was first produced about a millionth of a quintillionth of a quintillionth of a second after the big bang. They think the gravitational waves from that moment have traveled freely through the universe—without being absorbed or scattered by matter. The waves could therefore reveal what the universe was like at literally the beginning of time.

By comparison, electromagnetic radiation began propagating freely through the universe a million years after the big bang. Physicists have detected that radiation, and the discovery has led to important insights into the evolution of the universe.

Although theorists suggest that gravitational waves will emerge from spiraling neutron stars, exploding massive stars, colliding black holes and even the big bang, all their predictions are limited in a fundamental sense. They are based on information about the universe that astronomers have garnered from electromagnetic signals. Some predictions about gravitational-wave sources are almost certain to be wrong. To envision gravitational-wave sources by studying only electromagnetic radiation is somewhat like trying to guess the sound of an orchestra by watching its conductor.

Theorists also concede it is illogical to think all strong sources of gravitational radiation will emit enough electromagnetic radiation to be visible. The universe is very likely to contain sources that no one has anticipated. "The only thing that we can promise are coalescing neutron star binaries," Vogt declares. "I personally believe there is a hell of a lot more to see than that."

Vogt, now 62, has no illusions about his role. "I will be there just for the opening of the window. I will take one peek, and I will drop dead," he says with frightening conviction. "Others will learn things that I have never dreamed of and which I will never know about." But, he adds: "The first glimpse will be worth it—it justifies devoting the rest of my life to LIGO."

FURTHER READING

A JOURNEY INTO GRAVITY AND SPACETIME. John Archibald Wheeler. Distributed by W. H. Freeman and Company. Scientific American Library, 1990.

THE DETECTION OF GRAVITATIONAL WAVES. Edited by David G. Blair. Cambridge University Press, 1991.

THE U. S. LASER INTERFEROMETER GRAVITATIONAL-WAVE OBSERVATORY (LIGO) PROJECT. Rochus E. Vogt in *Proceedings of the Sixth Marcel Grossmann Meeting on General Relativity.* Kyoto, Japan, June 1991, LIGO 91-7; September 1991.


Begin a voyage of discovery that takes you from quarks to the cosmos, and beyond.
SLEEP
THE SCIENCE OF STRUCTURES AND MATERIALS
ANIMAL NAVIGATION
FROM QUARKS TO THE COSMOS
Tools of Discovery

Join the Scientific American Library now by examining *From Quarks to the Cosmos* for 15 days risk-free.

You are invited to embark upon a journey to the newest frontiers of science. From the mysteries of the brain . . . to the most spectacular phenomena in the universe.

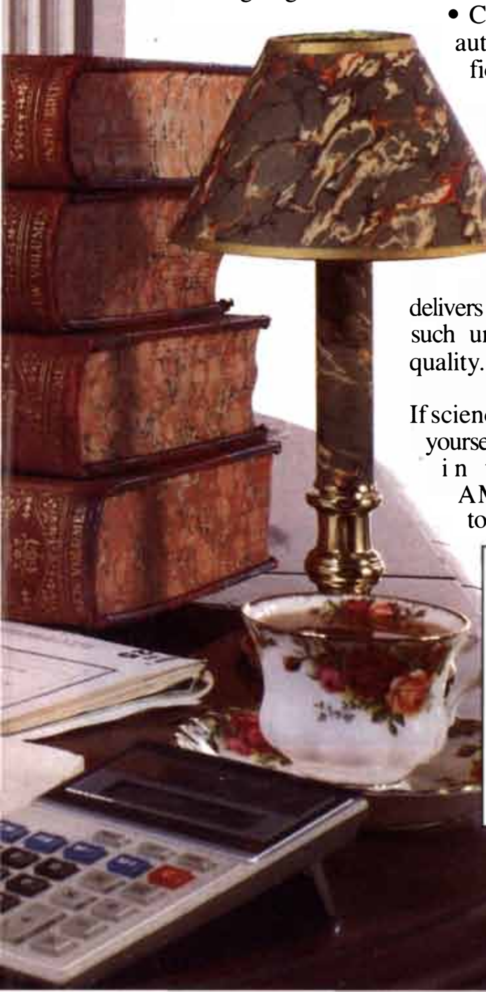
Join us and you will explore the most exciting discoveries on the scientific horizon today as a member of the SCIENTIFIC AMERICAN LIBRARY.

From Molecules, *a polarized light view of light heating oil; courtesy Manfred Kage/Peter Arnold, Inc.*

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 volume . . .

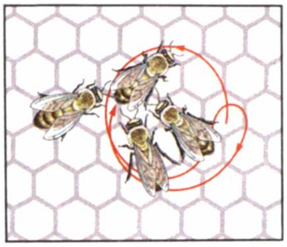
- Covers a topic of major interest with immediate relevance to ongoing research.
- Carries the authority of an author who has made significant contributions to the field.
- Is written in clear, compelling prose that's uniquely accessible even to the non-expert.



No other book club delivers the literature of science with such uncompromising editorial quality.

If science is your passion, reward yourself. Begin your membership in the SCIENTIFIC AMERICAN LIBRARY today.

Beautiful, richly illustrated volumes that are made to endure.



From The Honey Bee, *foragers dance to indicate a food source; drawing by John Hatzakis, adapted from K. von Frisch,* Tanzsprache und Orientierung der Bienen.

From the moment you receive your first book, you'll see that SCIENTIFIC AMERICAN LIBRARY volumes are produced to exacting standards. Each work:

- Features abundant photographs, paintings and drawings in full, vibrant color.
- Includes many helpful diagrams, charts, and maps to graphically illustrate the text.
- Is impeccably printed on high quality paper and clothbound for uncommon durability.

The books in the SCIENTIFIC AMERICAN LIBRARY are fascinating reading *and* they make a handsome addition to your library.

Join now by sending for your premier volume without risk or obligation.

From Eye, Brain, and Vision, *the neurobiology group at Harvard Medical School, 1963, including Nobel Prize winners David Hubel, standing right, and Torsten Wiesel, seated right. Photograph by Joseph Gagliardi.*

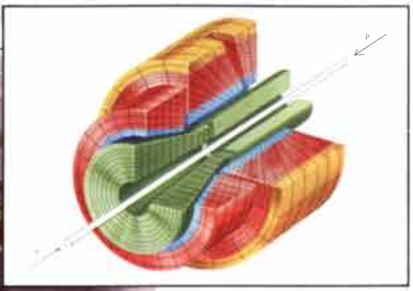
Join the SCIENTIFIC AMERICAN LIBRARY now by previewing *FROM QUARKS TO THE COSMOS,* written by 1988 Nobel Prize-winning physicist Leon Lederman and cosmologist David Schramm. This intriguing book explores the revolutionary discoveries made through the powerful merger of particle physics and cosmology.

This is just the first of many fascinating volumes you'll receive. In months to come you'll discover books covering a wide range of subjects, including *Molecules, Sexual Selection,* and *Eye, Brain, and Vision.*

As a member, you'll enjoy:

- A NEW BOOK EVERY TWO MONTHS or so, shipped direct to your doorstep.
- 15-DAY FREE EXAMINATION. Review each book without risk. You may keep it or return it for full credit.
- A LOW, MEMBERS-ONLY PRICE. Pay just $24.95 (25% off the retail price) for each book you keep.
- NO PREPAYMENT OR MINIMUM PURCHASE. Pay only for the books you choose to keep.
- OUR GUARANTEE that you'll be fully satisfied or *you may cancel at any time without penalty.*

TO JOIN, WRITE TO SCIENTIFIC AMERICAN LIBRARY, P.O. BOX 646, HOLMES, PA 19043, OR CALL TOLL-FREE 1-800-345-8112.



◄ *From* From Quarks to the Cosmos, *a schematic of a particle collider detector; drawing by George Kelvin.*

SCIENTIFIC AMERICAN LIBRARY

AN EXCITING NEW GENRE IN THE LITERATURE OF SCIENCE

Holey Silicon

Looking for light at the end of a chip

Silicon chips that reliably and efficiently produce light when stimulated electrically (rather than optically) could, in theory, bring about the marriage of optics and electronics, producing new generations of lightning-fast switches and other components for telecommunications and computers. Although wedding bells are not ringing yet, evidence is steadily building that so-called porous silicon will be more than a one-conference stand. "I'm guardedly optimistic," says Reuben T. Collins, a research manager at the IBM Thomas J. Watson Research Center.

Light-emitting silicon debuted at a Materials Research Society (MRS) symposium last year, when Leigh T. Canham and his colleagues at the Defense Research Agency in Great Malvern, England, claimed that silicon wafers, eroded into spindly structures tens of angstroms wide by a bath of hydrofluoric acid and excited by a light, would glow in a range of colors. The presentation generated many more questions than answers. The British investigators said they had also made wafers that would electroluminesce, or emit light when excited with electricity, but offered little proof. And they could only speculate on why silicon would light up at all.

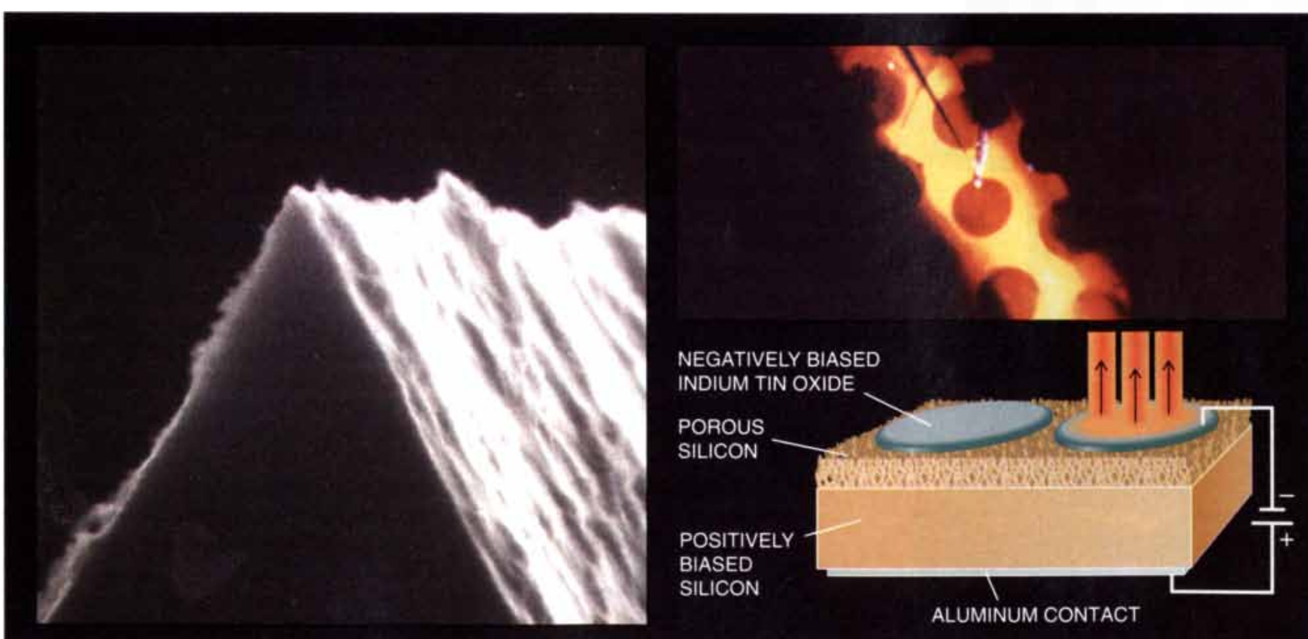
But within the past few months, many others have jumped into the fray. For instance, at the most recent MRS meeting held last December, researchers from the Institute for Solid State Technology in Munich and the Tokyo University of Agriculture and Technology presented evidence of electroluminescence. At the same time, scientists from IBM and Spire Corporation in Bedford, Mass., described making silicon devices that act like light-emitting diodes.

Building such devices is tricky because researchers must create sandwiches of positively and negatively charged layers of silicon and porous silicon, then lay down an electrical contact without damaging the material. "A lot of people have tried to make silicon electroluminesce and have had a lot of problems in doing it," Collins notes.

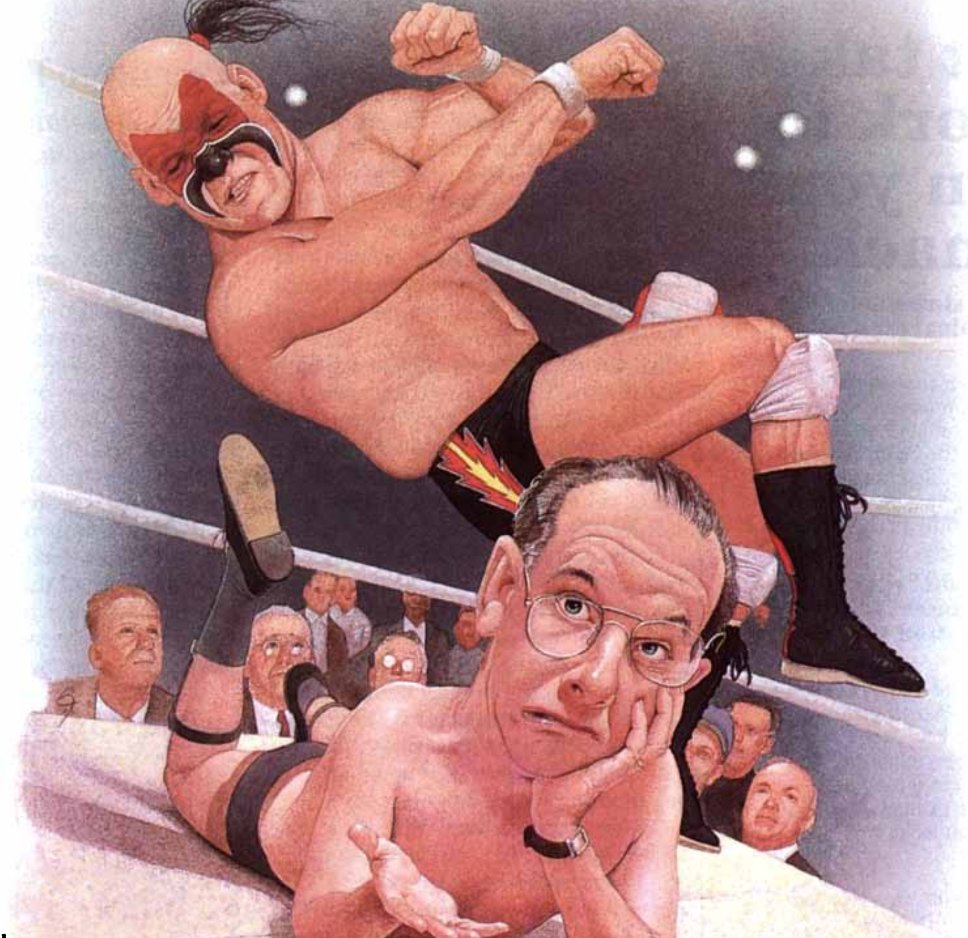
Extrapolating from their work on photovoltaic cells, Spire scientists added a negatively charged layer of indium tin oxide (ITO) on top of the porous silicon. ITO is transparent to visible light. So when the researchers applied a voltage, they could clearly see orange light coming from the top of the device, says Fereydoon Namavar, a senior scientist at Spire. IBM researchers built a more traditional diode that emitted light from the edge. They surrounded the porous silicon with layers of negatively and positively charged conventional silicon and then etched through the material.

From a pragmatic vantage point, not only must porous silicon turn electricity into light, it must do so efficiently. So far no researchers are willing to discuss the efficiencies they have achieved. Namavar's material, for instance, requires about 10 volts to produce "visible" light, he says. He hopes to lower that power requirement to about two volts by improving the contacts.

Apart from experimental efforts, workers are still puzzling out a theory to explain the unusual behavior of porous silicon. One school, led by Canham, has proposed that porous silicon is made up of "quantum wires," struc-



***UNUSUAL LIGHT-EMITTING SEMICONDUCTORS** include germanium* (left) *etched into the shape of "quantum wires" by researchers at the Research Triangle Institute. Porous silicon, fabricated by workers at Spire Corporation, also lights up when excited by ultraviolet light* (top right). *To make silicon luminesce with electricity, the researchers built a diode device* (bottom right). *Visible light from the porous silicon shines through the indium tin oxide caps.*



tures that confine the movement of electrons to one dimension, thereby promoting the recombination of electrons and positive charges. Those recombinations produce light. In contrast, physicists at the Max Planck Institute for Solid State Research in Stuttgart attribute the photoluminescence to siloxene, a silicon-oxygen-hydrogen compound. Siloxene has long been known to exhibit the characteristics now associated with porous silicon. Recently the researchers showed that siloxene on crystalline silicon (with an intermediate layer of calcium disilicide) similarly luminesces; such material would be far easier to manipulate than the brittle porous silicon, they suggest.

Rama Venkatasubramanian and his colleagues at the Research Triangle Institute in North Carolina decided to tackle the question from a different direction. "I thought, 'If the effect is valid in silicon, I should be able to do it in other semiconductors,'" he recalls. To test his idea, he used photolithography, a highly precise technique for etching quantum wires on germanium.

At the International Electronic Devices Meeting, also held last December, the Research Triangle investigators reported that the germanium returned a reddish glow when it was pumped with light. Even these results, however, do not conclusively point to a quantum effect. "There's still fluorine in there that may be causing the luminance," Venkatasubramanian says.

Equally curious is why such etching techniques have not produced light-emitting silicon when the less controlled hydrofluoric acid treatment has. Venkatasubramanian believes the explanation lies in the differences in the electronic (or band gap) structures of silicon and germanium. Still, he hopes that if he can make germanium light up with electricity, the work may point to ways of achieving the same result with silicon.

Even though optically excited silicon emits a wide range of colors, electrically powered silicon has so far offered a more limited range. Canham, for instance, says he and his colleagues have seen red to orangy-yellow colors in their electroluminescence experiments, but no green. Even so, "a single color would be good enough for a monochrome display panel," Namavar points out.

Everyone in the field is haunted by the roller-coaster rides of enthusiasm—and funding—other discoveries have endured. "We're trying to be very humble, very careful about our expectations at this time," Namavar says. Yet many researchers quietly predict that there will soon be light at the end of this silicon tunnel. —*Elizabeth Corcoran*

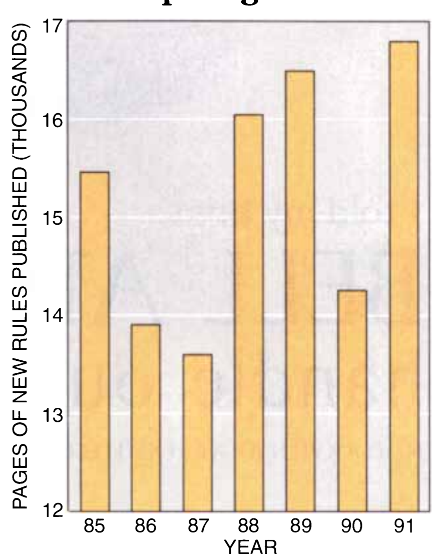
Executive Fix

Is the Competitiveness Council overstepping its bounds?

How could an administration with an aversion to "industrial policy" help U.S. manufacturers compete in the global marketplace? The Bush administration said it had one solution: ease the burden of costly regulations imposed by such executive agencies as the Food and Drug Administration and the Environmental Protection Agency. So in 1989 President George Bush created the White House Council on Competitiveness. With Vice President Dan Quayle at the helm, the council inherited the mantle of the Reagan-era Task Force on Regulatory Relief.

Instead the council has become a lightning rod in the power struggle between the Congress and the White House. Democrats protest that the council is a shadow government, operating in secret and coercing executive agencies into modifying regulations. Several congressional committees are now investigating the council's interventions in agency rule making and are locked in legal tugs-of-war over access to council documents.

Red Tape Fights Back



SOURCE: Federal Register

The council formally consists of seven high-ranking officials, including Quayle, Richard G. Darman, director of the Office of Management and Budget, Nicholas F. Brady, secretary of the Treasury, Samuel K. Skinner, the president's chief of staff, and Michael J. Boskin, chairman of the Council of Economic Advisers. Its operations are run by six members of the vice president's staff under the direction of Allan B. Hubbard.

The group has no connection to a private-sector group of industry and academic leaders that bears the same name.

Opponents of Quayle's council, aided by leaks from within executive agencies, charge that it skates around legally required procedures for rule making by leaning on agency heads to change proposed regulations in closed-door sessions. At the same time, conservative critics point out that the council has failed to stem the regulatory tide. The volume of new regulations published last year is 20 percent more than that in late Reagan years.

Moreover, critics from both ends of the political spectrum question whether the council's small staff has the expertise to assess the vast range of regulations proposed by the federal government. Lack of technical knowledge has, for example, been blamed for the council's foundering attempt to propose a legal definition of wetlands. Under that proposal, one third of the nation's wetlands would have lost their protection.

For their part, council staff members insist that they "work with" experts within the executive agencies. Ironically, it is the precise nature of that contact that is attracting the interest of congressional investigators. The White House insists that closed-door deliberations are protected by executive privilege. But Cass R. Sunstein, an administrative law expert at the University of Chicago, told a December hearing held by Congressman Henry A. Waxman of California that, even so, the authority for regulatory decisions must legally remain with agency heads. Sunstein further said that if the council influences regulations, its communications with outside organizations have to be a matter of public record.

The possibility that the council may be exercising improper regulatory authority has inspired at least two other congressional committees to examine how the council came up with the series of reforms last November to speed the drug approval process at the FDA. The council's announcement said the FDA would use external reviewers to scrutinize clinical trial results and would "immediately" lose its jurisdiction over early-stage testing of potential drugs in healthy subjects.

One company, Eli Lilly, took the council at its word and informed the FDA (over the objections of an FDA official) that it had asked an Indiana University review board to approve a planned trial of a drug for treating Alzheimer's disease. Lilly backed down when the FDA insisted that the new procedures would take effect only after a formal rule-making process.

Nor does it appear that the FDA fully agreed with the council. David A. Kessler, the FDA commissioner, said in December that he had originally opposed the plan to make use of external reviewers but later concluded it was worth trying.

Kessler is not the only one with reservations. The Health Research Group, an arm of Ralph Nader's Public Citizen organization, published in December the results of a survey of the FDA's in-house physician-reviewers. Of the 47 physicians who responded (38 percent of those polled), more than 90 percent opposed eliminating FDA jurisdiction over early-stage trials, and more than 80 percent opposed the plan to use external reviewers. Charles C. Edwards, president of the Scripps Clinic and Research Foundation in La Jolla, Calif., who conducted a major review of the FDA 18 months ago, doubts that external review will save either time or money because FDA staff will still have to oversee the activity.

Although there are few votes to be won in criticizing efforts to speed new drug approvals, environmentalists opposed to easing pollution regulations can count on strong public support. They charge that industry representatives used the Competitiveness Council for backdoor access to the rule-making process and are attempting to gut key provisions of the 1990 Clean Air Act. Under one of the council's proposed changes, now under review by the Justice Department, industries would be able to specify their own emission limits, and state agencies would have just seven days to object.

Council staff do not deny their frequent contacts with industry. And critics such as OMB Watch and Nader's Public Citizen have fanned the fires by raising the specter of conflict of interest. They argue that Hubbard and Quayle violated the law by influencing regulations that could affect companies in which they have holdings. Quayle, for example, has an interest in a newspaper company that owns a paper mill; Hubbard owns half a chemical company and had a stake in a utility until he donated the latter to charity late last year. He placed other holdings in a blind trust. Quayle issued Hubbard a "conflict-of-interest waiver" in 1990, but the House Government Operations Committee, headed by Congressman John Conyers, Jr., of Michigan, plans to inquire into whether the waiver was too broad to be legal—or too late. Hubbard's waiver did not take effect until a year after he took his job.

The controversy swirling around the council shows no sign of dying down. Congressman John D. Dingell of Michigan has now joined the fray. Dingell is examining the council's influence over EPA regulations to enforce the Resource Conservation and Recovery Act, an important piece of legislation governing solid waste that comes up for reauthorization this year. The agency has attempted to withhold from Dingell's committee certain documents on the subject, on the grounds that they involved meetings with the Competitiveness Council.

But Dingell does not back down easily. Another congressional committee has already used subpoenas to gain access to documents pertaining to the Competitiveness Council, and more subpoenas might be on the way. One way or another, some of the council's quiet intercessions may be dragged into public view. *—Tim Beardsley*

Vested Interests

Du Pont's Kevlar has enjoyed a virtual monopoly of the market for bullet-resistant vests for the past 20 years. The company claims that vests woven from that aramid fiber have saved the lives of more than 1,300 law enforcement officers, most of them from bullets. But in 1987 a new standard threatened to give the edge to a competitor, and Du Pont came out shooting.

The trouble began when the National Institute of Justice (NIJ), an arm of the Department of Justice, toughened up its voluntary standard for body armor and added a new test that required firing nine-millimeter bullets into the vests at an angle of 30 degrees. Fewer than 50 percent of Kevlar vests passed. In contrast, vests made with Allied-Signal's Spectra Shield, a high-density polyethylene composite, passed 97 percent of the time, NIJ officials say.

Furious Du Pont executives soon persuaded an industry group, the Personal Protective Armor Association, to create an alternative standard under which Kevlar vests performed well. Allied-Signal, unsurprisingly, continued to support the tougher NIJ standard. Police departments were caught in the cross fire, but the International Association of Chiefs of Police and the Fraternal Order of Police both came down on the side of Allied-Signal and the NIJ standard.

Still the sniping continued. Du Pont argues that NIJ-certified vests are uncomfortable and circulated a videotape showing a mannequin wearing a Spectra vest on fire. Allied-Signal protested that its product, with a flash point of 354 degrees Celsius, will withstand heat and flame more intense than that which would kill a human. The company filed suit against Du Pont last year, charging it with false advertising and illegally monopolizing the market.

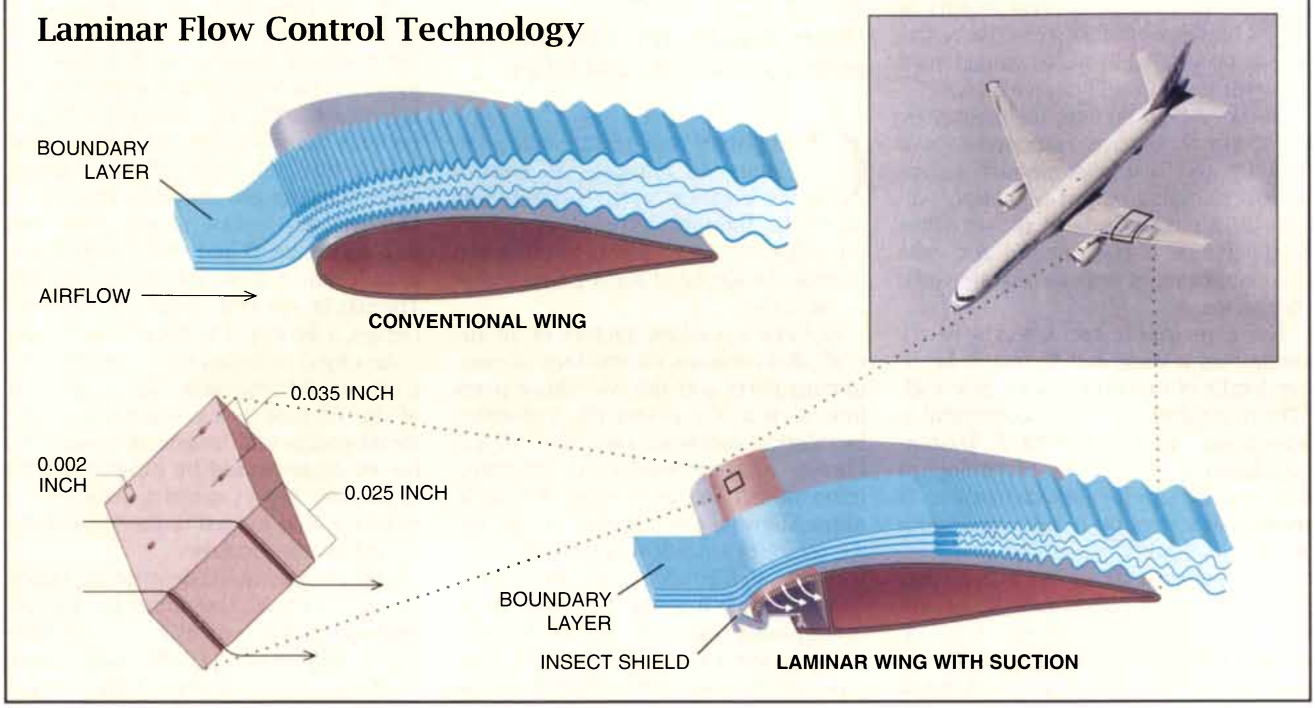
A group of concerned congressmen eventually called in the Office of Technology Assessment (OTA) to be an honest broker and lay out the complex technical issues of testing. Kevlar tends to "ball and bunch" after multiple shots against the clay support used in the tests, which reduces its effectiveness. The Du Pont–supported industry test standard solves the problem by allowing the material to be patted back into shape after each shot, a dispensation that Allied-Signal considers unrealistic. Du Pont says a more resilient backing material than clay would solve the problem—but no alternative has yet been shown suitable.

Du Pont also opposes wet testing, which is optional under the industry standard: Kevlar loses up to 40 percent of its ballistic resistance when wet. But Allied-Signal supports the NIJ's compulsory wet test, because Spectra Shield is unaffected by water.

Michael B. Callaham of the OTA says his study will not make determinations about which tests should and should not be used, although it may suggest improvements to the existing standard and test procedures. But congressional aides to Congressman Edward F. Feighan of Ohio say the report will probably be effective ammunition to support a bill introduced by Feighan that would make the existing NIJ standard—or something tougher—mandatory. If so, Du Pont's domination of the body armor market, estimated at $35 million per year for the civilian sector alone, might come to an end.

On the other hand, there could be a truce. A likely outcome, industry sources say, is a boom in vests that incorporate both Kevlar and Spectra Shield. Du Pont has previously discouraged efforts to combine the two, although they have properties that complement one another. *—Tim Beardsley*

Laminar Flow Control Technology



Drag Race

Aerodynamics tackles one of the final frontiers

The rush of air that streams across the wings on a commercial jet can quickly turn into a turbulent maelstrom that slows flight and causes fuel consumption to soar. Now aerospace researchers on both sides of the Atlantic are racing to equip wings and other surfaces with vacuum cleaner–like elements that can stave off these violent currents.

The payoff to the aircraft maker whose planes incorporate what is known as laminar flow control could be enormous. New aircraft might be designed with lighter airframes and smaller engines. Using less fuel generates tremendous cost savings and reduces nitrogen oxides and other harmful by-products of combustion.

The National Aeronautics and Space Administration estimates that a reduction in friction of 1 percent across the entire U.S. commercial air fleet could create a $100-million fuel savings. Laminar flow control may be able to achieve reductions in drag of up to 20 percent.

No other redesign measure can produce comparable economy. The shape of the modern jet airliner has remained largely unchanged since the 1950s because of its streamlined efficiency. "Some people say this is the last big thing we can do in aerodynamics for airframes," proclaims Joachim Szodruch, a general manager for research and technology at Airbus Industrie, the European aircraft consortium. Adelbert L. Nagel, a manager at Boeing Commercial Airplane Group, adds: "Laminar flow control is the last big carrot we know about in aerodynamics for subsonic commercial airplanes."

The goal is to preserve an orderly integrity in the boundary layer, the strata of air currents moving at varying velocities up to a tenth of an inch or so around the aircraft. Rivet heads, dirt and engine noise are enough to turn these nicely behaved layers of airflow into a chaotic torrent.

Smaller aircraft can better retain laminar flow because of shorter length of flight surfaces and because of wings that are more nearly perpendicular to the fuselage. Merely shaping and polishing wings and other surfaces of these aircraft are enough to retain the smooth, layered currents of laminar flow. Bigger airplanes—those that have 100 seats or more—need both wing alterations and a suction system that draws air into the wing through millions of tiny holes, a preventive against turbulence.

Although various experiments aimed at retaining laminar flow were first carried out during World War II, the difficulties of manufacturing and maintaining a laminar wing have acted to retard development. These concerns have been somewhat allayed by recent testing. The first suction system for large airliners was flown in 1990 and 1991 as part of a joint program with NASA, Boeing and the U.S. Air Force. Using lasers, engineers drilled millions of perforations of about 0.002 inch in diameter into a titanium panel that was fitted into the upper left wing surface of a Boeing 757 aircraft. A duct system for the holes was installed inside the wing, and a pump was placed in the engine strut.

The 757's wing was smoothed and painted, and a metal shield was extended over its leading edge during takeoff and landings to protect against insects clogging the holes. Although the panel in the 757 covered only the front 15 percent of the wing, sensor readings showed that laminar flow extended back over two thirds of the upper surface. "Even 50 percent would have been a roaring success," Nagel says.

In November, NASA and Rockwell International announced they had successfully demonstrated a similar suction system on an F-16 fighter at supersonic speeds. The research is directed toward commercial supersonic transports. "We feel this might be a make-or-break technology," says Fayette S. Collier, Jr., head of the laminar flow control project office at the NASA Langley Research Center.

Officials at Airbus acknowledge that the NASA flight tests have established a clear lead for the U.S. "We're some years behind," Szodruch admits. But Airbus has consistently tried to compete with its U.S. rivals by deploying new technology first. For its part, NASA has restricted distribution of some information about the various test programs. "We want to prevent export of the technology," Collier says. "We're trying to protect what we've gained with taxpayer money and not give it away."

So far tests of laminar technology by European aerospace companies have been limited to smaller aircraft. Airbus is now planning to put a suction system on the fin of one of its airplanes to determine if the ducting and other encumbrances impede maintenance operations.

The consortium also intends to try suction on a wing, but it, too, is keeping details of this program confidential. "The technology gives you a competitive advantage," says Bernhard F. Dziomba, chairman of Airbus's technological working group for aerodynamics. "I think that's why both sides are being so closed about it."

NASA believes about five more years of development work remain before the technology will be ready to be incorporated into a new airframe. Determining whether the wing's perforated surface will give the airlines a colossal maintenance headache—keeping the holes open and the wing surface clean—is just one of the issues to be addressed.

At the same time, aircraft builders are also exploring another less expensive—and less efficient—way of reducing drag. The technique involves clothing the aircraft surfaces in a kind of sharkskin suit. For a number of years, aircraft companies have been testing a tape made by 3M, the St. Paul–based diversified manufacturing concern, that is covered with a series of microscopic ridges and grooves.

Like the coarse flesh of a shark, these so-called riblets lessen friction caused by turbulence. But riblets are less effective than laminar flow control: they reduce drag by 2 percent or so. Once turbulence has erupted, they serve as a mild palliative. But they cannot restore the boundary layer to its original laminar state. Maintenance is also a big concern. Airlines would be loathe to have to strip off a tape and carefully reapply it for each periodic inspection.

No one is dismissing the tiny ridges yet. They could conceivably be applied where suction is impossible: over most of the fuselage, for example. Airbus, for one, has placed riblets on more than 70 percent of a test aircraft.

Air is not the only fluid where riblets have proved their worth. In its successful 1987 attempt to regain the America's Cup, the yacht *Stars & Stripes* had the 3M tape stuck to its hull. Drag reduction may now become a pivotal technology in winning the race to loft an ultimately efficient airliner. —*Gary Stix*

Objective Data

DARPA nudges development of object-oriented data bases

Computers now reside inside airplanes—to navigate, to control the engines, even to land. Surprisingly, though, software engineers are still wrestling with how to put a reasonable facsimile of an airplane inside a computer.

Keeping a working picture of an aircraft that contains the millions of component parts and the way those parts interact is a task awkwardly performed by existing computer software. What is known as object-oriented programming offers a partial solution. The technique allows a programmer to provide a straightforward description of an object, say, an aileron, and its behavior.

Still needed is a way to codify these descriptions into a data-base management system that would allow the various objects and their related operations to be stored, retrieved and used by several different people at once. Data are now stored as simple files—one for each part or assembly—or in a database management system. A relational data base, which consists of entries in tables, can be used to track simple relationships, but it falls short if a user tries to tie together the millions of subsystems and components that make up the airplane in the computer. "How can you describe a Boeing 777 with a collection of tables?" asks Stanley B. Zdonik, a professor of computer science at Brown University.

Zdonik is a participant in a program set up by the Defense Advanced Research Projects Agency (DARPA) to help solve the problem by guiding the development of object-oriented data bases to complement the programming language. Most of this software (sold by five or so small companies) has emerged in the past two years.

The potential importance of these data bases in both military contracting and battlefield command and control prompted DARPA to set up a three-year program to address both technical issues and lack of standards that could impede wider usage of the technology.

Last fall the agency began providing $22 million through contracts to Texas Instruments, a group of universities and the National Institute of Standards and Technology. "DARPA is worried that design in manufacturing takes too long and is too costly," says David Maier, a professor of computer science and engineering at the Oregon Graduate Institute of Science & Technology, which has received a contract from DARPA.

Object data bases attempt to overcome limitations of existing data storage methods. A relational data base has to search through entries in separate tables to assemble the myriad of subcomponents that make up an airplane—and then it dismantles the whole structure when the data are stored on disk. "It is like having to weld a plane together before each takeoff and then disassemble it each time it arrives at the gate," says Thomas M. Atwood, chairman of Object Design, a developer of object data bases.

An object-oriented data base, in contrast, would represent the subsystem of the airplane as a set of objects. The metal sheets and hinges of an aileron, for example, would be directly linked to another object called a wing, which in turn would be tied to the biggest object of all, the airplane.

This preassembled description, much like a schematic, resembles the way an engineer pictures a design. Object data bases can also accommodate more than the simple strings of letters and numbers present in conventional software: voice messages annotating an engineer's work, for instance.

Kinks remain, however. The complexity of these multimedia information caches makes access slow. And even answering a simple question can sometimes prove difficult. To extract information rapidly, university researchers under the DARPA contracts are exploring options that range from simple indexes to dissecting objects into their component parts so that the data they contain can be processed by several computers simultaneously. One scheme would place objects on a disk near to one another, corresponding to their arrangement in the physical world. An object for an aileron would be close to a wing so that the disk drive could access the objects at close time intervals.

Results of the research will be shared with companies developing object data bases. DARPA believes such data bases are so important that it cannot depend entirely on a coalition of small companies to develop standards for the technology, although the small subindustry is forging ahead with efforts to do so. Gio Wiederhold, a DARPA program manager, says object-oriented technology might eventually become the basis for a series of shared data bases that would allow engineers for different military contractors to access schematic diagrams or instructions for computerized machine tools.

Before that dream is realized, however, these systems must be able to provide simple answers to simple questions. —*Gary Stix*

Getting Out the Bugs

Wise management may keep a safe pesticide effective

Because insects have a maddening knack for evolving resistance to pesticides, sometimes within only a few years, the unofficial motto of the pesticide industry might as well be: "Use it and lose it." Now the biotechnology industry and environmentalists are scrambling to stay a jump ahead of bugs that threaten to overwhelm an important group of insecticides based on bacterial toxins of *Bacillus thuringiensis* (*Bt*). "We don't want *Bt* to be a flash in the pan that lasts for five years, makes profits for somebody, and then we're on to the next product," says Fred Gould of North Carolina State University.

Since at least the 1950s some farmers have been spraying *Bt* spores over their crops to discourage insects. Unlike broad-spectrum chemical insecticides, the various toxins in the spores are selectively lethal to beetles, flies, butterflies and moths—all common agricultural pests. Other insects as well as animals and humans are unharmed by *Bt*, making it ideal as a safe, minimally disruptive pesticide. *Bt* has one drawback, however: farmers must apply it much more often than they would chemicals, which raises its effective cost considerably. "After you spray it on, half the dose degrades within a day and a half," Gould explains.

As a result, *Bt* and all other biopesticides are estimated to hold less than 1 percent of the agricultural insecticide market. The biotechnology industry has therefore been developing new sprays that last longer than their predecessors or that contain higher levels of toxins. *Bt*-based pesticides are now marketed by many companies, including Abbot, EntoTech (a division of Novo Nordisk) and Mycogen.

Yet the more effective *Bt* products become, the more they increase the pressure on insects to develop resistance. During the mid-1980s, laboratory tests began to show that insects such as the Colorado potato beetle and the tobacco budworm had the genetic potential to develop *Bt* resistance under certain circumstances.

The problem ceased to be academic when Bruce Tabashnik of the University of Hawaii at Manoa and his colleagues discovered that a population of diamondback moths preying on a field of watercress in Hawaii had doubled their resistance to *Bt* between 1986 and 1989. "The irony is that the grower had been using a lot of *Bt* because he was environmentally conscious," Tabashnik says.

The moths in Hawaii and another group in Florida are the only well-documented cases of naturally arising *Bt* resistance, but observers are worried that they will not be the last. "Our concern is that if the biotechnology industry moves ahead with a lot of products, we will very quickly select for resistant insects," warns Edward Bruggemann of the National Audubon Society in Washington, D.C. "Farmers will then have no choice but to return to the use of chemical insecticides."

One simple approach to maintaining *Bt*'s effectiveness involves varying the microbial mixtures sprayed on crops. Various strains of *Bt* bacteria produce different combinations of toxins, and insects that are resistant to one may still be vulnerable to others. Tabashnik warns, however, that there might be "relatively few groups of toxins you could switch to, even though there are thousands of different strains."

Much recent work has also focused on inserting the toxin genes of *Bt* into plants or their symbiotic bacteria. Such transgenic crops would make and concentrate the toxins in their own tissues. Transgenic cotton, which Monsanto plans to introduce within the next five years, will probably be the first *Bt*-equipped plant, and potatoes and other crops should follow soon thereafter.

A scheme to counter the resistance pressure of transgenic plants would be to provide havens for *Bt*-susceptible insects by planting some ordinary crops among transgenic ones. The susceptible insects would live on the plants but would also breed with their more resistant fellows. If *Bt* resistance is a recessive genetic trait, as Tabashnik says some preliminary evidence suggests, most of the offspring in the next generation of insects would be vulnerable to *Bt*. A potential problem with this idea, Bruggemann says, is that no one yet knows whether a farmer would end up surrendering a large portion of the crop to insects.

A more exotic answer is to alter the expression of the *Bt* genes in transgenic plants. They might then be active only in certain tissues, during critical periods of plant growth or in response to chemical triggers applied by farmers. Regulating the *Bt* toxins in such ways, Gould explains, would lower the pressure for resistance. David Fischhoff of Monsanto points out, however, that the genetics of crop plants are too poorly understood for such approaches to be feasible for at least a decade. The first wave of transgenic products will therefore express *Bt* toxins continuously—and perhaps heighten the risk of resistance, Bruggemann worries.

Responsibility for resistance management will ultimately rest with the farmers, who must decide whether and how to use *Bt* products. Gould suspects the U.S. Department of Agriculture might be in the best position to offer incentives through programs that would link prudent *Bt* use with eligibility for aid. Such a proposal would be controversial, he concedes.

What farmers, researchers and industry need now, Gould continues, is more information about the consequences and effectiveness of the various approaches, alone and in combination. He is quite enthusiastic about exploring the use of *Bt* as one component of integrated pest management, in which farmers rely on both pesticides and natural predators to control insects. "You don't need to kill every bug," Gould remarks: some insects become pests only after their numbers pass a threshold.

More research on the interactions of insects, their natural enemies and crops should determine how best to use *Bt* while delaying resistance to it. Ecological balance seems to have preserved the effectiveness of *Bt* in nature for millions of years. That may be the only way to ensure its long-term survival in agriculture. *—John Rennie*

BUILT-IN INSECTICIDES* protect transgenic cotton (left) *when an ordinary plant* (right) *falls victim to pests. Photo: Monsanto.

Disarming Developments

They shall beat their swords into plowshares, and their spears into pruninghooks: nation shall not lift up sword against nation, neither shall they learn war any more.

—Isaiah 2:4

Even under the best of circumstances, transforming swords into plowshares—or more contemporarily, missile guidance systems into television sets—has a ring of improbability. Yet these days the world's major weapons producers do not have much choice. In both the U.S. and the republics that once made up the Soviet Union, arms makers are scrambling to divert their lethal arts to civilian ends.

With an effectiveness that pacificists only dreamed of, economic events have conspired to erode the appeal of the arms business. First, the centrally planned economies crumbled, freeing consumers to express long pent-up demands for fashionable shoes and videotape recorders. As peace broke out, the weight of the U.S. budget deficit began to depress any lingering yearnings for a return to lavish defense spending.

Second, the global market for armaments is sputtering. People around the world are still keen to bloody one another, but at least for now many of them do not have the cash to pay for much weaponry, says Michael D. Intriligator, an economist at the University of California at Los Angeles. Since the end of the Iran-Iraq conflict, the third largest war in this century in numbers of casualties, "there's been a substantial decline in demand for arms," he says.

As triumphant as the world's shift to plowshares may be, it is accompanied by tough challenges that have none of the glamour of a cinematic submarine chase. On one level both U.S. and former U.S.S.R. military contractors share some of the same problems, notes Lawrence R. Klein, a Nobel economist at the University of Pennsylvania. Weapons manufacturers on both sides of the Iron Curtain were accustomed to serving a single customer who had precise demands and the bottomless pockets of an indulgent uncle. In contrast, producing consumer goods means catering to the ever changing needs and whims of a diverse collection of price-conscious consumers.

In addition, the former Soviet military-industrial complex faces some unique—and critical—hurdles, suggests Julian Cooper, who directs the Centre for Russian and East European Studies at the University of Birmingham in England. For instance, the U.S. Congress will be largely responsible for making decisions to scale back American military production, but it is unclear who will make such calls in Russia, Ukraine or any of the other 10 nascent republics.

Moreover, because the republics are concerned that they may end up scrimmaging with one another, the changes they are making in the existing military infrastructures are mostly cosmetic. These halfhearted conversions are bound to be as successful as trying to open a fast-food restaurant from the kitchen of a soufflé chef, economists argue. "It's possible to convert people; it's much harder to convert institutions," points out Jeffrey B. Miller, an economist at the University of Delaware. Instead, he suggests, leaders should break up the old institutions, jettisoning much of the hardware and relying on the flexibility of the work force to power the move into commerce.

Commercializing former Soviet weapons factories holds great promise—and tough challenges.

Yet shuttering munitions plants to develop altogether new ventures such as tourism will be difficult. Soviet factories have traditionally provided not only jobs but often homes, foodstuffs and social activities. Few citizens are psychologically prepared to accept the idea of closing factories, Cooper adds.

Workers' fears are often justified. Because of housing shortages, they cannot casually quit their towns to search for work elsewhere. No wonder, then, that even the early efforts at *konversiya,* or conversion, have been dubbed *konvulsiya.* "Factory managers frequently observe that their enterprises have 'fallen under conversion' in the same way as one might speak of falling under a bus," Cooper says.

Further retarding efforts to reshape weapons producers into commercial ventures are the swirls of changes in other political and economic structures. Funding for research and development is in disarray. Top-flight scientists and engineers are heading for other countries, some possibly offering to sell their weaponry skills to the highest bidder. Critical subsystems and materials for producing consumer goods are scarce. Then there are the more pervasive problems: food shortages and a fragile confederacy of republics.

Yet the tattered arms producers may still hold some of the best hopes for catapulting Russia and the republics into the commercial fray. Workers in the military infrastructure have traditionally been the best trained and the highest paid of the Soviet work force. Many of them are eager to get into business to preserve—or improve on—their standard of living. As these people finally gain the authority to make decisions about production, new resources will be freed up, economists predict.

Bright spots are already emerging. In the absence of formal markets for goods, a host of barter exchanges is springing up, Cooper reports in his recent book, *The Soviet Defence Industry* (published by the Council on Foreign Relations). For instance, the former Soviet Space Flight Control Centre helped to set up the Russian Commodity and Raw Materials Exchange by providing the exchange with the dependable communications system and the spare facilities originally built to control the flights of the *Buran* space shuttle.

Foreign investors, particularly from South Korea, are beginning to lend funds and insight. Establishing separate currencies for each republic could also help stabilize local economies, Cooper suggests. "At least they'll have to stop blaming one another for their problems," he comments.

Cooper even spies the glimmer of a silver lining in the scientific exodus. If those who leave maintain ties with their colleagues at home, they can help integrate the remaining Russian scientific community into the international mainstream. Financial peace dividends are still too puny to show up on national balance sheets, but even that may eventually change as governments substitute more economics lessons for defense studies. —*Elizabeth Corcoran*

Seeing Infrared

We can feel heat and see its effect on the atmosphere—a mirage, for example. We can also detect hot objects in another way—from the infrared radiation they emit. Most infrared sensors are sophisticated, solid state devices [see "Infrared Video Cameras," by Jerry Silverman, Jonathan M. Mooney and Freeman D. Shepherd, page 78]. Recently I developed a way to build a simple infrared camera using a concave spherical mirror and a liquid crystal. Although my camera is not nearly so sensitive as state-of-the-art devices, it can view the infrared radiation from a hot object that is several feet away.

To test the design of my camera, I started with a heat lamp, a shaving mirror and a liquid crystal from a "mood sensor." I used the mirror to focus the radiation from the lamp onto the liquid crystal. Within a second or so, an image appeared on the liquid crystal. It bloomed rapidly into a large circle. Unfortunately, I could not tell whether the liquid crystal was responding to infrared radiation or to the heat generated when the crystal absorbed the visible light from the lamp.

Nevertheless, the experiment encouraged me to build a more sophisticated camera. I purchased a high-quality mirror and a set of liquid crystals that had good temperature sensitivity. (These items can be ordered from a science supply catalogue.) I also bought hardware to build a rigid support to hold the mirror and the liquid crystals. The materials cost a total of about $250. For comparison, commercial infrared cameras sell for thousands of dollars.

The most expensive part of my apparatus is the mirror (about $150). Its quality and size ultimately determined the resolution and sensitivity of the camera. I settled on a gold-coated mirror that was six inches in diameter and had a focal length of 12 inches. The gold coating reflects much of the infrared radiation while absorbing some of the visible light.

I chose the size and focal length of the mirror by applying basic principles of optics. I first calculated where the image would be focused in relation to the mirror. The image distance is equal to the following expression:

$$\frac{\text{focal length} \times \text{object distance}}{\text{object distance} - \text{focal length}}$$

Hence, a mirror with a focal length of 12 inches would focus an object 36 inches away to a point 18 inches in front of the mirror.

Next I figured out the magnification of the mirror, which is equal to the image distance divided by the object distance. When the object distance is 36 inches and the image distance is 18 inches, the magnification is 0.5—that is, the image will be 50 percent smaller than the object. I should also point out that the image will appear upside down.

The six liquid-crystal sheets, which together cost about $30, were each six inches wide, 12 inches long and 0.008 inch thick. Each sheet was sensitive to a different range of temperatures (as low as 20 to 25 degrees Celsius and as high as 40 to 45 degrees C). Eventually I realized that the sheets sensitive to low temperatures work best in a cool room, whereas those sensitive to higher temperatures are more appropriate for warm conditions.

All the liquid-crystal sheets are made of Mylar coated with two layers of ink. The inner layer is liquid-crystal ink, and the outer layer is black ink. The liquid-crystal ink is a mixture of three different compounds that contain cholesterol [see "Liquid Crystals," by James L. Fergason; SCIENTIFIC AMERICAN, August 1964]. The liquid-crystal ink is encapsulated so that the sheet does not leak ink when it is cut.

To make the support structure, known as an optical bench, I first joined two boards at an angle of 90 degrees, thereby creating a wide track [*see illustration on opposite page*]. At one end of the track, I then attached a small board to support the spherical mirror. Next I cut a triangular block that could slide along the track. In the top of the block, I drilled a hole to accommodate a threaded insert. The insert mated with a threaded brass post, which I had attached to a half-inch length of copper pipe one inch in diameter. Finally, I cut a piece of the liquid-crystal sheet to fit on the end of the pipe.

By sliding the block within the track, I could move the liquid crystal in relation to the mirror and therefore control the focus of the camera. The height of the liquid crystal could also be adjusted slightly by screwing the pipe assembly more or less tightly into the threaded insert. (The coupler is used as a locknut when the desired height is achieved.)

I originally designed the apparatus so that infrared radiation would reflect off the mirror and directly illuminate the liquid-crystal ink. I discovered that I could obtain much better results by directing the radiation onto the black surface of the liquid crystal. The black ink presumably absorbs the infrared energy and heats up the liquid-crystal ink. As the ink is heated, it changes color.

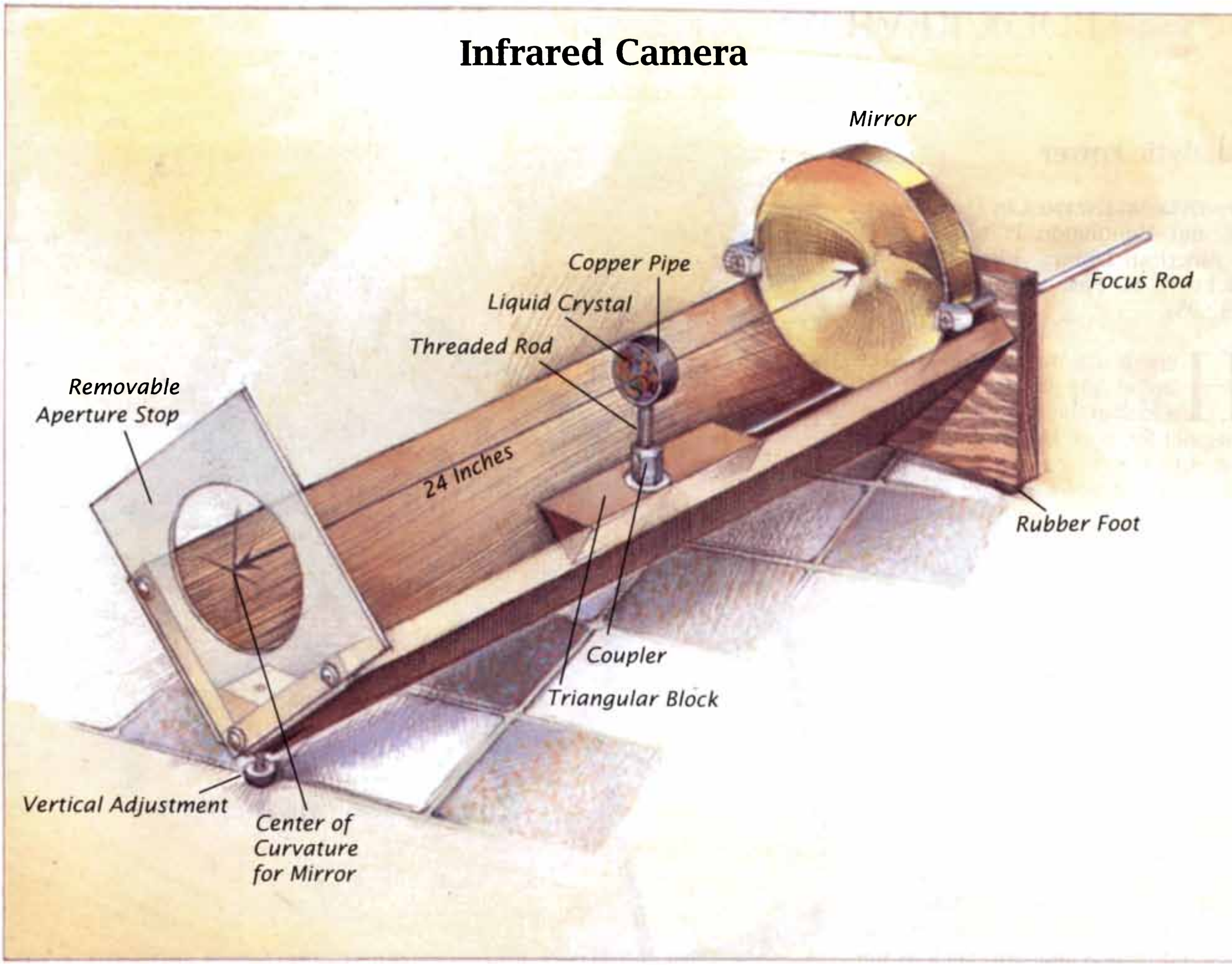
To test the apparatus, I searched my home for sources of infrared radiation. I tried a hot teakettle, a 250-watt heat lamp, the burners on an electric stove and a bright flashlight. (I should warn all who attempt these experiments that they can burn themselves severely if they mishandle various sources of infrared radiation.)

The liquid crystal did not register the teakettle with boiling water in it. When exposed to the heat lamp, the camera displayed several hot spots that bloomed rapidly. A red-hot burner swamped the crystal instantly. The liquid crystal did detect the radiation from the flashlight, and as I swung the light around, I could draw patterns on the crystal.

Although I was excited by these results, I worried that the liquid crystal might be detecting visible light instead of infrared radiation. To prove that this was not the case, I decided to use a piece of silicon, which absorbs visible light but only attenuates the intensity of infrared radiation. I placed a silicon chip that was four inches in diameter and 0.02 inch thick in front of the aperture of the camera. With this setup, I tested several sources and found that the images still appeared and, as expected, were somewhat dimmer.

Eventually I discovered that a steam iron was the ideal source of infrared ra-

DONALD G. MOONEY is a senior lens designer at Rockwell International and has designed optics for infrared systems for the past 30 years. His first engineering assignment was to develop an optical system for the *Apollo* spacecraft. He is the father of Jonathan M. Mooney, one of the authors of "Infrared Video Cameras," page 78.



diation for my experiments. I could set the temperature of the iron fairly accurately, within a range of 50 to 250 degrees C. (The temperature of the iron could be measured adequately with a deep-fry kitchen thermometer.) The iron was a good source for another reason: its bottom was made of Teflon, which when heated readily emits infrared radiation.

I had a few problems with the steam iron. When tilted for an extended period, it turned off automatically, and other features did not work at all. After convincing my wife we should buy a new iron, I dismantled the old one and modified it as a test target. I spent some time experimenting with the iron and my infrared camera. The liquid crystal displayed a clear image, 0.75 inch high, when the iron was 12 feet away.

I then set out to determine the resolving power of my camera. In front of the steam iron, I positioned a plastic spatula that had four slots, each four millimeters wide. I reasoned that the spatula would block some of the infrared radiation and therefore form an image consisting of several bars. It took quite some effort to produce this image.

I discovered I needed to focus the image very precisely. To do this, I illuminated the spatula from behind with a flashlight. I could then observe the visible light from the flashlight on the black side of the liquid crystal. I focused the image so that I could clearly see the shadow of the spatula. (Because the infrared camera is a reflecting system, the visual and infrared radiation come to the same geometric focus.)

After I replaced the flashlight with the steam iron, I put a piece of cardboard in front of the infrared camera to block the radiation. I then heated the liquid crystal with the flashlight and allowed it to cool. The liquid crystal is most sensitive at the instant the crystal turns black. At that point, therefore, I removed the cardboard, and the image of the spatula's shadow could be seen for a second or two.

To record the images produced on the liquid crystal, I used a conventional camera. But I had some difficulty bringing the camera lens close to the liquid crystal. For this reason, I mounted a mirror on the support structure so that the image on the liquid crystal could be seen from the side of the apparatus. This arrangement allowed me to position the camera a few inches away from the image.

I have no doubt that other amateur scientists will find ways to improve the sensitivity and range of the infrared camera. I hope they will enjoy, as I have, observing the world in a new light.

For more information, please write to: The Amateur Scientist, Infrared Camera, Scientific American, 415 Madison Avenue, New York, NY 10017-1111.

FURTHER READING

OPTICS AND OPTICAL INSTRUMENTS: AN INTRODUCTION WITH SPECIAL REFERENCE TO PRACTICAL APPLICATIONS. B. K. Johnson. Dover Publications, 1960.

BASIC OPTICS AND OPTICAL INSTRUMENTS. U.S. Navy Bureau of Naval Personnel. Dover Publications, 1969.

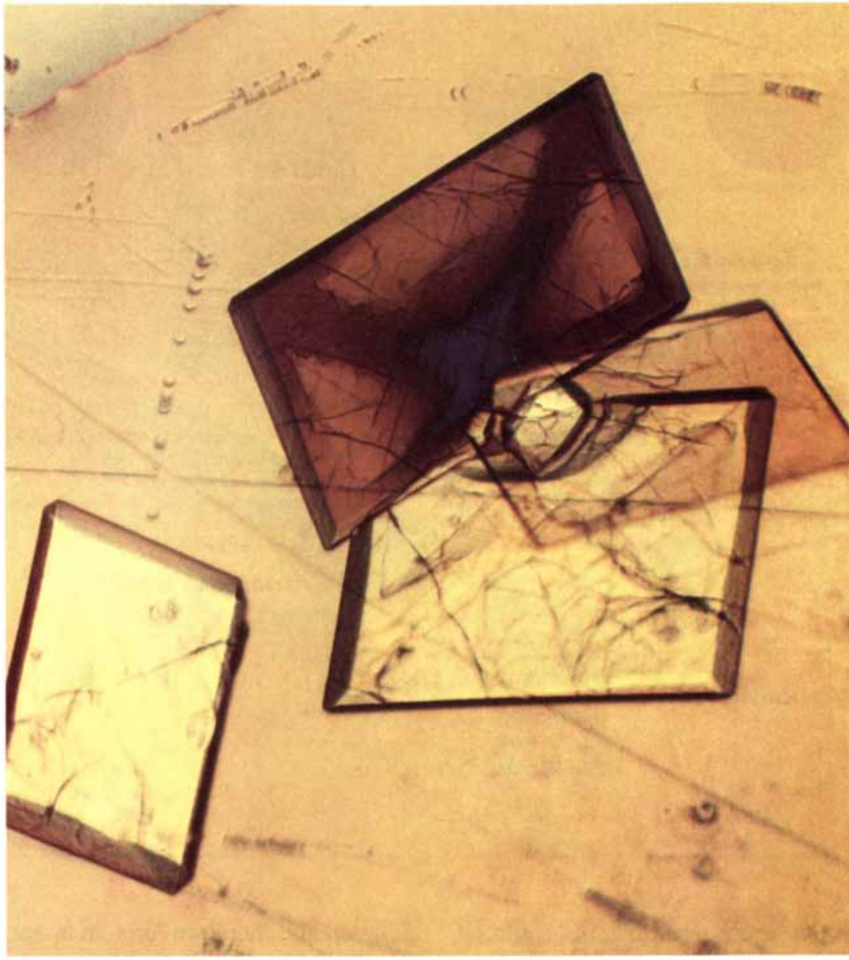
Catalytic Power

DISCOVERING ENZYMES, by David Dressler and Huntington Potter. Scientific American Library, 1991. Distributed by W. H. Freeman and Company ($32.95).

Here is an introduction to the broad powers of biochemistry, a look at molecular biology that does not focus on the fateful program coded in the cellular DNA. This book examines the business end, by which alone the life plan can be carried out. Its battery of molecular tools, jigs and dies are the enzymes, almost all of them proteins.

The first three of seven beautifully illustrated and smoothly written chapters, by two Harvard researcher-teachers, open the history of chemical experiment for the uninitiated reader. It was in 1835 that the great Swedish chemist Jöns Jakob Berzelius saw the possibilities for a chemical grasp of life's diverse nature. "We find solid reasons to assume that, in living plants and animals, thousands of catalytic processes occur within the tissues and fluids, generating a multitude of substances." He cited old results and new, such as the release of oxygen gas by gold and platinum metal from that new, weakly bound liquid, hydrogen peroxide. In all these changes the reaction-promoting additives were themselves unchanged, mere bystanders, yet strangely forceful. He named them catalysts. Even a generation later Justus von Liebig and Louis Pasteur were united in disbelief of such powerful molecular bystanders. Berzelius was once dismissed with the remark that to call a phenomenon catalytic was merely to replace a common word by a Greek one. "In the nineteenth century, chemists were cantankerous—and great chemists were greatly cantankerous."

By about 1900 catalysis was accepted, but enzymes were still a puzzle; their very efficacy makes them hard to analyze. An active enzyme preparation might contain irrelevant impurities that much outweighed the milligram of enzyme present, so that good analysts would report inconsistent results. Enzymes did not seem to be proteins. But in 1926 a young Cornell chemist, James B. Sumner, crystallized an enzyme for the first time. His small, uniform crystals were all protein with a little water; maybe the enzyme was hidden in the water portion? Ten years' work led to firm conviction: protein crystals, however separated or inactivated, always went in parallel with the catalytic power.



CRYSTALS of the digestive enzyme chymotrypsin: more is known about its structure and mode of catalysis than is the case for almost any other enzyme.

In the latter half of the book, the protagonist appears, not enzymes in general, nor even a long catalogue of them, but one specific digestive enzyme, the protein chymotrypsin. It is generated within the vertebrate pancreas and used in the intestine to begin breakdown of ingested protein molecules. The anatomy of chymotrypsin is presented here in a variety of artful mappings. The primary string of amino acid monomers was read out 20 years ago, by taking the links delicately apart step by step. (You must be alert; some enzymes are one long chain of links, but this one has several chains with side branches and loops. That, too, was learned by witty disassembly.)

The backbone of links wanders in three-dimensional space, a ball of yarn squeezed into an asymmetric atomic lump five-millionth of an inch across. "The ridges and grooves on the surface... are as unique as the mountains and craters of the moon." (This molecule joins 4,000 atoms, all in specified position, arranged into 241 amino acid links. Only complex x-ray diffraction analysis of pure crystals of the enzyme enabled this detailed 3-D mapping, as it did for helical DNA. Every single molecule of the stuff is almost the same, or else the thousands of diffraction spots would blur. (The text does not tell us enough about the limits to that precise identity; of course, the various elemental isotopes are set around at random.)

A marvelous experiment was carried out in the 1960s that illuminates the nature of specific enzyme form like a lightning flash. A sample of an enzyme, one rather like our hero, was placed into reagents able to break all the cross-links and undo all the internal folding, but not to sever the main links of the backbone chain. The protein was reduced to a mere random coil and lost all its enzymatic activity. Then the "denatured" protein was incubated within various near-natural environments, in an earnest effort to heal the poor injured thing by trial after trial. Finally, the therapy worked; the random coils of protein refolded, the loops were remade and full enzyme activity was regained. "No outside agency or force in the cell is needed." Error-free spelling is enough to yield the entire sculptured form and function.

All a catalyst can do is increase the likelihood of a given reaction; it cannot permanently add or take away energy, since it acts without lasting change in itself. Tolerable temperature changes will bring a modest increase in reaction rate, say by 10- or 100-fold, but chymotrypsin dissolves its target proteins at a rate one billion times faster than the spontaneous rate. Enzymes are the masters of time.

The catalytic specificity of enzymes is extreme; one enzyme, one reaction—or anyhow not many. Chymotrypsin is huge compared with its target protein threads; only at some active sites do contact and reaction actually take place. The tricky chemical logic of inactivation, substitution and competition has been able to locate just where the enzyme surface binds with its target molecule. The range of specificity arises from the surface form; the shape must keep out the wrong sites but admit the right ones. The fitting portion has a kind of springy atomic jaw that aligns the target link; the process is referred to as a lock-and-key mechanism. Chymotrypsin usually catches the charged end of the protein thread and severs the long thin chain. The parts move off, the spring resets and the unchanged enzyme molecule can act again.

The enzyme responsible for clearing away the neurotransmitter signal that swiftly bridges the synapses between neurons in the brain is related to chymotrypsin. It is strange and somehow moving that the terrible nerve gases act to bind and degrade these very enzymes. It is still stranger that the tiny, brownish deposits that form in the brain during Alzheimer's disease are pools of two proteins that are specific inhibitors, able to turn off the active sites of this decisive family of enzymes. The tools of life are joined in a profound geometric unity that sometimes goes ironically and tragically awry.

From Accretion to Zodiac

THE ASTRONOMY AND ASTROPHYSICS ENCYCLOPEDIA, edited by Stephen P. Maran. Van Nostrand Reinhold (in North America) and Cambridge University Press, 1992 ($89.95).

A thousand and two large pages make up this hefty single volume, its tall tales as filled with wonders as were the thousand and one nightly spinnings of Scheherazade, if here expressed in soberer, less personal accounts. The illustrations, all in black-and-white and most of them tightly packed graphs and diagrams, are given perhaps a tenth of the page area. (The dust jacket is ornamented by one big, colorful view of many-ringed Saturn seen by the Hubble Wide Field Camera in orbit.) There are about 400 articles, by a few more authors than that, just about all of them researchers at work today in a dozen countries of five continents. The alphabetized entries add up to a participant's collective summary of current astronomical interests, written for the most part at the level of the more detailed articles found in this magazine. There are a good many simple formulas (hardly two integral signs are to be found in the entire volume), with many tables of data and outlined scenarios, but few of the richer pedagogical aids frequent in *Scientific American.*

Sets of entries cluster around important current topics. For instance, almost 20 pages under the letter A treat "Active Galaxies and Quasistellar Objects." Those pages present nine concise pieces, each by an expert who writes on one lively subtopic of that field. The pieces address such topics as gas infall, jets, superluminal motion and x-rays, all as they are in active galaxies, and one fine essay sketches a unified view of the whole diverse class.

A slow, enjoyable browse was sufficient to encounter a few narratives high enough and brief enough to report. One article, written by Frank H. Shu of Berkeley and exceptional for its vivid style, opens with a sharp reminder: without the dense central furnaces of stars, there would be only one molecule in nature, hydrogen; all matter would be a structureless gas. Indeed, the main omission from the entire volume is a fuller account of that very understanding, the long-convincing evidence of the formation of the whole run of the elements, especially those heavier than iron, by identifiable processes *within the stars.* The subject is still unfinished, although this pot simmers slowly these days; hard-pressed editors have omitted the topic.

In "Moon, Origin and Evolution," H. Jay Melosh sets aside all the older views, that our moon is either the sibling, the spouse or the child of Earth. It is now held that the satellite was formed in a glancing collision between a Mars-size projectile and the new-formed Earth. The moon formed from the outer mantles of both partners; it took little central material from either, and so it lacks much of an iron core. A wonderful drawing, recently prepared by George L. Withbroe, now at NASA, shows the globe of the sun bristling with lines of magnetic force swept outward by the solar wind, the patchy surface pattern recognizable in an adjoining x-ray photograph made of the solar corona at the right time. Anthony P. Fairall carries deep cosmic mapping one long step outward; his maps identify the nearer galaxy superclusters and then go beyond to show in a preliminary way superclusters that lie a few times farther out than what we see in more familiar surveys.

These symposiumlike clusters of articles and many more individual articles, from accretion to zodiacal light, constitute the book. A certain staccato quality is inescapable; perhaps, too, there is less introductory matter than might be optimal. But the sense of engagement, authority and timeliness is unmatched. You are among astronomers talking today's astronomy, at a nontechnical level for those who will carefully attend.

There are household names here as senior advisers. But the 400 whose words we read are mainly a muster of the young and able, the men and women closest to current data and theory. (That is not to neglect a substantial fraction of notable, more senior contributors.) Neither rank beginners nor those who want to know about problems not now on the front burner are well served. This confident encyclopedia is authoritative for its day, for keen amateurs and writers, for students and teachers, and will itself in due time become a prime object for the historians who one day will assess the shortcomings of today's luminous insights.

Narrative Gravity

CONSCIOUSNESS EXPLAINED, by Daniel C. Dennett. Illustrated by Paul Weiner. Little, Brown and Company, 1991 ($27.45).

Decades back Alan Turing himself proposed a brilliant test for any claim that a machine might rival us conscious humans, a test not unfairly tilted by such inessentials as voice or appearance. He suggested searching interviews conducted by wire, to elicit a sequence of Q's and A's, all recorded on a printout, whereby a clever conversational computer might be compared line by line with a witty human control. That test appears fair and practical even now (a version was tried out lately in Boston). But today we could enrich the wire link, to admit exchange of two-way drawings, musical snatches and mathematical forms as well as plain-language text, for surely those are only texts in different symbol systems.

In this book, 500 pages of sustained and lively argument, a professional philosopher from Tufts University has spun an entire method of inquiry out of Turing's hint and used it to wonderful advantage. His sophisticated discourse is as savvy and articulate about good beer or the Boston Celtics as it is about parallel processing, modern cognitive experimentation, neuropathology, echolocation by bats, or Ludwig Wittgenstein. He is a rare advocate, willing to examine his own mistakes, cite and confirm or dispute fairly a long list of his contemporaries, and allay the commonsense doubts of nonspecialist readers with appropriate experimental results, crystal-clear analysis and vivid metaphor. He does all this with verve in a persuasive philosophical work, the best examined in this column for decades.

To understand the scheme he has built over 20 years, begin with the "fauna and flora" of conscious experience, thoughts and memories as well as "pains or aromas or daydreams." For Dennett does not ignore our evidence of mental experience, beginning with his own. The dark, old pitfall remains. How can we know the inner experience of others? Yet how can we hope to construct a testable and growing science out of subjective material alone? Dennett's answer is plainly akin to Turing's test: we accept as incontrovertible the varied textual accounts our conscious friends give of what they consciously feel and believe. That kind of account is as objective as tape recorder and print can be, and it is fully open to all critics. Its humanity is beyond doubt. Nearly as evident is the inference that animals lacking elaborate text-making ability can hardly provide us with the rich data we need to assess consciousness. A fine chapter on the evolution of consciousness adds the plausibility of continuity in time.

All accounts of mental life are data, but what narrators *conclude* about mental experience need not become part of our theories. Sincere and well-reported introspection yields valuable data (not answers), sources for extrapolations and eventual model making. The task of explaining the conscious mind is not unlike the task of literary criticism: we do not believe Sherlock Holmes really looked out at busy Baker Street, yet we can reasonably make sense and context for the stories. Every student will agree, even from mere secondhand accounts of the actual texts, that "Holmes was smarter than Watson; in crashing obviousness lies objectivity."

In the same way, the many convinced reports that witness every kind of mind stuff, the medium in which a purple cow is rendered, the "thinking thing" itself, the sources of love, hate and enjoyment, even moral responsibility, are only input for a serious theorist of consciousness. "The subject's... world will be... stable, intersubjectively confirmable... having the same metaphysical status as... Holmes's London." Those narratives tell what seems to be there, to be considered and assessed just as the Turing test answers must be.

What Dennett does most of all is to unpack carefully but vigorously the simple terms and explanations that we hold. Most narratives of mental life suggest a single agent of mind, single as the body. We all recognize the danger of a contradiction there; what goes on, in acts of speech or memory or even in the judgment of enjoyment, is more contingent, more diffuse in time and space, more ambiguous, than the pat descriptions we freely use. Just as the organized termite colony has no Soul but rather a set of inborn rules by which every one of 100,000 little agents does its thing, so "our human consciousness, and our narrative selfhood," is more the product of all that we constantly tell ourselves and others than it is the single source.

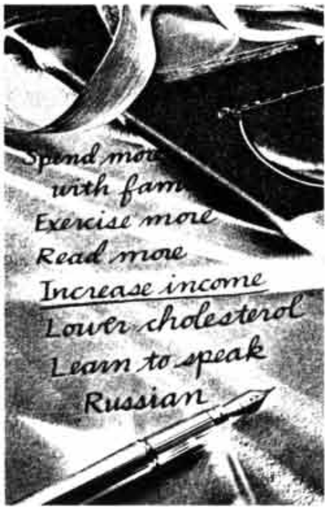
The Self is a grand abstraction, like the physicists' center of gravity. Any mechanical system moves in such a way that a certain point (it is the average position of all the constituent masses) moves with striking and predictable simplicity, however the rest of the weights, springs, gears and shafts tumble, spin and shake. There need be no matter at all at the position of the center of gravity. So we may argue for the Self. It is no point of mind, but a great abstraction "defined by the myriads of attributions and interpretations... that have composed the biography of the living body whose Center of Narrative Gravity it is."

The stream of consciousness is no single flow, not even a Joycean one, but a turbulent, gappy, much-branching stream whose full texture we cannot sense or recall. Many cognitive experiments, much observation of damaged minds, inform that view. What is dominant in forming that consciousness is a noisy chorus of specialized, even competing, elements of mind, "feature detectors," logically much elaborated from the simplest model—wired retinal elements that respond, for example, to a T-shaped feature in the visual field but not to a bright point. Choices among acts of decision, often squeaky judgments, influence future ones, to build that abstraction of a Center, "just as other structures keep track of information on Boston, or Reagan, or ice cream."

"The trouble with centers of gravity is that they aren't real; they're theorists' fictions," the author's built-in critical voice objects. But the rejoinder is brilliant: "That's not the trouble with centers of gravity; it's their glory. They are *magnificent* fictions," like the fictional characters in our best novels.

Can a mere machine, then, ever form a conscious self? The author says a resounding yes. What sort of silicon "person" will we know one day? Professor Dennett has only made a beginning with his wise, conjectural description. His is brave, knowledgeable discourse, neither proof nor example; his title might well have worn a question mark. He plainly hopes and expects such conscious machines to come, probably not singly, but in their families and bands. None of them will be "artificial intelligences," patient, logical von Neumann computers, serially working out swift inference on inference infallibly forever. They are more likely to resemble fuzzily parallel computers, using probabilities to open heaps of uncertain gates, learning by corrected experience as much as by design—not easily predictable, all too human! No single machines have yet had much of what is needed for self-surveillance and multiple computations wired in. That time will come.

Skeptics, supporters and the undecided should proceed at once to find and read this good-humored, imaginative, richly instructive book.



WOMEN AND AIDS by Kenneth H. Mayer and Charles C. J. Carpenter

Although most of the Americans living with AIDS and the virus that causes it are male, the epidemic is rapidly spreading among women. Several major urban hospitals in New York and Florida report that more than five percent of child-bearing women are infected with the human immunodeficiency virus (HIV). Screening of recent entrants into the Job Corps and the military demonstrates near parity between the numbers of HIV-infected men and women. On a global scale, among the more than 10 million people infected with HIV, approximately one in three is female. In the U.S., of approximately one million infected people, one in eight is female. But in several eastern states the distribution of cases is approaching the global ratio.

As is the case with several other sexually transmitted diseases, women are more susceptible to acquiring HIV from an infected male partner than the other way around. Their biological vulnerability is compounded by their frequent lack of awareness of their partner's HIV status, which often leads to a misperception of their personal risk. Women are therefore less likely than men to know they are infected until they manifest symptoms. And even when they do show symptoms, physicians are often slow to diagnose HIV infection.

Infected women are also more likely than men to be from poor communities. They are therefore less likely to have access to therapy that has been demonstrated to prolong life—for example, prophylaxis against *Pneumocystis carinii* pneumonia and antiretroviral medication.

These biological and social realities have combined to produce a marked underrepresentation of women in clinical research related to HIV. Of these factors, poverty may be the most significant. Many women with HIV infection are the sole providers for their children, and they confront a health care system that may often be insensitive to their specific concerns, such as child care, transportation costs and culturally appropriate services.

As several studies have documented, when HIV-infected women from communities of color who meet clear clinical indications for receiving zidovudine (formerly named azidothymidine, or AZT) or *Pneumocystis* prophylaxis are compared with white homosexually active males, a much lower percentage of the women have been offered the treatments. Many HIV specialists have not developed extensive expertise in the care of women, although more than 100,000 women in the U.S. are infected with HIV.

Regional differences in the clinical signs of advanced HIV-related immunodeficiency have been shown to vary around the world, depending on the prevalence of local opportunistic pathogens as well as nutritional and immunosuppressive factors. Gender-related differences, however, have not been fully delineated. Furthermore, many federal and state programs are predicated on a diagnosis of AIDS based on criteria developed at a time when very little was known about the natural course of HIV infection in women.

Nevertheless, some studies are beginning to elucidate the progression of the disease in women. In Rhode Island, we have observed that recurrent vaginal infections with *Candida albicans* have been the first clinical signs in over one third of a group of more than 200 women. Investigations at Downstate Medical Center in Brooklyn, N.Y., have found cervical tumors in many women who are infected with both HIV and human papillomaviruses (HPVs). Many researchers have also noted a high prevalence of recurrent and recalcitrant genital warts (which are caused by certain strains of HPV) in HIV-infected women.

The oncogenic interactions between HPV and HIV need careful study. So does pelvic inflammatory disease. It is not known whether women with HIV infection are more likely to acquire this condition and, if they do, whether they are less likely to respond to standard therapy. Failure to recognize symptoms specific to women may delay diagnosis and thus the ability to monitor a woman's immune status and offer appropriate therapies.

Even after showing signs of an HIV-associated illness that meets the current criteria for the diagnosis of AIDS, a woman may encounter other hurdles that keep her from receiving promising experimental therapies. Many research protocols mandate routine pregnancy tests and eliminate women from trials if they become pregnant. Women may consequently be forced to choose between pregnancy and a potentially effective treatment. Many of them may have personal convictions that do not permit the termination of pregnancy. In an ongoing study of the heterosexual spread of HIV in southeastern New England, we have found that more than one third of HIV-infected women anticipated becoming pregnant.

The conflict between reproductive choice and therapeutic options may represent remnants of a paternalistic era; few research trials dictate the responsibility of males with regard to insemination of female partners while the men are participating in studies. And the effects of drugs on semen are not routinely evaluated. Reticence about including women with reproductive potential in tests has been linked to the considerable body of corporate law regarding liabilities for drug-induced birth defects. The situation has become even more complicated because certain drugs, such as zidovudine, are now being studied to determine whether they might decrease the rate of HIV transmission from mother to fetus.

Scientists are increasingly recognizing the need to learn more about HIV-specific illnesses in women, and they are trying to incorporate women's concerns in the design of drug trials and patterns of care delivery. But a great deal still needs to be done to address the inequities facing women infected with HIV. In certain parts of Africa, the heterosexual epidemic has been called grandmother's disease, because the older generation assumes responsibility for caring for increasing numbers of children whose mothers have died of AIDS. As the demographic pattern of the AIDS epidemic in the U.S. more and more resembles the African experience, we must move quickly to learn the lessons from Africa so that we can respond in a sensitive and forthright manner to minimize both the societal and the medical havoc caused by this lethal virus.

KENNETH H. MAYER is director of the Brown University AIDS program and chief of the Infectious Disease Division at Memorial Hospital of Rhode Island. CHARLES C. J. CARPENTER is associate director of Brown's AIDS program and professor of medicine at the university.

WE COULD BUILD AN AUTOMOBILE WITH AN EVEN SMOOTHER RIDE, BUT THAT WOULD MAKE US UNCOMFORTABLE.

It seems the relentless pursuit of the plushest possible driving experience has blinded some of the world's carmakers to a very important point:

If you completely insulate the driver from the realities of the road, you've effectively removed his ability to contend with the realities of driving.

First and foremost, a driver must always remain in control of his vehicle. And the false luxury of an ocean-liner-like ride should never come at the expense of a driver's safety.

It is exactly this most critical, most fundamental issue — enhancing driver safety — against which the engineering decisions at BMW are measured.

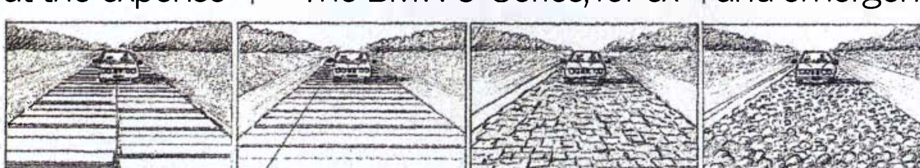
◆ Coil springs and shock absorbers are mounted separately to the chassis for improved ride comfort on uneven roads.

AN AUTOMOBILE MUST COMMUNICATE WITH ITS DRIVER.

The ability of a car to hold the road — and thus, the ability of a driver to maintain control over his automobile — are largely dictated by the car's suspension system. Which is why at BMW — a company long recognized for its achievements in engine technology — every new automotive design is begun not from the engine down, but rather, from the suspension system up.

The BMW 5-Series, for example, offers a perfect case in point. Both the 525i and 535i are equipped with highly developed, internationally patented front and rear suspension systems so refined, "the communication with the driver is perfect" according to the editors of Sports Car International.

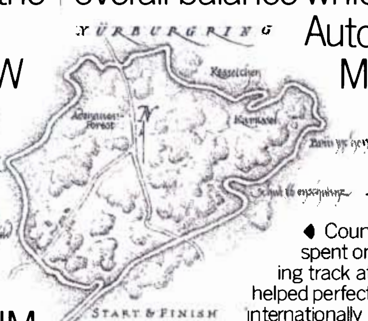
Its acoustically decoupled front suspension employs a unique double-pivot system with two independent lower control arms rather than the conventional single-arm design. This increases stability, minimizes dive, and provides improved control and feel under all driving conditions. While the track link rear suspension system, also an exclusive BMW design, assures crisp, predictable handling in both cornering and emergency situations.



◆ To achieve its outstanding ride quality, the 5-Series is tested and refined on all types of road surfaces.

The extremely rigid steel body of the 5-Series enables the suspension to operate more accurately from a firmer platform (while also optimizing passenger protection in the

* See your BMW dealer for details on this limited warranty. ** Services provided by Cross Country Motor Club, Inc., Boston, MA 02155, except in CA, where services are provided by Cross Country Motor Club of Calif., Inc., Boston, MA.


NÜRBURGRING
Kesselchen
Karussel
START & FINISH