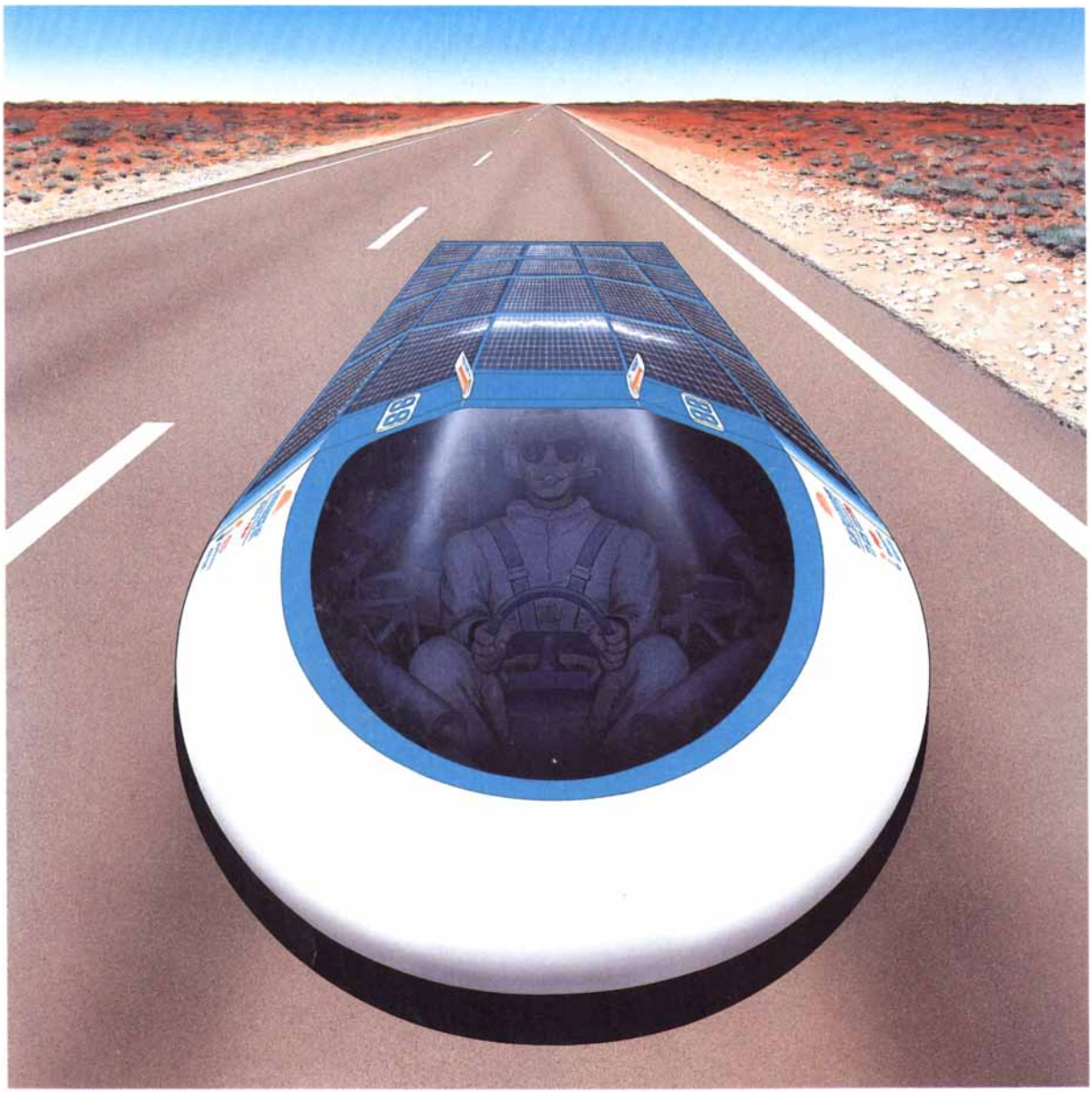
SCIENTIFIC AMERICAN

MARCH 1989
$2.95

The shuttle is flying, but the U.S. space program is not.

A molecular pump enables tumors to resist chemotherapy.

Accelerators that promise more power than the Supercollider.



***Solar-powered car** wins a race—and points the way toward practical electric vehicles.*

0 74820 08715 03

"My favorite GM car I worked on was the '54 Buick. Got one for myself and kept it all these years, even after I retired. Every time my son looks at it, he smiles. Very popular at the time, I tell him.

My boy's at GM now, building beauties like this '89 Olds. And though most things have changed, one thing hasn't–we still build 'em to last."

Fact: GM cars have held their resale value better than any other U.S. make based on average value of 3-to-5 year old cars over the past nine years.

GM

MARK OF EXCELLENCE

Chevrolet, Pontiac, Oldsmobile, Buick, Cadillac, GMC Truck

SCIENTIFIC AMERICAN

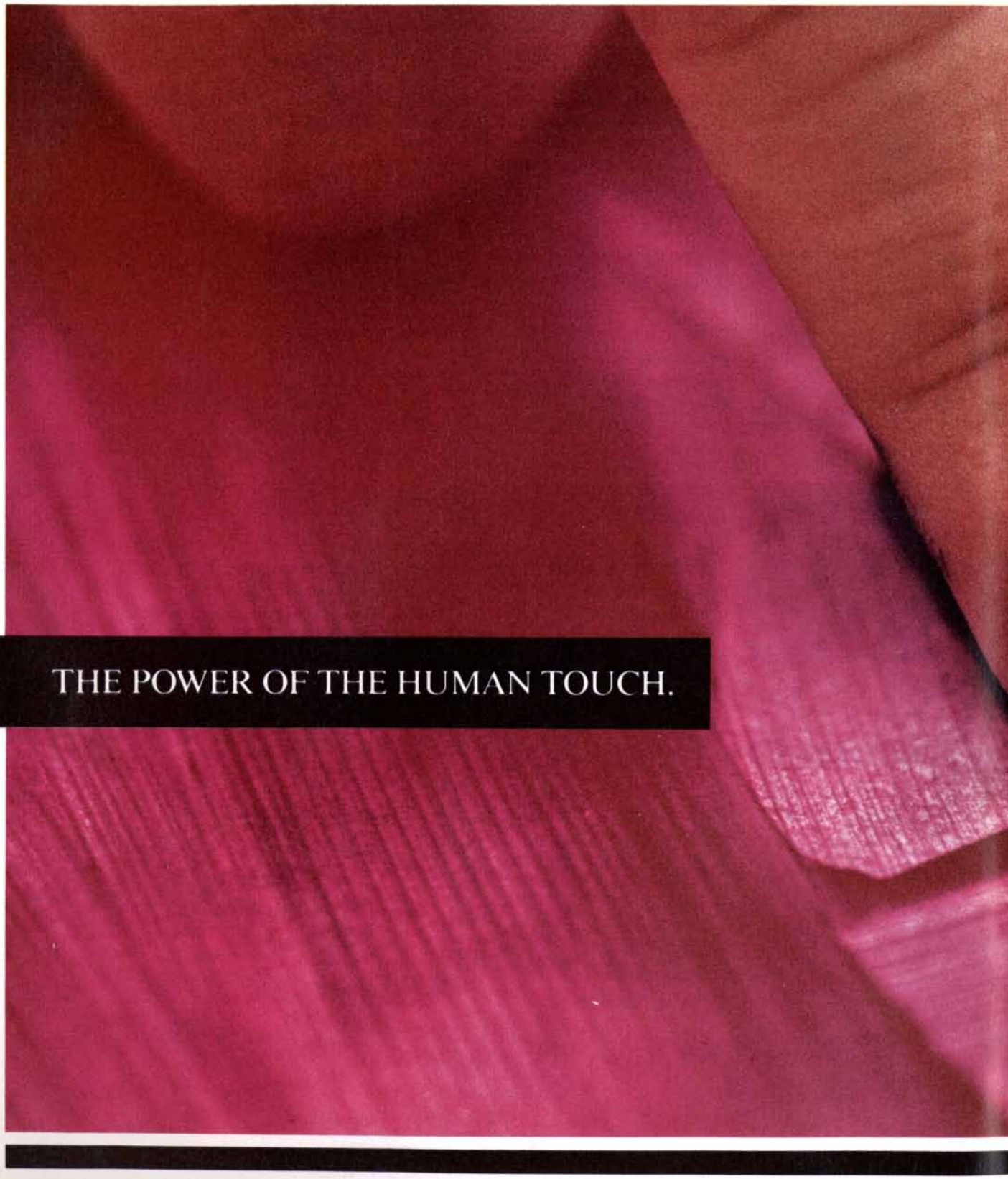
March 1989 Volume 260 Number 3

DEPARTMENTS

Imagination. Creation. Application.

At GTE, we give you the power of people. The incredible human energy that powers new ideas into realities. The dedication behind compassionate, effective service.

In emergencies like the October, 1987, California earthquake, you can count on our people to respond immediately and work tirelessly to restore the telephone service so comforting to families, so vital to

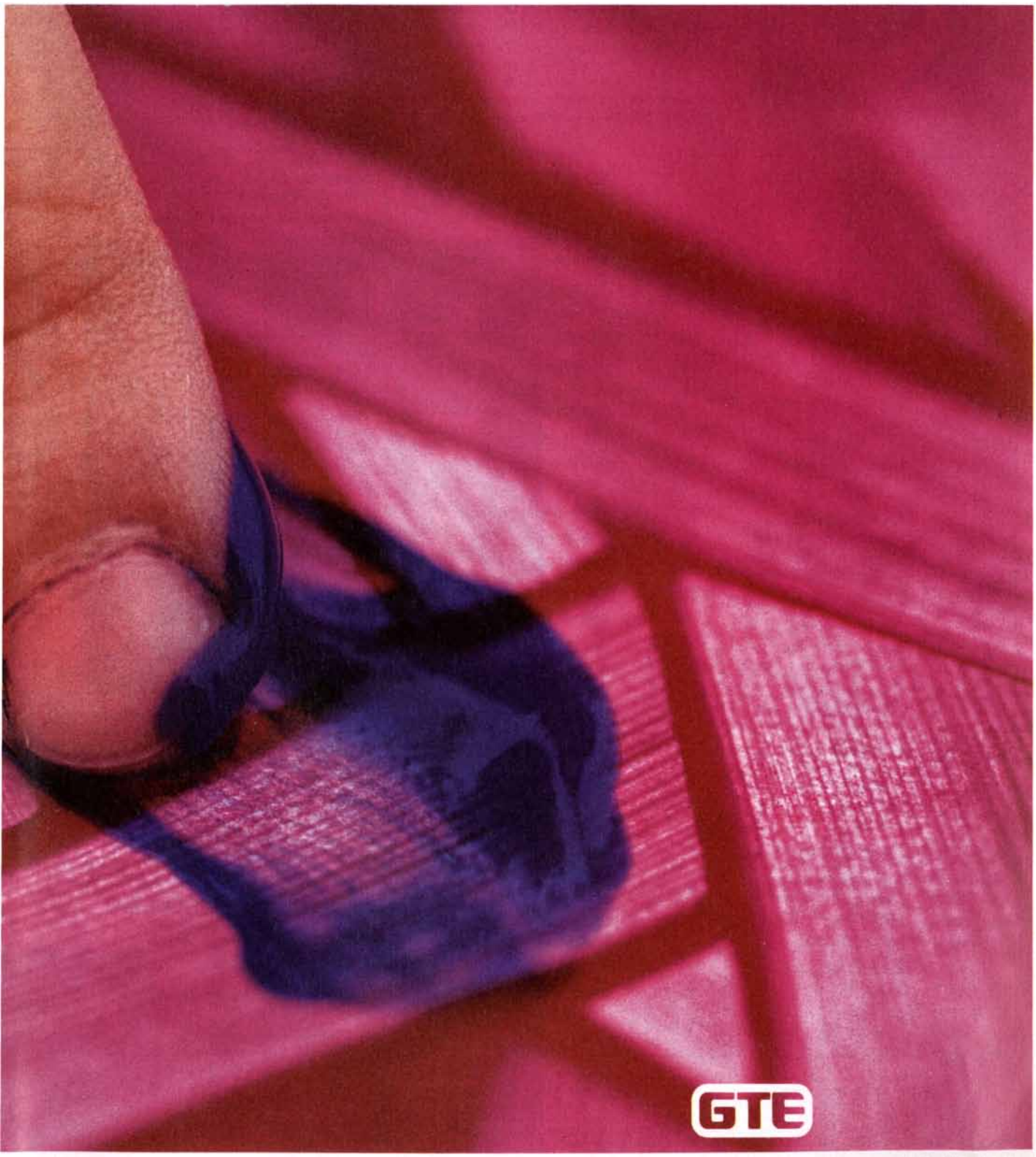


businesses. Right now, our Sylvania lighting people are working closely with Therakos, a Johnson & Johnson company, to develop systems for a new technology utilizing light-activated drugs to treat a certain type of cancer.

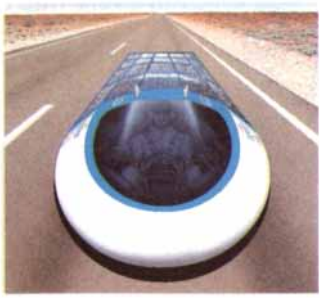
One to one, person to person, GTE to you.

Not just in these instances, but in every one of our companies, at GTE our people are our power.

And at GTE, the power is on.



THE POWER IS ON



THE COVER painting shows a head-on view of *Sunraycer* cruising across the Australian desert during the Pentax World Solar Challenge race (see "Lessons of *Sunraycer*," by Howard G. Wilson, Paul B. MacCready and Chester R. Kyle, page 90). Powered by 9,500 solar cells, the car won the 1,867-mile Darwin-to-Adelaide race on an amount of energy equivalent to five gallons of gasoline. The Plexiglas canopy was coated with gold film, which kept the driver cool under the blazing desert sun.

THE ILLUSTRATIONS

Cover painting by Hank Iken

Page	Source
34-35	Gabor Kiss
36	Johnny Johnson
37-39	Gabor Kiss
45-50	Ian Worpole
51	Grace Bradley, Ontario Cancer Institute and University of Toronto
55-57	George V. Kelvin
58	John M. Dawson
59	George V. Kelvin (*top*); Thomas C. Katsouleas, University of California, Los Angeles (*bottom*)
60	George V. Kelvin
61	James D. Simpson, Argonne National Laboratory
62-63	Jon Brenneis (*top*), Alexander McPherson (*bottom*)
64-65	Andrew Christie (*top*), Alexander McPherson (*bottom left and right*), Jon Brenneis (*bottom middle*)
66	Alexander McPherson
67	Jon Brenneis (*top*), Alexander McPherson (*bottom*)
68-69	Alexander McPherson
75	Brian J. Skinner, Yale University
76-77	Joe Lertola
78	Johnny Johnson (*top*), Joe Lertola (*bottom*)
79	Joe Lertola
80	H. Hughs, Four By Five
82	Courtesy of National Portrait Gallery, London
84	Johnny Johnson
85	Neil O. Hardy
86	Judith L. Rapoport
87	The Bettmann Archive
88	Neil O. Hardy
89	Judith L. Rapoport
90-91	Hank Iken
92	Andrew Christie
93	Laurie Grace
94	General Motors Corporation
95	Laurie Grace
96	Courtesy of Howard G. Wilson (*top*), Laurie Grace (*bottom*)
98-99	Peter Frey, The Image Bank
100	Tom Prentiss
101A	George F. W. Hauck
102-102B	Tom Prentiss
104	George F. W. Hauck (*left*), Tom Prentiss (*right*)
107-109	Michael Goodman
111	Andrew Christie
112	Johnny Johnson

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017. Second-class postage paid at New York, N.Y., and at additional mailing offices. Authorized as second-class mail by the Post Office Department, Ottawa, Canada, and for payment of postage in cash. Subscription rates: one year $24, two years $45, three years $60 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. only 800-333-1199; other 515-247-7631/32. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51593.

SCIENTIFIC AMERICAN

Established 1845

EDITOR: Jonathan Piel

BOARD OF EDITORS: Armand Schwab, Jr., *Managing Editor;* Timothy Appenzeller, *Associate Editor;* Timothy M. Beardsley; John M. Benditt; Laurie Burnham; Elizabeth Corcoran; Gregory R. Greenwell; John Horgan; June Kinoshita; Philip Morrison, *Book Editor;* Tony Rothman; Ricki L. Rusting; Russell Ruthen; Paul Wallich; Karen Wright

ART: Samuel L. Howard, *Art Director;* Murray Greenfield, *Associate Art Director;* Edward Bell, *Assistant Art Director;* Johnny Johnson

COPY: Sally Porter Jenks, *Copy Chief;* M. Knight; Michele S. Moise

PRODUCTION: Richard Sasso, *Vice-President Production and Distribution; Managers:* Carol Eisler, *Manufacturing and Distribution;* Carol Hansen, *Electronic Composition;* Leo J. Petruzzi, *Manufacturing and Makeup;* Carol Albert; Madelyn Keyes; William Sherman; Julio E. Xavier

CIRCULATION: Bob Bruno, *Circulation Director;* Lorraine Terlecki, *Business Manager*

ADVERTISING OFFICES: NEW YORK: Scientific American, 415 Madison Avenue, New York, NY 10017; Robert F. Gregory, *Advertising Manager;* Kate Dobson, *Advertising Manager;* Lisa Carden; John Grant; Meryle Lowenthal. CHICAGO: 333 N. Michigan Avenue, Chicago, IL 60601; Patrick Bachler, *Advertising Manager;* Litt Clark, *Midwest Manager.* DETROIT: 3000 Town Center, Suite 1435, Southfield, MI 48075; William F. Moore, *Advertising Manager;* Edward A. Bartley, *Detroit Manager.* CANADA: Fenn Company, Inc. DALLAS: Griffith Group. PRINCETON: William Lieberman, Inc. WEST COAST: Frank LoVerme & Associates

ADVERTISING SERVICES: Laura Salant, *Sales Services Director;* Diane Greenberg, *Promotion Manager;* Ethel D. Little, *Advertising Coordinator*

INTERNATIONAL: FRANKFURT, GENEVA, LONDON, PARIS: Infopac. HONG KONG/SOUTHEAST ASIA: C. Cheney & Associates. SEOUL: Biscom, Inc. SINGAPORE: Cheney Tan Associates. TOKYO: Nikkei International, Ltd.

ASSOCIATE PUBLISHER/BUSINESS MANAGER: John J. Moeling, Jr.

PRESIDENT OF MAGAZINE DIVISION AND PUBLISHER: Harry Myers

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

PRESIDENT AND CHIEF EXECUTIVE OFFICER: Claus-Gerhard Firchow

EXECUTIVE COMMITTEE: Claus-G. Firchow; *Executive Vice-President and Chief Financial Officer,* R. Vincent Barger; *Senior Vice-President,* Harry Myers; *Vice-Presidents:* Linda Chaput, Jonathan Piel, Carol Snow

CHAIRMAN OF THE BOARD: Georg-Dieter von Holtzbrinck

CHAIRMAN EMERITUS: Gerard Piel

TWA's Business Class. More cities. More flights. More often.

If you like flying business class—and who doesn't—TWA's Ambassador Class® was made for you. It's the only business class with over 700 domestic departures each week. And it's the only one that flies to more than 18 major U.S. business centers.

You'll enjoy all the amenities you've been missing on other airlines. A separate Ambassador Class cabin, so you can work or sleep, free of distractions. Seats so wide, with legroom so ample, they rival first class seats on many other airlines. Complimentary cocktails and headsets. Meals served on fine china and linen.

With TWA's Ambassador Class, you're treated to more service, more comfort, and more convenience. A lot more often.

TODAY'S TWA.
FIND OUT HOW GOOD WE REALLY ARE.℠

TWA

LETTERS

To the Editors:

Chiye Aoki and Philip Siekevitz ["Plasticity in Brain Development," SCIENTIFIC AMERICAN, December, 1988] raise the important question of why the brain is "plastic" (susceptible to environmental influences) only in early infancy and not in adulthood.

If one eye of a young kitten is briefly deprived of vision ("monocularly deprived"), that eye becomes functionally disconnected from the brain, so that most cells in the visual cortex become dominated by the experienced eye. Normally this happens only during the first two or three months of life—the so-called critical period. But Max S. Cynader has shown that if kittens are reared in total darkness, their visual system can be kept plastic even for two years; if an adult dark-reared cat is monocularly deprived for a short period, cells in the cortex become dominated by the experienced eye just as they would in a kitten. More recently Baruch Kupperman and I were able to completely *reverse* the effects of monocular deprivation in adult dark-reared cats by simply closing the experienced eye and reopening the deprived eye. The originally deprived eye now "takes over" cells in the cortex and begins to dominate their activity. The curious implication is that the visual cortex of a dark-reared cat is almost as plastic as that of a young kitten. Such experiments have obvious relevance to amblyopia ("lazy eye"), which affects 5 percent of U.S. children.

Since dark-rearing can keep the visual system plastic indefinitely, we can conclude that visual experience itself is required for terminating the critical period. Aoki and Siekevitz have identified a protein, MAP2, that may be involved in mediating plasticity during the critical period. They offer the intriguing hypothesis that the protein can mediate plasticity only in a phosphorylated state and that exposure to light during infancy dephosphorylates the protein and terminates the critical period. They suggest that even a brief bout of visual experience in early infancy can act as a "switch" and set in motion a biochemical cascade that terminates the critical period.

My experiments with Kupperman contradict the "switch" hypothesis. We gave four kittens normal visual experience for from three to five weeks *before* dark-rearing them for six months. When we monocularly deprived these adult dark-reared cats for short periods, we found their brains were "plastic" even though they had been exposed to light before the dark-rearing. These experiments suggest that light cannot simply function as a "switch" to trigger a sequence of events that terminate the critical period. The presence of continuous, ongoing visual experience for several months may be required for the critical period to "exhaust" itself and presumably to adequately dephosphorylate MAP2.

Our results do not rule out the basic hypothesis that there is a protein that mediates plasticity; they do imply that the notion of a simple light-dependent "switch" needs to be examined more carefully. Further experiments may reveal why the infant brain is so malleable whereas older brains (including those of some older scientists) are so notoriously resistant to forming new impressions.

V. S. RAMACHANDRAN

Department of Psychology
University of California, San Diego
La Jolla, Calif.

To the Editors:

Dr. Ramachandran's results help to remind us that many more exciting

questions remain to be addressed by biochemists. It would be interesting to examine whether MAP2 becomes progressively more dephosphorylated after longer exposures to light or after light-exposed animals are returned to darkness. Readers interested in details of dark-rearing effects might read a paper by George D. Mower and William G. Christen (*Journal of Neurophysiology,* Vol. 53, No. 2, pages 572–589; February, 1985) proposing that light stimulus has different effects at different phases of cortical development. Drs. Ramachandran and Kupperman's results are consistent with this model, which predicts that one month of abnormal rearing would allow plasticity to reach its peak and the subsequent dark-rearing would retard the decline in plasticity that normally ensues.

CHIYE AOKI

To the Editors:

Your analysis of the deficiency in knowledge of science in "Unscientific Americans" ["Science and the Citizen," December] overlooked a major point. The deficiency is not simply a matter of lack of facts, and therefore of science education. The problem is that a significant number of people do not believe in science or in the scientific method. Pseudoscience and antiscientific religion hold significant sway.

Either the universe works in a predictable, analyzable way or it works spasmodically, with miracles, action at a distance and wishful thinking as the three fundamental forces. People tend to take one view or the other. Until Americans embrace the former world view, U.S. science education will continue to be lamentable.

FRANK D. KIRSCHNER

Alexandria, Va.

To the Editors:

"The Geometric Phase," by Michael Berry [SCIENTIFIC AMERICAN, December], is a fine example of the service you do in making concepts generated within specialized areas of inquiry available to the literate public. As a psychiatric therapist who has done some work also in theories of consciousness, I am struck by the utility of the concept of parallel transport, as developed in the article, as a model for certain experiences that result in insight, attitude change or more fundamental transformations of a psychological nature. One can move through a series of states or experiences that do not appear in themselves to alter the continuity of attitude, and at the conclusion of which one has returned to what seems the original state, only to discover that one's attitude toward everything is irrevocably changed—a kind of "psychological anholonomy."

RUSSELL C. HUFF

Southbury, Conn.

To the Editors:

Professor Berry can observe anholonomy without taking pencils to the ends of the earth. He could stay in Bristol, where I once studied anatomy, and simply observe the motion of the human shoulder.

Stand with your arm beside you, the palm of your hand facing in. Keeping the arm straight, lift it sideways until it is horizontal. Move it forward until it points directly in front of you. Finally bring it down to your side. Lo, the palm is facing backward!

RICHARD CHERRY

East Birmingham Hospital
Birmingham, U.K.

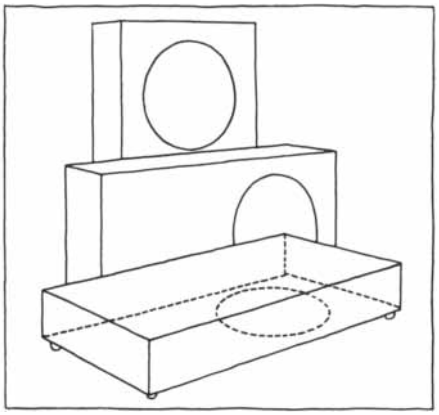
"They Were Designed To Play Music This They Do Very Well, In At A Bargain Price...It's Hard To Ima

It has always been true that placement in the listening room has a profound effect on the sound of any loudspeaker, regardless of its inherent qualities. Cambridge SoundWorks has confronted this factand created Ensemble,™ a speaker system that can provide in your home, the superb sound once reserved for the best conventional speakers under laboratory conditions. And because we market it directly, Ensemble costs far less than previous all-out designs. Perhaps best of all, it virtually disappears in your listening room.

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

The best sound comes in four small packages.

Ensemble consists of four speaker units. Two compact low-frequency speakers reproduce the deep bass, while two small satellite units reproduce the rest of the music. Separating the low bass on both channels from the rest of the range makes it possible to reproduce just the right energy in each part of the musical spectrum without turning your listening room into a stereo showroom. With clumsy conventional systems, you can either strive for that balance by letting loudspeakers dominate your room, or sacrifice it for less conspicuous speaker placement.



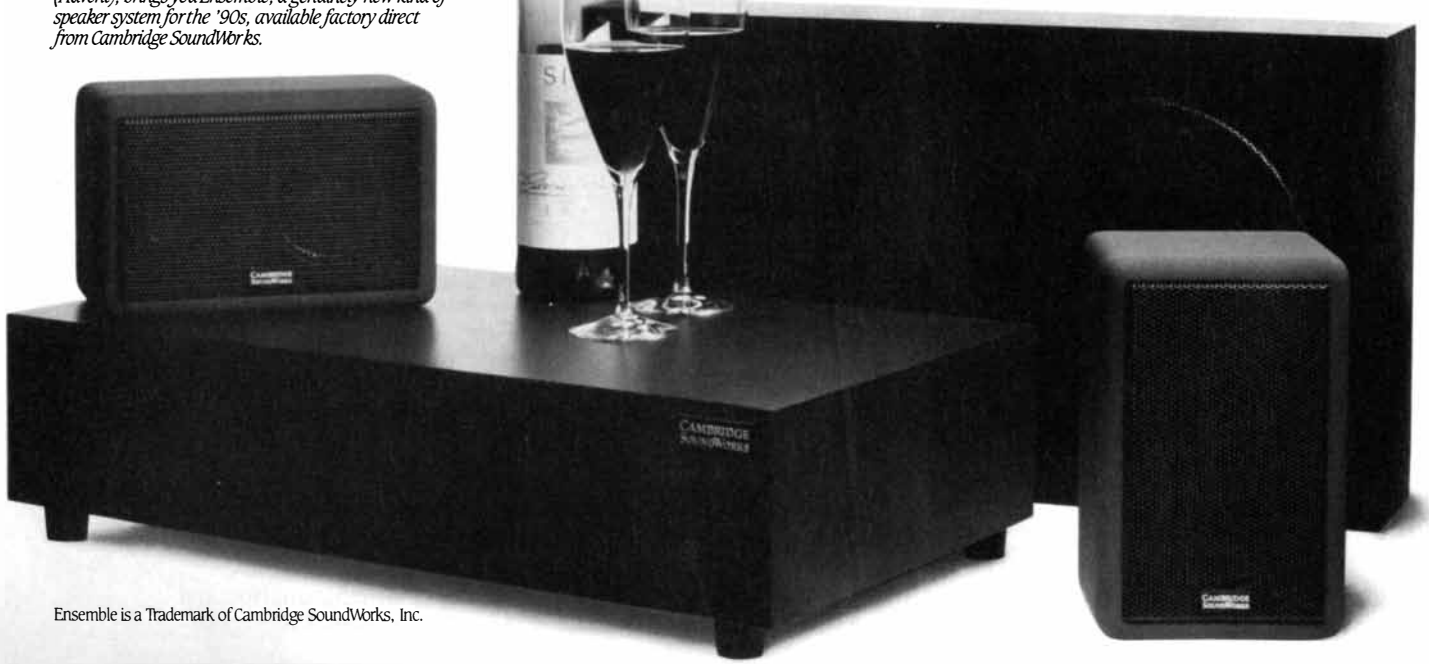
Because low frequencies are non-directional, Ensemble's bass units can be installed horizontally, vertically, facing upwards, or facing downwards.

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

Room acoustics emphasize and de-emphasize various parts of the musical range, depending upon where the speaker is placed in the room. If you put a conventional speaker where the room can help the low bass, it may hinder the upper ranges, or vice-versa.

Ensemble, on the other hand, *takes advantage* of your room's acoustics. You put the low-frequency units where they provide the best bass, whether or not that location is good for the high frequencies (and it usually

Unlike satellite systems which use a single large subwoofer, Ensemble features separate compact bass units for each stereo channel. They fit more gracefully into your living environment, and help minimize the effects of the listening room's standing waves.



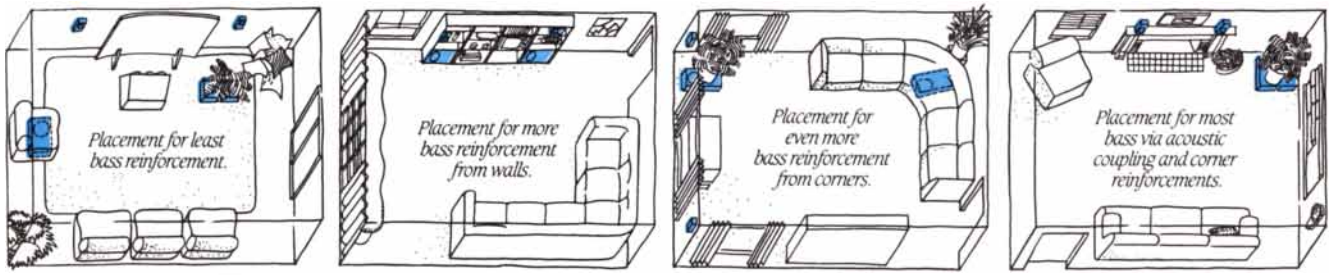
Ensemble is a Trademark of Cambridge SoundWorks, Inc.

–And Make It Sound Like Music. A Most Unobtrusive Way, gine Going Wrong With Ensemble."

Julian Hirsch
Stereo Review, Sept. '88

isn't for any speaker). Then you put the satellites where they provide a well-defined stereo "stage."

The ear can't tell where bass sounds come from, which is why Ensemble's bass units can be tucked out of the way—on the floor, atop bookshelves, or under furniture. The satellites can be hung directly on the wall, or placed unobtrusively on windowsills or shelves (among other possibilities). The result is extraordinary. There are no bulky speaker boxes to dominate your living space, yet Ensemble reproduces the satisfying deep bass that *no* mini speakers can.



You can put Ensemble's low-frequency units exactly where they should go for superb bass. You can't do this with conventional speakers because you have to be concerned about the upper frequencies coming from the same enclosures as the low ones.

Not all the differences between Ensemble and other speaker systems are as obvious as our *two* subwoofers.

Unlike three-piece satellite systems that may appear similar, Ensemble's four-piece design doesn't cut any corners. We use premium quality components for maximum power handling, individual crossovers that allow several wiring options and cabinets ruggedly constructed for proper acoustical performance. The low-frequency units use the classic acoustic suspension design, and are finished in black laminate. The satellites are finished in gunmetal gray Nextel, a suede-like finish highly resistant to scratching. We even gold-plate all connectors to prevent corrosion. But perhaps an even bigger difference between Ensemble and other speakers is how we sell it...

What Henry Kloss tells his friends:

Every time I came out with a new speaker at AR, KLH, or Advent, my friends would ask me, "Henry, is it worth the extra money for me to trade up?" And every time I would answer, "No, what you've already got is still good enough."

But today, with the introduction of Ensemble, I tell them, "Perhaps now is the time to give your old speakers to the children."

The best showroom of all: your living room.

Choosing a loudspeaker after a brief listen at a dealer's showroom is like deciding on a car after one quick trip around the block. Therefore we make it possible to audition Ensemble right in your own home. In fact, Ensemble is sold *only* by Cambridge SoundWorks directly from the factory.

That only makes sense. You get to match Ensemble specifically to your listening room in a way no other system permits. You get to listen for hours without a salesman hovering nearby. And if after 30 days of all that you're not happy, you can return Ensemble for a full refund (we'll even reimburse the original UPS shipping charges in the continental U.S.).

You also get to save.

At only $499—complete with all hardware, 100′ of speaker cable, and free ongoing assistance—Ensemble costs hundreds of dollars less than it would in a retail store.

Call 1-800-AKA-HIFI* (1-800-252-4434)

Our toll-free number will connect you to a Cambridge SoundWorks audio expert. He or she will answer all your questions from why (or why not) to buy Ensemble to those you may have about related equipment. Your audio expert will take your order (you can use Visa, MasterCard or American Express), and arrange surface shipment via UPS ($7 to $25 anywhere in the continental U.S.). You should have Ensemble within one week. And your Cambridge SoundWorks audio expert will continue as your personal contact with us, to answer questions which might come up after you've begun to enjoy Ensemble at home. We think you'll like this new way of doing business.

*In Canada, call 1-800-525-4434. Audio experts are on duty Mon.-Fri., 9AM–10PM, Sat., Sun., 9AM–6PM Eastern Time.

CAMBRIDGE SOUNDWORKS
154 California St., Newton, Massachusetts 02158

☐ Send more information and test reports.
☐ Send Ensemble risk-free for 30 days, for $499*
☐ Send an Ensemble Gift Certificate for $499*
I'm paying by ☐ Check ☐ MC ☐ Visa ☐ AmEx

Acct. Number_____ Exp._____

Signature_____

Name_____

Address_____

City_____ State_____ Zip_____

Phone (Area Code)_____ Number_____

FOR IMMEDIATE SERVICE: 1-800-AKA-HIFI
MA residents add 5% sales tax
*Plus freight. Call and ask.

Dept 105M

50 AND 100 YEARS AGO

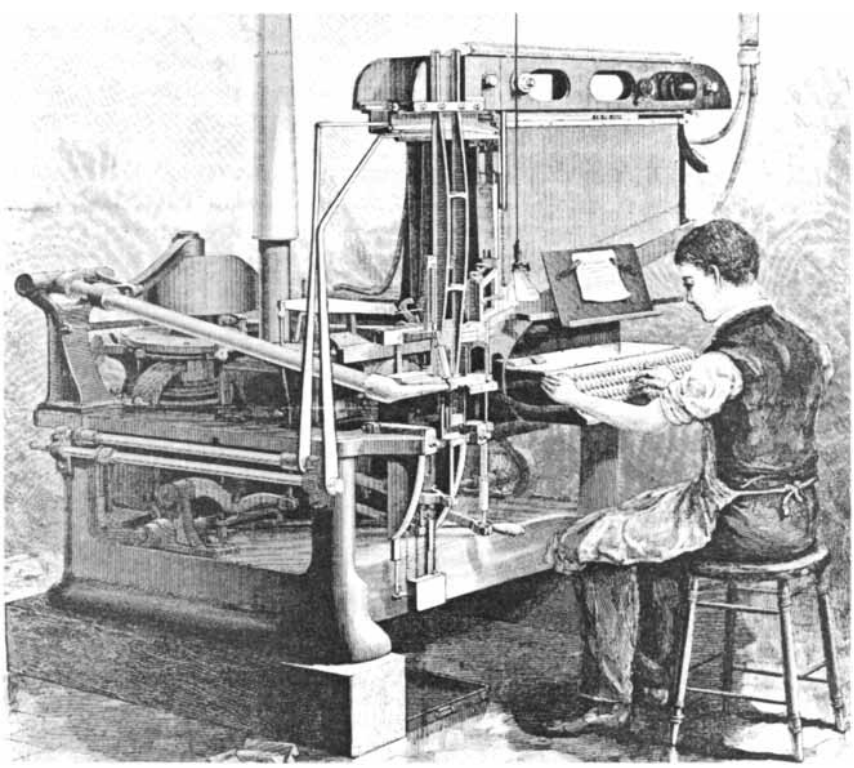
SCIENTIFIC AMERICAN

MARCH, 1939: "The problem of stream pollution is of direct concern to those who gain a livelihood in the fisheries industries, to consumers of shellfish and fin fish, and to those who at intervals fortify their health and strength with recreational fishing. To anyone who has noted the increased demands of the public in recent years for the effective cleaning of streams and coastal waters, it must be apparent that the time is rapidly approaching when an intelligent and comprehensive plan for dealing with the problem must be evolved. If industry and government co-operate in finding a solution, it will be easier, less costly and less disruptive than if a thoroughly aroused public, irked by delays and subterfuge, finally insists on immediate, drastic and precipitate action."

"The Firestone Tire & Rubber Company announces the introduction of a new motoring safeguard—the tire X ray. This new application of the X ray, a benefactor of man ever since its discovery in 1895, will enable car owners to checkmate most of the 40 million flats that are now occurring annually."

"Although the science of surveying snow is still in its infancy, its extent and importance are increasing rapidly. To grasp its significance in our western states, picture a series of snow-covered mountain ranges that by their streams make life possible in dry intervening valleys. Dearth of snowfall in the mountains means dry river beds in August, machinery standing idle, thirsty cattle and dying plants."

"Most people think of helium as an inert gas, discovered in the sun's spectrum before it was found on earth, whose sole use is for filling dirigibles. But there is a further use. Helium weighs only one seventh as much as nitrogen. Hence a mixture of 21 percent oxygen and 79 percent helium has only one third the density of air and, more important, requires only about half the effort to pass through a narrow orifice. If overtaxed lungs are allowed to breathe a helium-oxygen mixture, they need to do only about half the work they do in breathing ordinary air. The helium-oxygen mixture is of use in asthma, certain anesthesia difficulties, any obstruction of the air passages and asphyxia of the newborn."



Setting type mechanically on the Linotype machine

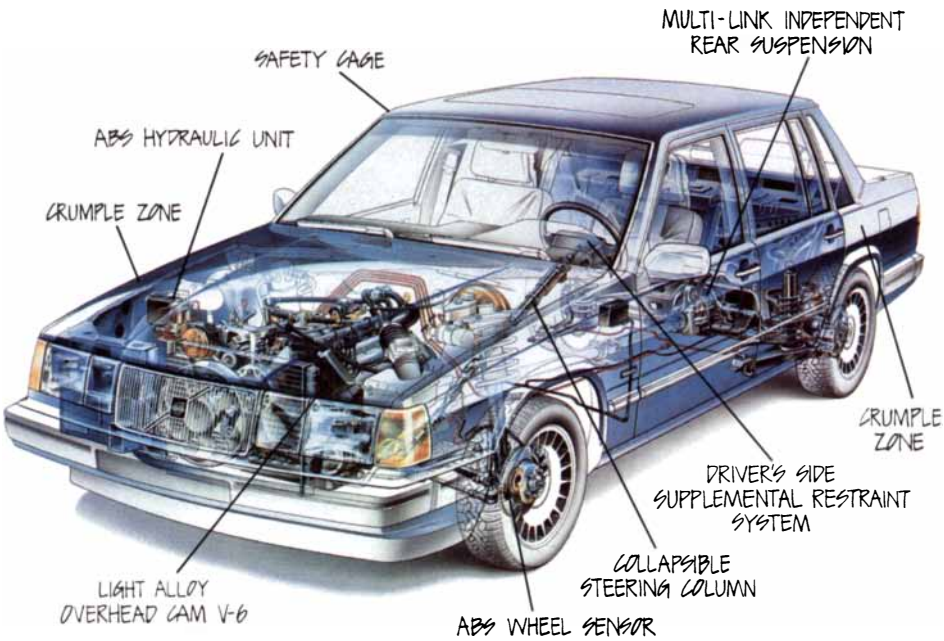
MARCH, 1889: "One of the most interesting problems for persons who devote themselves to the study of geology is this: Can we estimate, at least approximately, the age of the globe? A number of physicists and geologists have undertaken this task, but the discordance in the results obtained plainly says that the methods employed offered no certainty. The majority have found improbable figures—these being hundreds of millions and sometimes thousands of millions of years."

"It is claimed that two professors connected with the Pasteur Institute have discovered the generative microbe of diphtheria, and that a preventive of this disease by means of vaccine is expected to follow. The prosecution of the search for a diphtheria germ, which has engaged the attention of scientific experts for the past few years, does away entirely with the popular theory that the disease can be caused by sewer gas or filth."

"The direct use of electricity as a labor-saving machine has been applied at the great steel works, Cleveland, Ohio, where a large electro-magnet is used, suspended from a crane, to pick up steel bars and billets. It will pick up 800-lb. billets and drop them where wanted by the touch of a key, the movement of the crane being done by steam."

"The accompanying illustration represents the latest, and in many respects the most remarkable, of the numerous machines which inventors and mechanics have from time to time devised in their long-continued efforts to find some practical means by which to supersede or cut short the tedious work of typesetting. It is known as the Linotype machine, from the nature of its product. It is not, strictly speaking, a typesetting machine, but forms type bars, each of the length, width and height of a line of type, and the exact counterpart of that which a compositor would set up, except that each line is formed of one entire piece of metal. The actual performance of the machine at present on newspaper work is about equal to that of three ordinary compositors."

TO FIND OUT WHAT SEPARATES THE VOLVO 760 FROM OTHER LUXURY CARS, YOU HAVE TO GET INSIDE IT.



Inside the Volvo 760 you'll find more than just leather-trim seats, power windows, and a high-end stereo system.

You'll find a level of automotive engineering that exceeds that of many cars in its class.

Consider, for instance, Volvo's unique Multi-link independent rear suspension. Unlike conventional systems, Multi-link allows each wheel to react individually to varying road conditions. The result is an exceptionally smooth ride coupled with precise road handling.

Also consider the 760's choice of power plants. You can have a highly responsive, overhead cam V-6. Or an intercooled, turbocharged four that can rocket the 760 from 0-55 MPH in less time than it takes many so-called performance sedans.

But even more important than 0-55 is the time it takes to go from 55-0. Which is why the 760 comes with a state-of-the-art anti-lock braking system (ABS). With ABS, you can practically stand on the brakes with little chance of skidding or losing control. Even on wet surfaces.

For additional safety, the 760 is equipped with a driver's side Supplemental Restraint System. And, of course, it is replete with all the safety features Volvo has long been famous for.

All of which goes to prove that if you're in the market for a luxury car, you should look into the Volvo 760.

After all, no other car in its class has such an impressive interior.

VOLVO

A car you can believe in.

The guys at Pola to shoot each oth

Picture it. Their data system was good.

But now, with the instant success of their Spectra® camera, good suddenly wasn't good enough.

!@#$%&*( )! Trouble.

So Polaroid called in NYNEX—one of the information industry's favorite troubleshooters.

And NYNEX designed an integrated voice and data network that connects all eleven Polaroid locations.

And makes everything from inventory to finance to technological analysis faster. Easier. And more cost-effective.

There were simply no negatives.

The NYNEX family of companies would like to work with you, too.

roid were ready

ner.

We offer you everything from computer networks and software, to mobile phones, to the services provided by New England Telephone and New York Telephone. And more.

Call us at 1 800 535-1535.

Discovering the answer is NYNEX could be a long-term boon to your business.

We're hardly a flash in the pan.

Need to communicate? Need to compute? The answer is

NYNEX

For Bill Demby, the difference means getting another shot.

When Bill Demby was in Vietnam, he used to dream of coming home and playing a little basketball with the guys.

A dream that all but died when he lost both his legs to a Viet Cong rocket.

But then, a group of researchers discovered that a remarkable DuPont plastic could help make artificial limbs that were more resilient, more flexible, more like life itself.

Thanks to these efforts, Bill Demby is back. And some say, he hasn't lost a step.

At Du Pont, we make the things that make a difference.

Better things for better living. DU PONT

REG. U.S. PAT. & TM. OFF

SCIENCE AND THE CITIZEN

Clearing the Air

Chemical weapons can be banned, given the political will

Even in an age of nuclear arms, chemical weapons evoke revulsion. Yet such devices have been unleashed on contemporary battlefields, accusations of their use are a staple of international invective and the possibility of a terrorist chemical assault is a universal nightmare. Although the Pentagon wants to bolster the U.S. arsenal, President Bush's support of a global ban on chemical weapons is one hopeful omen. What are the prospects for a ban?

Recent accusations that arms dealers have sought to acquire nerve-gas bombs and that chemical companies knowingly helped Libya to build an alleged chemical-weapons factory do not give much cause for optimism. To be sure, Julian P. Perry Robinson of the University of Sussex, a leading expert on chemical weapons, considers the results of a conference held recently in Paris to be encouraging: representatives of 149 nations reaffirmed the 1925 Geneva Protocol banning the use of chemical weapons and recognized ongoing talks in Geneva as the legitimate forum for negotiating a total ban. But the conference produced a final communiqué that fell far short of the hopes of some Western nations (including the U.S.). Although the conference did support a ban, it did not back sanctions against violators of the existing protocol.

Still, substantial progress has been made in the past year. There is a proposed treaty on the table at Geneva that calls for the destruction of both chemical-weapons plants and weapons stockpiles. The existence of plants capable of making supertoxic agents for research use (principally nerve gases such as Sarin and VX, which are fatal in doses of less than a milligram) would have to be disclosed. A country could have no more than one such facility, and production would be limited. Plants able to make the immediate chemical precursors of supertoxics would also have to be disclosed and would be subject to inspection. The treaty's big enforcement stick is a provision stating that nondeclared plants suspected of weapons production would be subject to mandatory "challenge inspections."

Matthew Meselson of Harvard University points out that there are now no major remaining disagreements between the U.S. and the Soviet Union delegations at the conference. The Soviets have even accepted U.S. proposals for short-notice "anytime, anywhere" challenge inspections. For its part France no longer insists on maintaining "security stocks" of weapons while a treaty takes effect. The remaining political differences fall along a north-south divide; India, for example, is still reluctant to accept challenge inspections.

Meselson and Perry Robinson argue that verifiability is no longer an obstacle. The search for 100 percent reliability, they say, is a red herring. They say the crucial issue is whether, given the high degree of confidence in verification that is attainable, the danger of having a treaty is greater than the danger of not having one. "Verification can work and is practical, as long as there is reasonable access," declares Alastair W. M. Hay of the University of Leeds, who has worked on methods for detecting trace amounts of chemicals. Present techniques are able to detect nerve gases at concentrations

On-site inspections as well as satellite data will be needed to monitor a chemical-weapons treaty

***LIBYAN INDUSTRIAL FACILITY** 40 miles south of Tripoli has been alleged by U.S. officials to be the site of a new chemical-weapons plant. Libya denies the allegation. The satellite photograph was provided by the SPOT Image Corporation.*

lower than one part in a trillion. According to Meselson, that is low enough that a factory could not be "cleaned up" in order to avoid detection after making a batch of supertoxic chemicals. Gas chromatography in conjunction with mass spectrometry is the favored approach.

Indeed, an entire suite of analytic procedures for chemical weapons has been developed over the past 15 years by Marjatta Rautio and her colleagues at the University of Helsinki. They say that 200 critical precursors of the supertoxics can be reliably measured, even against a background of other chemicals. Such work has gone beyond analysis. Recently several of the countries taking part in the Geneva conference initiated national trials of different inspection protocols. The results are expected to lead to further trials of a single internationally accepted protocol. Some of the monitoring that would be needed under a treaty may well be done by tamperproof sensors that would be left in place to record the presence of key substances over extended periods.

If adequate verification is possible, what are the remaining obstacles? One is that verification could be onerous to industrial concerns. Some of the precursors that can easily be turned into supertoxics (or used in binary weapons) also have legitimate industrial uses. As a result many chemical plants would fall within the inspection regime, including most of those that make pesticides. According to Perry Robinson, "the main issue now is protecting confidentiality."

Chemical manufacturers are understandably apprehensive about the possible disclosure of commercially sensitive information, which might not be patentable. A treaty would require the disclosure of key details about plants that make possible precursors; such information could be used by an industrial competitor for market research. In addition, challenge inspections might discover details of a proprietary process. After all, as Kyle B. Olson of the U.S. Chemical Manufacturers Association points out, inspectors are likely to be experts who will probably eventually return to the chemical industry.

The details of how and when challenge inspections would be done have yet to be worked out. One principle that is already recognized, however, is that analytic procedures should be designed to test only for the presence of prohibited substances. Other ideas are also being pursued. Workers in West Germany have devised a system known as "sample now, analyze later" that stores minute samples on a moving tape. The approach makes it easy to take samples at preset times without an inspector's having to be present. The Chemical Manufacturers Association has proposed that some data on plants be stored in tamperproof enclosures on site to avoid their being spread around offices at the inspectorate's headquarters.

The chemical industries of the U.S., Britain and Japan say they support a treaty and are coordinating their positions on inspections. Although they accept the need for inspection of plants making precursors, they are reluctant to extend the list. Substances such as hydrogen cyanide and phosgene are, according to the manufacturers, simply too widespread for every production plant to be subject to inspection. Is it possible to find a compromise? "If we didn't think it was doable, we wouldn't be investing all this time and energy in it," Olson says. —*Tim Beardsley*

To Our Health

Policy recommendations for a national health-care program

Although the U.S. spends more of its gross national product on health care (11 percent) than any other nation, 37 million Americans remain financially unprotected from medical expenses. As the Bush Administration takes office, three proposals have been put forward for a universal health-insurance program that would guarantee all Americans affordable, high-quality medical care.

After a 30-month study the National Leadership Commission on Health Care, whose 34 members are leaders in the field as well as in economics and public affairs, have recommended a program that provides universal access to a nationally determined basic package of health services. Employers would be given strong financial incentives to provide health-care benefits for all their employees, according to Margaret M. Rhoades, the commission's executive director. The cost of the package for those not insured would be shared by all employers, and by employees whose incomes exceed 150 percent of the poverty level.

The commission also recommended that existing private and public agencies develop national guidelines that set standards for appropriate treatment of particular health problems. Patients and insurance companies could consult the guidelines to monitor the adequacy and cost of care.

Another proposal, devised by Alain Enthoven and Richard Kronick of Stanford University, focuses on providing greater incentives in the health-care industry to discover and implement medical practices that produce the same health benefits as existing practices at less cost. The key features of the Enthoven-Kronick plan, published in the *New England Journal of Medicine,* are institutions known as sponsors that act as health-care brokers. A sponsor would continuously monitor the health industry and make informed, cost-conscious decisions on behalf of its membership to contract with health plans that deliver high-quality, economical care. This arrangement is intended to exert a downward pressure on price by creating competition among providers. Large employers would insure their full-time employees through private sponsors; public sponsors that are funded by taxes would insure all others.

The sponsors would subsidize 80 percent of the average health-care plan, and beneficiaries would pay the difference according to their means or desire for more expensive care. The health-care plans would hire physicians to act as managers. A patient would not be entirely free to pick his or her own physician; the choice would have to be made among the physicians in a particular plan.

Physicians for a National Health Program (PNHP), directed by David U. Himmelstein and Steffie Woolhandler of the Harvard Medical School, proposed a different approach in the same issue of the *New England Journal of Medicine.* The PNHP plan is modeled after the Canadian system. "Our plan is financially more radical than Dr. Enthoven's, but it is less disruptive of the doctor-patient relationship," Himmelstein said. All Americans would be insured under a single, comprehensive program that the Federal Government would ultimately control and finance. The Government would pay hospital operating costs and other capital costs. Physicians would be salaried employees of a hospital or health-maintenance organization, or they could work on a fee-for-service basis. The fees would be set through negotiation between the Government and physicians. The Government consequently could establish overall spending limits on all health care.

Enthoven and Kronick are skeptical: "An attempt to enact a system like the Canadian one, involving a virtually complete government takeover of

Saab passenger cars are intelligently priced from $16,995 to $32,095. Manufacturer's suggested retail prices not including taxes, license, freight, dealer charges or options. Prices subject to change. Sunroof wind deflector optional at extra cost. © 1989 Saab-Scania of America, Inc.

One of the 10 Best. From day one.

There are two truths about "10 Best" lists. Car magazines like making them. And car makers like appearing on them.

Saab is no different. Where we do differ is in reserving our euphoria for those times when a car magazine's passing fancy moves close to permanent endorsement.

Since we first put our Saab 9000 Turbo on the road, *Car and Driver* has put it on their list. That's four straight years as one of the world's ten best cars.

Considering how good we are at making their list, perhaps we'll find our way onto yours.

SAAB

The most intelligent cars ever built.

health care financing, would represent far too radical a change to be politically feasible in this country."

Will any of these plans survive as they are batted between the Executive branch, Congress, practitioners and the health industry? "It is hard to predict their fate," writes Arnold S. Relman, editor-in-chief of the *New England Journal of Medicine,* "but it is safe to say that they and others like them will receive increasing attention from policy makers as they grope for ways to repair or replace our present disastrously inadequate health care financing system." —*Russell Ruthen*

A Gothic Tale

How Oxford is combating the effects of the Thatcher cuts

To step off The High (one of Oxford's main thoroughfares) into almost any of the streets that branch from it is to step into the 15th century. The walls of cloistered colleges create prospect after prospect of Gothic stonework, complete with finials and gargoyles. Arched gateways open into grassy quads. For 800 years the University of Oxford has educated men and women who contributed to all facets of knowledge, including science. That tradition is threatened at Oxford and other British universities by the Thatcher government's cutbacks in support for higher education. At risk is a universal cultural resource: Roger Bacon, John Locke, Edmund Halley and Lewis Carroll, to name just a few, all studied at Oxford. To find out how Oxford is coping, Tim Beardsley, a New College man and Washington correspondent for SCIENTIFIC AMERICAN, *returned to his university.*

For its first 800 years the University of Oxford relied on its reputation to attract benefactions and, in this century, government support. Now the government of Margaret Thatcher has forced the dons to look beyond their "crocketed pinnacles and indented battlements." Borrowing from U.S. university tradition, they have organized the Campaign for Oxford. The campaign, the largest ever launched by a British university, aims to put £220 million in the coffers over the next five years in order to protect future Nobelists and prime ministers (Oxford educated 24 of each, among them Thatcher) from the whims of the government of the day.

Like other British universities, Oxford had already been suffering from a brain drain, brought on by a number of problems. One, ironically, is its celebrated one-to-one tutorial teaching system, which means there are few professors who are unencumbered by heavy teaching loads. Worse, only 7 percent of its chairs are endowed.

The most pressing difficulty, however, is shared by all British universities. Prime Minister Thatcher has earned the enmity of many academics by countenancing substantial reductions in the value of general university support in the past decade. (In 1985 Oxford made its feelings known by voting down a proposal to grant her an honorary degree.)

Although Oxford has been unusually successful in attracting government and industry funds for scientific research, the cuts in general support (used to pay salaries and maintain the university's buildings) have nonetheless forced an 11 percent "retrenchment." Consequently the main chemistry building is said to be "almost unsafe." The physics department has no adequate lecture hall and has lost seven academic posts in recent years. The Campaign for Oxford, which was conceived by the outgoing vice-chancellor, Sir Patrick Neill, aims to provide flexibility by endowing both existing chairs and new chairs (including three for physics). Among the proposed new projects are an institute for theoretical physics and, fashionably, an environmental-change unit.

Oxford, which has 13,500 students, is also the first British university to develop a fund-raising strategy geared to corporate donors as well as to individuals, according to Henry M. Drucker, director of the campaign. Drucker, noting the early response has exceeded expectations, believes Oxford will open up the field. Indeed, Oxford reached its interim target of £51 million within months of the campaign's launching last summer. The largest donation so far is for £20 million, from E. R. Squibb and Sons, Inc., of New Jersey, which will establish a new pharmacology research building.

Not all the news reaching Oxford from Whitehall is bad. Government funds allocated for science will increase by 13 percent in 1989–90, after a long period of level funding, and it was recently announced that Britain (having negotiated a lower subscription) will after all remain a member of the European laboratory for particle physics (CERN). Government officials say the emphasis placed on applied research in recent years will shift toward more basic work.

Yet there is not likely to be a com-

mensurate increase in general university support, and a long-term shift toward funding research through direct grants to individual investigators by the result-oriented research councils is expected. Kenneth Baker, U.K. Secretary of State for Education, recently advocated a transition to a more "market led" higher education sector along U.S. lines.

In the meantime the cutbacks have already begun to take their toll at universities throughout the U.K. Detailed subject reviews have led to the closing of eight earth-science departments, and from 15 to 20 chemistry and physics departments are threatened. Oxford may be blazing a trail that others will be forced to follow.

PHYSICAL SCIENCES

Cosmic News

1987A, the Milky Way's core, dark-matter lenses and more

In January, John Horgan attended a meeting of the American Astronomical Society in Boston. His report on four topics discussed at the meeting follows.

Supernova 1987A. Two years after it first appeared in the Large Magellanic Cloud, the supernova still commands the attention of astronomers worldwide. "This supernova is very different from what I learned about in classical textbooks," said Rashid A. Sunyaev of the Soviet Institute for Cosmic Research as audience members nodded in assent.

The expanding "envelope" of the explosion, Sunyaev and others note, appears to be ragged and lopsided in form and motley in composition. The early and ongoing observations of heavy elements, such as iron and cobalt, forged in the core of the progenitor star suggest that in the explosion they have become mixed with light elements, such as hydrogen and helium, from the progenitor's outer shell. The blurred Doppler shifts, simultaneously red and blue, seen in some readings provide evidence of fragmentation; apparently matter on the far side of the envelope, hurtling away from the earth, as well as on the near side has come into view.

Employing a technique called speckle interferometry, workers at the Harvard-Smithsonian Center for Astrophysics have made images that show the envelope bulging on one side, in roughly the same region where the

Save 20%
on clothbound books

Special Offer

NEW MOLECULAR DESIGN OF LIFE

Lubert Stryer

Here is an information-packed introduction to key recent advances in biochemistry. This new paperback brings together in a convenient format the first eight chapters of Lubert Stryer's *Biochemistry,* the world's bestselling biochemistry text. These up-to-date chapters reflect the growing interrelatedness of molecular biology and biochemistry and show how techniques of protein and nucleic-acid chemistry reinforce each other. Highly informative, beautifully rendered color illustrations support the concise, lucid text.
240 pages, illus., $14.95 paper (ISBN 2050)

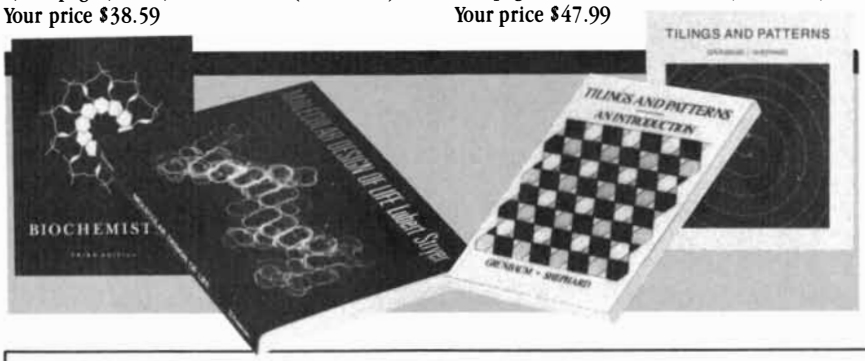
Also available:
BIOCHEMISTRY, Third Edition, Lubert Stryer
1,089 pages, illus., $47.95 cloth (ISBN 1843)
Your price $38.59

NEW TILINGS AND PATTERNS: AN INTRODUCTION

Branko Grünbaum
and G. C. Shephard

This paperback presents the most essential and broadly useful material from *Tilings and Patterns,* the authors' "encyclopedic, definitive" reference *(Mathematical Intelligencer).* Omiting the more specialized chapters of that volume, this abbreviated edition makes the authors' pioneering contributions to tiling theory and its applications fully accessible to a wide audience for the first time. The hundreds of illustrations form a valuable visual reference in themselves.
464 pages, illus., $28.95 paper (ISBN 1998)
Available April 15, 1989.

Also available:
TILINGS AND PATTERNS, Grünbaum & Shephard
700 pages, illus., $59.95 cloth (ISBN 1193)
Your price $47.99



THE MATHEMATICAL TOURIST

Snapshots of Modern Mathematics
Ivars Peterson

Peterson guides us on an unforgettable tour through the fascinating world of modern mathematics, where exciting discoveries are being made at every turn. "The only popular book on mathematics which covers many of the really new developments. . . ."—Ronald L. Graham, Director, Mathematical Sciences Research Center, AT&T Bell Laboratories
240 pages, illus., $17.95 cloth (ISBN 1953)
Your price $14.59

ENERGY AND THE EVOLUTION OF LIFE

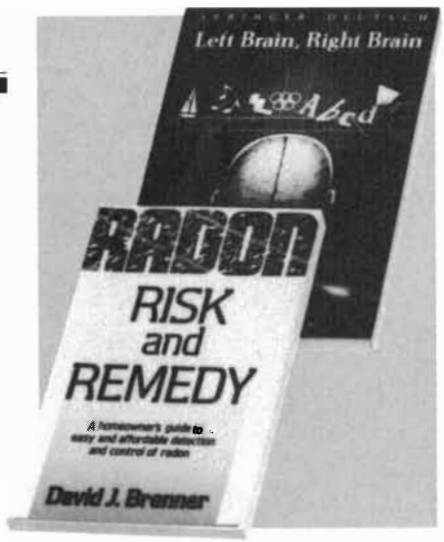
Ronald F. Fox

In this compelling examination of biological evolution, its physical bases, and its mechanisms of expression, Fox proposes provocative answers to two questions: How did life arise from inanimate matter? How and why did living organisms become progressively more complex as they evolved? "An exciting scientific book. . . .Will inevitably raise debate. . . ."—Ronald B. Spector, Rockefeller University
182 pages, illus., $26.95 cloth (ISBN 1849)
Your price $21.59

NEW RADON: RISK AND REMEDY

David J. Brenner

Radon is a radioactive gas present to some degree in virtually all households. This valuable new book—required reading for all homeowners—offers the scientific facts and the practical advice you need to protect yourself and your family from this potentially deadly toxin. As the author shows, radon is a dangerous problem, but one that is easy and affordable to detect and control.
224 pages, illus., $10.95 paper (ISBN 2030)
$16.95 cloth (ISBN 2029)—Your price $13.59



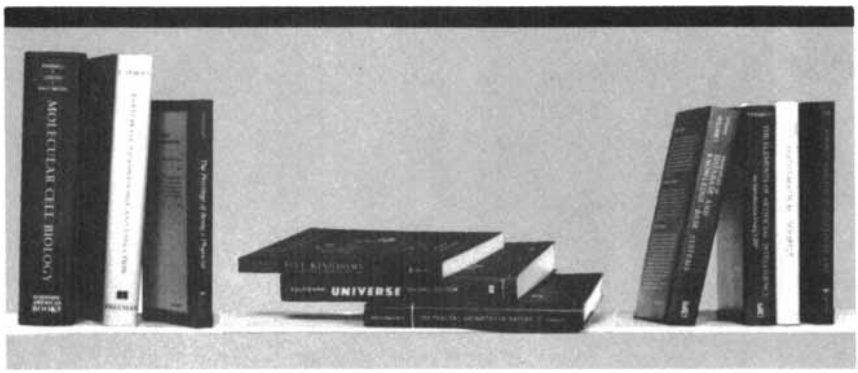
NEW LEFT BRAIN, RIGHT BRAIN

Third Edition
Sally O. Springer
and Georg Deutsch

The new edition of this award-winning bestseller brings together important recent developments in brain science that shed light on how mental and physical functions are divided between left and right hemispheres of the brain. The authors present critical findings on asymmetry in brain-damaged, split-brain, and normal subjects and examine their implications for human behavior and cognition.
384 pages, illus., $14.95 paper (ISBN 2000)
$24.95 cloth (ISBN 1999)—Your price $19.99

to Scientific American readers

from W. H. Freeman and Company, the book publishing arm of Scientific American



Our most popular references for your professional library

THE FRACTAL GEOMETRY OF NATURE
Benoit B. Mandelbrot
460 pages, $34.95 cloth (ISBN 1186)
Your price $27.99

MOLECULAR CELL BIOLOGY
James Darnell, Harvey Lodish, and David Baltimore
1,187 pages, $46.95 cloth (ISBN 1448)
Your price $37.59

VERTEBRATE PALEONTOLOGY AND EVOLUTION
Robert L. Carroll
698 pages, $52.95 cloth (ISBN 1822)
Your price $42.59

FIVE KINGDOMS
An Illustrated Guide to the Phyla of Life on Earth
Second Edition
Lynn Margulis and Karlene V. Schwartz
376 pages, $24.95 paper (ISBN 1912)

UNIVERSE
Second Edition
William J. Kaufmann III
634 pages, $36.95 (ISBN 1927)
Your price $29.59

NEW
INTRODUCTION TO GENETIC ANALYSIS
Fourth Edition
David T. Suzuki, Tony Griffiths, Jeffrey Miller, and Richard C. Lewontin
784 pages, $39.95 cloth (ISBN 1956)
Your price $31.99

From Computer Science Press

PRINCIPLES OF DATABASE AND KNOWLEDGE-BASE SYSTEMS

Volumes I and II
Jeffrey D. Ullman

This two-volume set integrates classical database concepts and the new trends in knowledge systems and object-oriented systems. Volume I (1988) covers data models and their commercial languages, as well as design theory, physical storage, and security. Volume II (1989) covers old and new optimization techniques for efficient implementation of query languages, including the more expressive logic languages of knowledge-base systems.
Volume I: 631 pages, $41.95 cloth (ISBN 8158) Your price $33.59
Volume II: 576 pages, $41.95 cloth (ISBN 8162) Your price $33.59 NEW

THE ELEMENTS OF ARTIFICIAL INTELLIGENCE

An Introduction Using LISP
Stephen L. Tanimoto

"Clearly the best AI book available. . . . LISP and LISP programs are beautifully incorporated. . . . so that LISP and AI can be learned together."—Leonard Uhr, University of Wisconsin, Madison

"Should be on the bookshelf of every programmer interested in this dynamic and fascinating field."—*Optical Engineering*
537 pages, $39.95 cloth (ISBN 8028)
Your price $31.99

THE PRIVILEGE OF BEING A PHYSICIST

Victor F. Weisskopf

Weisskopf—recent recipient of the Enrico Fermi Award—brings his intellectual integrity to bear on a diverse range of topics, from the origin of the universe to the making of the atomic bomb to his own life as a Hitler refugee and as a leader in theoretical physics. These essays reveal a pioneering scientist closely attuned to the moral issues of the nuclear age—an age he helped to create.
256 pages, illus., $17.95 cloth (ISBN 1982)
Your price $14.59

Your satisfaction matters. Return any book in good condition within 15 days of receipt and we'll send you a prompt refund.
Offer expires May 31, 1989.

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 _____

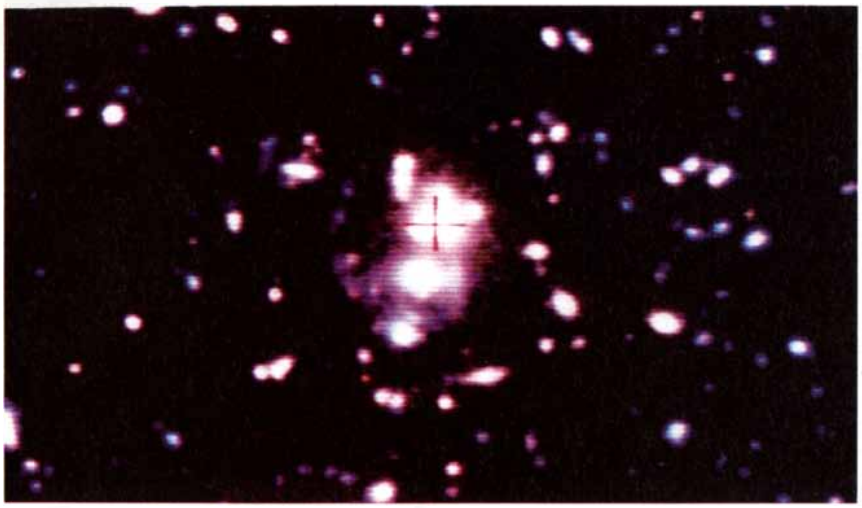
City/State/Zip _____

Qty.	Author	ISBN	Special Price	Total

Postage & handling*(Add $1.95 for first book, $1.25 for each additional book) $_____
NY, NJ, MD, CA, and UT residents add sales tax $_____
TOTAL $_____

*For Canadian residents, postage & handling is $2.25 for first book, $1.50 each additional book. Please allow four weeks for delivery. Items ordered together may be shipped separately.

0335



CLUSTER OF GALAXIES distorts the images of more distant galaxies near the same line of sight in a charge-coupled-device image made by J. Anthony Tyson of the AT&T Bell Laboratories **(left).** *Supernova 1987A appears egg-shaped in a speckle-interferometry image made by workers at the Harvard-Smithsonian Center for Astrophysics* **(right).**

group observed a "mystery companion" to the supernova soon after it appeared. The observations have fed speculation that Sanduleak −69 202, the giant blue star thought to be the progenitor of 1987A, may have had a companion.

The compact core that most supernovas are thought to leave behind remains hidden. But the *International Ultraviolet Explorer* satellite has recently observed a slight "tailing off" of the long and steady decline of light from the supernova, according to Robert P. Kirshner of the Harvard-Smithsonian Center. The readings could be the first sign of a remnant: an ordinary neutron star, a rapidly rotating one (a pulsar) or a black hole, Kirshner says.

The galactic center. Does a black hole lie at the center of the Milky Way? Observations made back in 1977 by a balloon-borne gamma-ray detector suggested that it might. The instrument, designed by the Sandia National Laboratories and the AT&T Bell Laboratories, detected a powerful spike of gamma rays coming from the galactic center. The radiation's energy level, 511,000 electron volts (511 keV), indicated that it had been generated by the mutual annihilation of electrons and their antimatter counterparts, positrons. Theorists said only the accelerative forces of a black hole could produce antimatter in such copious amounts.

Two years later, however, the radiation abruptly ceased, and for almost a decade it remained undetected. Indeed, only last April an instrument made at the California Institute of Technology did not detect the 511-keV spike, although it did spot an abundance of lower-energy gamma rays and X rays coming not from the galactic center but from a position some .7 degree, or at least 340 light-years, distant. Then one month later the Sandia and AT&T team, joined by investigators from the National Aeronautics and Space Administration, found that the 511-keV radiation had "turned on" again. Observations last November confirmed the finding.

Because their detector does not have the resolution of the Caltech device, the Sandia-AT&T-NASA workers cannot tell whether the 511-keV radiation is coming from the exact center of the galaxy or from a nearby source, such as the one observed by the Caltech group. The workers hope to pinpoint the location with observations later this year. In any case, they think the radiation comes from a black hole at least 1,000 times as massive as the sun that only intermittently drags matter from a surrounding cloud of gas. The ultimate source of this gas may have been identified by Paul T. P. Ho of Harvard University. Using the Very Large Array radio telescope, Ho detected a vast river of gas some 15 light-years long funneling toward the galaxy's core.

Dark-matter lenses. Astronomers have long suspected that the matter they can monitor directly represents only a fraction of the total stuff of the universe. How can the "dark matter" be detected? One way is to examine distant galaxies for signs of distortion, or "lensing," caused by the gravitational field of a nearer mass. J. Anthony Tyson of the AT&T Bell Laboratories is a pioneer in this approach.

With charge-coupled devices Tyson has been able to detect swarms of galaxies some 10 billion light-years distant, at the edge of the visible universe. He has also helped to develop pattern-recognition software that can automatically search images of those galaxies for signs of lensing by nearer objects. (A powerful lens typically will make more distant galaxies appear to bend into arcs around it.) By combining these techniques, Tyson says, he has found several regions in which distant galaxies have been distorted by closer galaxies or clusters of galaxies. The degree of the distortion suggests that the lenses consist largely of dark matter. Tyson has recently organized an international project to search for more lenses. Eventually, he says, the project may provide an estimate of just how much dark matter the universe contains.

The Boötes "void." Another cosmic question concerns the distribution of matter in the universe. Most models suggest it should be spread fairly smoothly through space. But in recent years three-dimensional maps of galaxies (whose distances are derived from their red shifts) show them concentrated in "clumps," "sheets" or "bubbles" separated by immense voids. In the late 1970's workers found what seemed to be the greatest void of all—a roughly spherical region some 400 million light-years across—beyond the constellation Boötes.

It appears that the so-called Boötes

void may not be as empty as had been thought. While examining readings of the region made by the *Infrared Astronomical Satellite* (*IRAS*), Gregory D. Bothun and Gregory S. Aldering of the University of Michigan at Ann Arbor found hundreds of optically dim galaxies—which tend to be dustier and less evolved than ordinary bright galaxies—that had never been catalogued. The apparent luminosities of the galaxies, together with red-shift data gathered subsequently by optical telescopes, indicated that many of them were within the void.

Might other "voids" observed in surveys of optically bright galaxies actually be filled with dim, dusty galaxies? Not necessarily. Other investigators studying *IRAS* data recently reported that, in an area this side of the Boötes void, dim galaxies tend to cluster in the same regions as optically bright galaxies. It seems only a much more thorough survey of both bright and dim galaxies will determine just how lumpy or smooth the universe is.

The Summer of '88

A closer look at last year's drought

During the summer of 1988 more than 70 percent of the continental U.S. suffered from scorching heat and drought. Fields turned brown, lakes and ponds disappeared and crop yields fell by 30 to 40 percent. Was this the beginning of irreversible global warming brought on by the greenhouse effect, as many Americans feared, or was it merely a one-summer phenomenon, a natural perturbation in the earth's climate?

Although the planet may well be in the early stages of global warming, most climatologists think the onset of last summer's hot weather was much too abrupt to reflect the slower, more insidious change in climate resulting from the buildup of carbon dioxide and other greenhouse gases. Instead the 1988 drought is widely viewed as the result of naturally occurring variation, much like that responsible for the dust bowl of the 1930's and the devastating drought of the 1950's.

In a recent issue of *Science* Kevin E. Trenberth and Grant W. Branstator of the National Center for Atmospheric Research and Phillip A. Arkin of the National Oceanic and Atmospheric Administration offer a compelling explanation for the drought's origin. According to Trenberth and his colleagues, the drought can be linked to cyclical events in the tropical Pacific that periodically result in an El Niño event: the appearance of an anomalous stretch of warm surface water in the eastern Pacific near the Equator.

Alternating with the El Niño and of about equal duration is a cold episode, sometimes referred to as La Niña, during which surface waters in the eastern Pacific become abnormally cold. The cold episode is triggered by the onset of unusually strong trade winds that push the warm El Niño waters westward across the Pacific; as the warm water moves west, cold water wells up to replace it, in turn affecting patterns of atmospheric circulation. Eventually the winds die down and warm water from the western Pacific flows back toward South America. When that happens, the El Niño re-forms and the entire cycle, which lasts for about four years, starts over.

Trenberth and his colleagues believe, based on sophisticated climate modeling, that the recent mass of unusually cold water, which began to form in the spring of 1988 (following the El Niño of 1986–87), combined with warmer waters southeast of Hawaii and displaced the intertropical convergence zone (so called because it is where trade winds from the Northern and Southern hemispheres meet) to the north. There thunderstorm activity in the zone interfered with the North American jet stream, an eastward-flowing air current that normally brings spring and summer rains to North America. Consequently the jet stream was pushed into Canada, leaving behind it a large, stalled high-pressure system marked by clear, hot weather. The drought then became self-perpetuating: plants stopped releasing moisture into the air, surface temperatures rose and residual water in the soil evaporated.

The theory, while suggestive, is still open to criticism. Mark A. Cane of the Lamont-Doherty Geological Observatory agrees that the connection between drought and the cold episode is likely but says history has shown that one can occur without the other. "Although Trenberth's work strikes me as sound, it is not clear at this point what the precise connection is between the cold phase of the El Niño and last summer's drought," he says.

Although not all cold episodes are associated with drought, Trenberth points out that the recent cold episode was unusually cold—from two to three degrees Celsius below the temperature of the surrounding water—and unusually large, extending almost a fourth of the way around the globe. Moreover, because the El Niño event is known to trigger a chain of interconnected oceanic and atmospheric events, it is reasonable to suppose La Niña also mediates climatic change.

What, if anything, do these findings say about the summer of 1989? An accurate prediction is impossible, but Trenberth expects to see a return to more normal weather conditions. "Although we won't know for certain until late March or April, when the trade winds are weakest and the trend toward warm or cold sea-surface temperatures is set, signs are already encouraging that next summer will be more normal," he says.

Tim P. Barnett of the Scripps Institution of Oceanography agrees that last summer's drought may in fact be attributed to the cold episode of the El Niño. But he emphasizes that "even if the summer of 1989 is relatively cold and wet, we will be making a serious mistake to assume that our climate problems are over. It is essential that the long-term consequences of the increasing buildup of greenhouse gases not be forgotten once the weather returns to normal." —*Laurie Burnham*

A Snowball's Chance

Support is evaporating for a theory on tiny comets

Science is a heartless process. For every theory that survives, many more perish in infancy. That seems to be the fate of the "cometesimals" posited by Louis A. Frank of the University of Iowa.

Several years ago, while examining images from the *Dynamics Explorer I* satellite, Frank noticed that every few seconds a dark spot appeared briefly against the bright background of the earth. He proposed that balls made primarily of ice and averaging some 10 meters in diameter were hurtling into the atmosphere and vaporizing at a rate of millions per year. The proposal flew in the face of conventional cosmogony. It implied, for example, that the oceans were filled by the cometlike objects and not by water in the earth's primordial materials.

Investigators suggested that Frank's spots represented noise, or random fluctuations, in the satellite data. Disbelief mounted when other satellites and telescopes failed to find any sign of the cometesimals. Then, in late 1987, Clayne M. Yeates of the Jet Propulsion Laboratory seemed to have found corroborating evidence. While scanning the skies with an electronic

COMFORT JUST GOT

INTRODUCING A COMPLETELY DIFFERENT MERCURY COUGAR XR7.

Mercury Cougar XR7 is a completely redesigned automobile that combines comfort and control with uncommon quickness. Its adrenaline takes the form of a 210-hp intercooled supercharged 3.8L V-6 engine. XR7's longer wheelbase contributes to a more comfortable ride, while its wider stance enhances stability and control. It also gives you the comfort and control of standard four-wheel disc anti-lock brakes and an adjustable automatic/firm ride control suspension. Plus an interior that was designed by people who sit behind the wheel, not just behind a desk. Experience the comfort and control of a Mercury Cougar XR7 at your Lincoln-Mercury dealer today. For more Mercury Cougar information, call 1-800-822-9292.

LINCOLN-MERCURY DIVISION Ford Buckle up—together we can save lives.

A SHOT OF ADRENALINE.

MERCURY

WHERE COMFORT AND CONTROL ARE ONE.

detector at the Kitt Peak National Observatory in Arizona he recorded faint streaks in numbers matching those predicted by Frank (see "Science and the Citizen," June, 1988).

Yeates submitted his findings early last year to *Geophysical Research Letters,* the same journal that published Frank's original paper on cometesimals in 1986. Alexander J. Dessler of Rice University, the editor of the journal, asked four experts in remote sensing to review the article. Three reviewers concluded that the streaks observed by Yeates were produced by noise; one surmised that the streaks showed the passage of orbiting debris. Dessler rejected the paper.

Yeates insists that his data are good and says he is considering protesting Dessler's decision. Frank plans to submit a paper of his own to another journal; he is also considering setting up an observatory in Iowa to look for the cometesimals. Dessler suggests that these efforts may allow the cometesimals, like the Loch Ness monster, to persist in the public's imagination. "But the overwhelming view among astronomers," Dessler says, "is that they don't exist." —*J.H.*

Beyond Understanding?

Computers are changing the spirit of mathematics

Once upon a time a mathematical proof had to be short, simple and intelligible to the human mind. That ideal has been fading fast. In 1976, for example, Wolfgang Haken and Kenneth Appel reported that a computer program had searched 1,482 graphs that represented all possible map configurations. It had found that four colors sufficed to mark any map so that no adjacent regions shared the same color. Although the computation took more than 1,000 hours, the results could still in theory be checked by hand.

Then, late in November of last year, publication of a solution to just part of a different problem demolished the notion of hand checking forever. After about 3,000 hours of computation on a CRAY-1 supercomputer (shoehorned into the machine's idle moments over two years), Clement W. H. Lam of Concordia University in Montreal announced that his team of mathematicians and computer scientists had proved one example of a conjecture first posed by Carl Friedrich Gauss nearly 200 years ago.

Practically speaking there is no way to check their result, which is that there are no finite projective planes of order 10. (Such a finite projective plane, if it existed, could be thought of as a matrix of 111 rows and columns, in which each row has exactly 11 positions filled and any two rows have only a single filled position in common.) To prove the negative result, the CRAY examined roughly 100 trillion different possibilities in its final two-year run. Worse yet, notes John McKay, one of Lam's colleagues, the result did not come from a single program but from a collection of "little bits" of software, each designed to attack a single facet of the problem. "I was hoping for a positive result, because then you could check it," Lam says. Instead he found a sea of negative cases, any one of which could be in error.

Yet Lam has confidence in his answer, and the mathematics community appears to have confidence in Lam. Part of the reason for this confidence, Lam says, is that the programs used contain consistency checks, so that simple errors in the calculations would be readily detectable. Mathematical intuition also favors Lam. It had already been proved, for example, that an order-10 projective plane, if it existed, would have no symmetry, something almost unheard of for such a complex object.

The response to Lam's research yields a significant sociological result: after millenniums of preferring only the simplest mechanical aids, mathematicians have come to accept the computer as a partner in coping with complex problems. No mathematician today would flatly refuse to accept the result of an enormous computation, as some did when Haken and Appel announced their four-color proof, says Ronald L. Graham of the AT&T Bell Laboratories. Haken is a bit blunter: he notes that many of the older mathematicians who raised philosophical objections to the four-color computations have retired.

Indeed, computers have become the only way to make progress in certain areas of mathematics. Studies of chaos and fractals, for example, rely almost entirely on computing. Rather than proposing theorems and formulating proofs, mathematicians in these disciplines frame hypotheses and perform experiments.

For those who prized the certainty and abstraction of traditional mathematics, such new developments may be disheartening, but they appear to be inevitable. Although some mathematicians hold a "religious belief" that a simple theorem ought to have a short, simple proof, Haken says, some simple theorems have enormous proofs. Graham agrees. He notes that in some restricted subsets of mathematics, proving a two-symbol statement can require 16 symbols. A three-symbol statement can require 256, and proof of an eight-symbol statement can demand more symbols than there are fundamental particles in the universe.

Furthermore, Graham says, mathematicians working alone may have only scratched the surface of the interesting theorems to be found in the world. Computer-based exploration can point out promising avenues for formal analysis. "Knowing what to prove," Graham observes, "is more than half the battle." —*Paul Wallich*

BIOLOGICAL SCIENCES

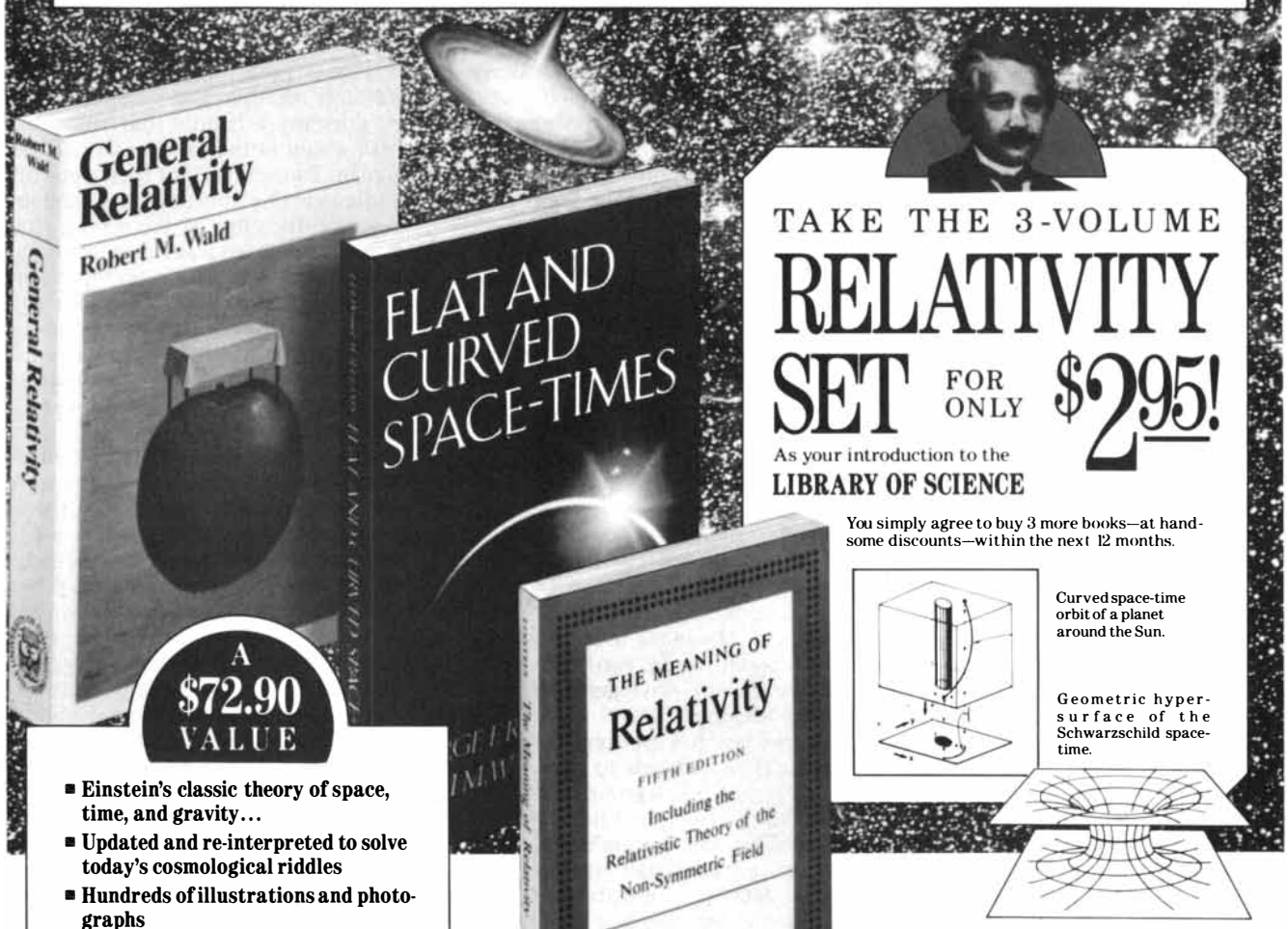
Time Bomb

War breaks out in the field of evolutionary biology

With the publication of two articles, a conflict in the field of evolutionary biology has finally escalated into an officially declared war. The dispute involves a crucial issue: What is the best way to determine the phylogeny, or evolutionary heritage, of a given species? One way is to compare its genes with those of other species. The assumption is that after any two species branch off from a common root, their genes diverge at a roughly constant rate as a result of random mutations.

Two of the most prominent practitioners of this so-called DNA-clock approach are Charles G. Sibley and Jon E. Ahlquist, former collaborators at Yale University who have recently moved to San Francisco State University and Ohio University respectively. The bulk of their research has focused on birds (see "Reconstructing Bird Phylogeny by Comparing DNA's," by Charles G. Sibley and Jon E. Ahlquist; SCIENTIFIC AMERICAN, February, 1986). After doing thousands of experiments in the 1970's and early 1980's, they felled many family trees devised by other ornithologists (most of whom had been relying on anatomical rather than molecular comparisons) and replaced them with extraordinarily intricate new ones. More recently, on the basis of a much smaller number of experiments, Sibley and Ahlquist revised the uppermost branches of the hominoid tree. They concluded that

EINSTEIN'S THEORY IN TODAY'S PHYSICS AND ASTRONOMY



TAKE THE 3-VOLUME RELATIVITY SET FOR ONLY $2.95!

As your introduction to the **LIBRARY OF SCIENCE**

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

Curved space-time orbit of a planet around the Sun.

Geometric hypersurface of the Schwarzschild space-time.

- **Einstein's classic theory of space, time, and gravity...**
- **Updated and re-interpreted to solve today's cosmological riddles**
- **Hundreds of illustrations and photographs**

Over 80 years after Einstein first proposed his theory of Special Relativity, gravitational research is as vital a field as ever, with implications reaching from inflationary cosmology and superstring unification to the ongoing search for gravity waves. The Relativity Set brings you two comprehensive modern treatments of Special and General Relativity, along with Einstein's original ground-breaking work where he first presented relativity. Best of all, you can save almost $70.00 on this 3-volume set. Just complete and return the postage-paid reply card and we'll send you the $72.90 Relativity Set for only $2.95. For faster service, call 1-800-888-4843.

THE MEANING OF RELATIVITY
Fifth Edition
Albert Einstein
A classic collection of Einstein's revolutionary papers on space, time, and relativity. Selected writings focus on both Special and General Relativity, as well as the Relativistic Theory of the Non-Symmetric Field—Einstein's last published scientific work.

Publisher's Price: $7.95

FLAT AND CURVED SPACE-TIMES
George F. R. Ellis and Ruth M. Williams
An intriguing approach to the geometrical space-time of Einstein's theories which clearly shows how many seemingly disparate features of our physical universe are in fact one. Contraction, time dilation, and the twin paradox are included in the flat space-times development of the special theory of relativity. Topics from expanding universes to black holes are then introduced in the curved space-times of general relativity to show how scientists might reach new breakthroughs.

Publisher's Price: $35.00

GENERAL RELATIVITY
Robert M. Wald
"Required reading for anyone who wishes to have an understanding of the modern approaches to general relativity."—*S. Chandrasekhar*
A comprehensive, up-to-date adaptation of Einstein's theory to today's quest for a quantum theory of gravity and cosmic evolution. Einstein's theory provides a classical description of space-time structure and gravitation. The theory of relativity is described as it relates to quasars, singularities, and the quantum particle creation in the gravitational fields of black holes.

Publisher's Price: $29.95

Call **1-800-888-4843** to join now!

THE LIBRARY OF SCIENCE is the nation's oldest and most respected science book club. Dedicated to serving both professionals and serious amateurs, it surveys the entire field of published sources on scientific subjects to make the very finest works available to club members at substantial savings.

MEMBERSHIP BENEFITS • In addition to getting the Relativity Set for only $2.95 when you join, you keep saving substantially on the books you buy. • Also, you will immediately become eligible to participate in our Bonus Book Plan, with savings of 65% off the publishers' prices • At 3–4 week intervals (16 times per year), you will receive the Library of Science News, describing the coming Main Selection and Alternate Selections, together with a dated reply card. • 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 the Library of Science, Dept. 2-CL5-00956, Riverside, NJ 08075, for membership information and an application.

Scientific American 3/89

human beings and chimpanzees are each other's closest relatives and that gorillas branched off from the hominoid trunk earlier. Previous molecular studies had indicated that the three species diverged so nearly simultaneously that they form an unresolvable trichotomy.

Sibley and Ahlquist have had many critics, none more hard-hitting than Vincent M. Sarich of the University of California at Berkeley, Jon Marks of Yale and Carl W. Schmid of the University of California at Davis. For the past two years this trio has challenged both the methodology and the conclusions of Sibley and Ahlquist. Until recently the critics' charges have been confined to letters, meetings, remarks to science writers (stories on the controversy appeared in *The Scientist* last June and in *Science* last September) and two unpublished papers. Now two peer-reviewed journals, *Cladistics* and the *Journal of Human Evolution,* have formalized the dispute by publishing the papers. The *Cladistics* article focuses on the bird research and the other article on the hominoid work.

Both articles begin by saying that Sibley and Ahlquist exaggerate the accuracy of their DNA-clock technique. The technique, known as DNA hybridization, calls for combining single strands of DNA drawn from two different species into "hybrid" double strands. The greater the similarity between the two types of DNA is, the more tightly they bind to each other and the more heat is needed to "melt," or break the bonds between, the hybrid strands. The hybrid strands are labeled with radioactive iodine, which enables workers to measure the melting rate. Sarich, Marks and Schmid assert that the technique is fraught with uncertainties and possible sources of error that are not adequately acknowledged by Sibley and Ahlquist.

The critics do not stop there; they also accuse Sibley and Ahlquist of having misrepresented their original laboratory data. The critics base this claim on their analysis of small portions of the original data obtained through a third party. Some of the data did not match the results published by Sibley and Ahlquist, according to Sarich et al.; they also concluded that the raw data supported neither the human-chimpanzee-gorilla relation described by Sibley and Ahlquist nor some of the detailed bird phylogenies.

Sibley and Ahlquist admit they altered some of their raw data, but they insist these were legitimate "corrections" of readings that—owing to their long experience with DNA-hybridization procedures—they knew were flawed. "We should have reported the changes," Ahlquist remarks. He notes that since he and Sibley reported that humans and chimpanzees are each other's closest relatives, numerous investigators have reached the same conclusion. As a further check, Ahlquist says, he and Sibley are reanalyzing their hominoid data without making corrections; so far the reanalysis supports their original finding.

As for the bird research, Ahlquist suggests that some of the more intricate branchings devised by him and Sibley may indeed turn out to be based on flawed interpretations of their data. "We didn't always indicate whether our data were tentative or definitive," he says. He concedes that redoing the bird experiments might lead to the rearrangement of a few branches or to a realization that some of the relations are too close to call, but he contends that "most of the major features will remain the same." An independent confirmation of Sibley and Ahlquist's avian phylogenies is unlikely to be forthcoming soon, however: the phylogenies are based on more than 25,000 separate DNA-hybridization experiments that took more than a decade to complete.

In a speech to a group of rapt graduate students at the City University of New York last December, Marks said bluntly: "Sibley and Ahlquist fudged their data." He called the controversy a "tragedy" that has hurt the entire field of evolutionary biology. But Sibley called it "a tempest in a teapot" that both DNA hybridization and his and Ahlquist's work will survive. "Twenty years from now more people will be doing this, the precision will be better and the soft spots will be rooted out," he says. "Nobody knows who the pioneers are until much later." —*J.H.*

Thinking in Circles

Workers take a picture of a brain in action

What does a thought process actually look like? Experimental psychologists have long maintained that many mental operations closely mirror operations in the real world. We work out how to maneuver a chair through a doorway, for example, by carrying out the procedure in our mind beforehand. Such an "analogue" mental process may now have been imaged directly, as a changing pattern of nerve-cell activity in the brain of a monkey.

Writing in *Science,* Apostolos P. Georgopoulos of the Johns Hopkins University School of Medicine and his coworkers describe the monkey's task, which was designed to make the animal perform the mental analogue of a physical rotation. The monkey began by grasping a handle that coincided with a light at the center of a circular screen. The center light then went off and a new one appeared at a random point on the edge of the screen. The monkey had been trained to move the handle in the direction of this stimulus if the light was dim; if the cue was bright, the animal was supposed to move the handle toward a point 90 degrees counterclockwise (toward nine o'clock if the light appeared at 12 o'clock, for example).

Based on previous behavioral studies of human beings, Georgopoulos and his colleagues hypothesized that in order to respond correctly the monkey first had to imagine a "movement vector" and mentally rotate it 90 degrees counterclockwise from the direction of the stimulus. How might such a movement vector be represented in the brain of a monkey, and how can the representation be detected?

The answers derived from earlier work, in which another group led by Georgopoulos had recorded the activities of hundreds of neurons in the motor cortex of a monkey as the animal moved its arm in various directions. Changes in firing rates showed that each neuron had a "preferred direction," for which its firing rate increased the most. During any one movement, however, neurons with many different preferred directions showed changes in activity. Cells tuned for directions close to the direction of the movement showed the largest increases in firing; those tuned to different directions showed proportionately smaller changes.

The direction of the actual movement, it appeared, reflected the collective activity of many different neurons. When the workers represented the change in each neuron's firing rate as a vector pointing in the cell's preferred direction, the sum of the vectors—the "population vector"—accurately matched the direction of the movement. Moreover, the population vector developed well before the actual movement.

Here, then, was a neuronal representation of an upcoming movement—a representation that might rotate, the investigators thought, as the monkey prepared to respond in the new task. To test the hypothesis, they monitored the activity of 102 neurons in

the monkey's motor cortex and determined their preferred directions. For each trial they calculated the population vector at intervals of 10 milliseconds beginning when the stimulus light appeared.

When the light was dim, directing the monkey to move the lever in the same direction as the cue, the population vector that developed in advance of the movement pointed steadily toward the target. The vector that developed in response to a bright light, however, initially pointed toward the stimulus but then rotated 90 degrees counterclockwise, toward the direction of the correct response. The monkey, it seemed, had rotated a mental representation of movement in deciding how to respond.

Like rotations in the real world, the mental rotation was continuous and measurable. The direction of the population vector changed steadily during the 100 milliseconds or so just before the monkey moved the lever; the rotation rate averaged about 730 degrees per second. The results, Georgopoulos says, represent "the first direct visualization of a cognitive process in the brain"—of what in humans would be called thinking. He and his colleagues believe studies of population vectors may soon yield equally precise pictures of other kinds of spatial cognition. —*Tim Appenzeller*

The Blood-Brain Barrier

More evidence links genes and intelligence

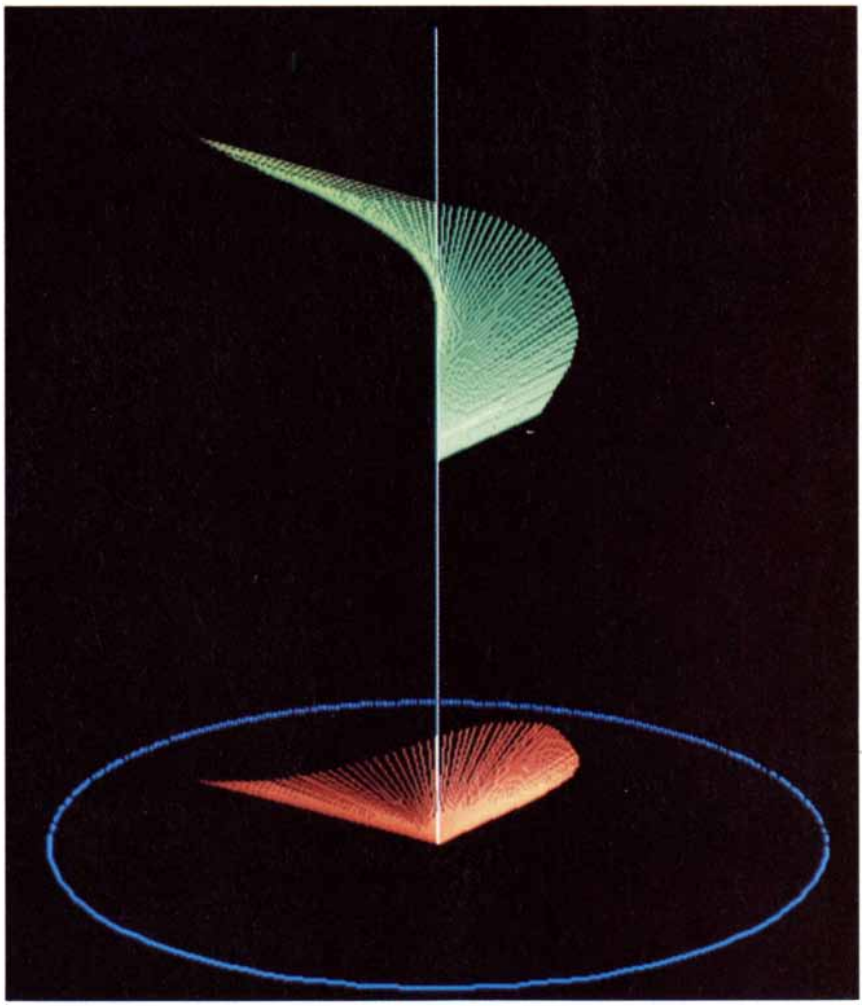
Johann Sebastian Bach had three sons who became accomplished musicians and composers. Irène Curie won a Nobel prize several decades after her parents Pierre and Marie were so honored. Erasmus Darwin, physician, philosopher and inventor, inspired his grandson Charles. The Bernoulli family produced eight exceptional mathematicians in three generations. History demonstrates that talent begets talent, but its examples cannot help geneticists to measure the relative influences of heredity and environment.

Evidence amassed over the past 60 years suggests there is a genetic component to intelligence, and the most recent piece of evidence comes from a study reported in *Nature*. The authors—David W. Fulker and John C. DeFries of the University of Colorado at Boulder and Robert Plomin of Pennsylvania State University—tested 245 adopted children and their biological and adoptive parents on various measures of cognitive ability such as perceptual skills, memory and abstract reasoning, all of which are represented by the so-called intelligence quotient (IQ). The children were tested when they were one, two, three and four years old and again after they had completed first grade.

Fulker and his colleagues found that the children's scores on the tests are more closely correlated with those of their biological parents than with those of their adoptive parents. Correlations with either set of parents are erratic when the children are from one to three years old, but the differences become more pronounced as the children get older. In the Colorado study correlations between the children and their biological parents increase from .15 at age three to .28 at age seven, whereas correlations between the children and their adoptive parents decrease from .16 to .06 during that time. (Fulker says the correlation between parents and children in nonadoptive homes is usually about .40.)

From those figures Fulker's group calculated that 36 percent of the variation in mental ability among individual seven-year-olds can be ascribed to genetic makeup. The rest of the variation can be ascribed to environmental influences. Like the correlations, the percentage of variance due to heredity increases with age. The project will follow the children until they are 16,

Pirouette of a vector reveals a brain at work



POPULATION VECTOR* (green), *representing the collective activity of neurons in a monkey's motor cortex, rotates counterclockwise by 90 degrees as the animal solves a spatial problem. The white line is the time axis; the vector's rotation is projected onto the base in red. Photograph courtesy of Apostolos P. Georgopoulos.

Celebrity. $11,495.* A great way to teach your kids the value of the dollar.

- Front-drive, four-door, 6-passenger family sedan.
- Electronic Fuel Injection
- All-season steel-belted radials.
- EPA estimated MPG city 23, highway 30.
- 3-year/50,000-mile Bumper to Bumper Plus Warranty.†
- Standard power steering and brakes.

THE Heartbeat OF AMERICA TODAY'S CHEVROLET

Show them how much you get in a Celebrity. Like room. A huge trunk. And lots of standards. Which is our way of giving you more for less than you'd expect. Then, show them the bottom line that's less than many family cars. Leaving you more to spend on your family. They'll like that part. Best.

*M.S.R.P. including dealer prep. Tax, license, destination charges and optional equipment additional. †See your Chevrolet dealer for terms of this limited warranty.

Chevrolet and the Chevrolet emblem are registered trademarks of GM Corp. ® 1988 GM Corp. All Rights Reserved. Let's get it together...buckle up. GM

and so the workers will get a chance to see if the trend continues.

They suspect it will. Other adoption studies have yielded heritable variations higher than 36 percent, and a longitudinal adoption study published in 1949 also suggested that genetic influences on intelligence become more apparent between early childhood and adolescence. Fulker says his study is probably more conclusive because the children in the Colorado project were not placed with adoptive parents who matched the educational level or socioeconomic status of the biological parents. Consequently correlations between children and their biological parents are more likely to reflect heredity rather than the effects of similar upbringings.

Fulker cautions against misinterpreting the group's results. "I would not like to derive any social-policy agenda from studies such as this," he remarks. "They don't really tell you anything about the malleability of intelligence." Indeed, some critics maintain that the studies do not tell you anything, period. Leon J. Kamin of Northeastern University cites cases in which markedly different correlations have been calculated from the same body of data. "You're really looking at the biases of the investigator," he says. —*Karen Wright*

The Syphilized World

A recent survey suggests that syphilis is a New World disease

The origin of syphilis has long been a contentious subject. One theory holds that the disease originated during the evolution of the genus *Homo* and has therefore been present in all human populations. According to the opposing theory, syphilis originated in the New World and was brought to Europe by early explorers, perhaps by Columbus' crew. A new survey of the available evidence by Brenda J. Baker and George J. Armelagos of the University of Massachusetts at Amherst finds a New World origin to be the most probable. Among the recent data are immunological tests indicating that a bear living in Indiana more than 11,000 years ago was infected with the syphilis agent.

Baker and Armelagos' review, published in *Current Anthropology,* covers documentary evidence and observations of skeletal material from both the Old World and the New. Syphilis, which is caused by a class of anaerobic bacteria called treponemes, leaves characteristic marks on bones, particularly the skull and long bones such as the tibia and fibula. These lesions serve as definitive identifying marks of the disease. The authors found such traces in abundance in New World specimens but not in Old World skeletons. Indeed, they found that there are no unambiguous syphilitic skeletal lesions among European skeletons that have been securely dated to before 1492.

"I was very surprised by the absence of definite material from Europe," Armelagos said recently. "The predominant scholarly view in recent years has been that syphilis was present in Europe well before Columbus. If that were the case, I would have expected to be hit with all kinds of securely dated, unambiguous cases from the European material. But that wasn't the case. There's a real paucity of material from Europe and every case is ambiguous—either in terms of its date or in terms of the diagnosis."

In the New World the situation is quite different. Well-dated and unambiguous cases of treponemal infection have been reported in every major region of the Americas. The first scattered cases, Armelagos said, come from South America and are in human bones that are from 3,000 to 4,000 years old. By the Mississippian period, he added, in about A.D. 600, the characteristic treponemal lesions appear to be quite widespread, particularly in what is now the eastern part of the U.S.

The origin of syphilis might have taken the following course, Armelagos hypothesizes. It is known that treponemes can be spread by nonvenereal means. In pre-Columbian America the disease might have been a mild, nonvenereal infection spread, for example, by children huddling together for body warmth on cold nights. By the Mississippian period agricultural communities with sizable population centers had come into being, and such communities would have provided an excellent setting for the nonvenereal transmission of treponemes.

The bacteria were capable of venereal transmission as well, however, and that was probably how they were acquired by the crew on Columbus' first voyage. The Spanish crew brought the pathogen back with them to their homeland. Not long afterward Spanish mercenaries were involved in the siege of Naples carried out by Charles VIII of France. In that siege the mercenaries were devastated by a plague thought to be syphilis. When they disbanded, the disease spread to the rest of Europe in venereal form.

But where had human beings first acquired the bacteria? The answer may be by zoonosis, that is, from other species. In a paper in *Nature* in 1987 Bruce M. Rothschild of the St. Elizabeth Hospital Medical Center in Youngstown, Ohio, and William Turnbull of the Field Museum of Natural History report applying modern immunologic techniques to the bones of a bear found in Indiana and radiocarbon-dated to about 11,500 years ago. The bones, which showed typical treponemal lesions, also tested positive for the characteristic antigens, or proteins, of the treponemes, according to Rothschild and Turner.

Although the bear data are still controversial (some scholars think the positive results are due to later admixtures from the surrounding soil), they do suggest that treponemal infection might have entered human beings through animal species in the New World before being transmitted to Europe during the era of exploration. At any rate, the debate over the origins of syphilis seems to be heading toward resolution. Final resolution will probably come, Baker and Armelagos note, when immunologic methods are more widely applied to human bones from both the Old and the New World. —*John Benditt*

MEDICINE

tat's Surprising

The HIV protein tat *may do unexpected damage*

A clue that may help to explain how the AIDS virus replicates and damages cells has been discovered. The AIDS-causing human immunodeficiency virus (HIV) carries a gene known as *tat.* Investigators have long known that when *tat* is activated in an infected cell, the protein it encodes enters the nucleus and stimulates massive viral replication, probably by activating a segment of the viral genome called the promoter (which itself stimulates other HIV genes). Such replication can ultimately kill a cell. Workers have assumed that the *tat* protein is active only in the individual cells in which it is produced. The new finding raises the possibility that the substance may do mischief in other cells as well.

Alan D. Frankel of the Johns Hopkins University School of Medicine and Carl O. Pabo of the Howard Hughes Medical Institute there and, independently,

You've stopped smoking, cut your cholesterol and reduced your caffeine.

When are you going to start taking care of yourself?

All it takes is the Lifecycle® Aerobic Trainer and 12 minutes a day, a few days a week!

NOW ONLY $39.95 A MONTH!*

After all you've done for your health, DON'T YOU HAVE 12 MINUTES A DAY FOR THE EXERCISE YOU NEED to look better, feel better and live longer? Recent clinical studies confirm THAT'S ALL THE TIME IT TAKES—12 MINUTES A DAY!

You can get a challenging, exciting and effective aerobic workout in only 12 minutes on the computer-controlled, variable resistance Lifecycle® aerobic trainer. And training longer can help you lose weight and keep it off! That's why millions of exercisers have made it THE most popular computerized bike in health clubs, nationwide.

AND NOW YOU CAN HAVE YOUR OWN LIFECYCLE® TRAINER IN YOUR HOME FOR ONLY $39.95 A MONTH, INTEREST-FREE, NO MONEY DOWN!*

That makes exercising convenient *and* affordable. Everyone in your family can use it at their convenience—no commuting, no babysitters, no problems with the weather!

TRY IT IN YOUR HOME FOR 30 DAYS, RISK-FREE!

We'll give you an UNCONDITIONAL 30 DAY, NO-RISK TRIAL. If after 30 days you aren't convinced, we'll pick up the bike at our expense and refund any money you've paid, no questions asked!

SPECIAL LIMITED TIME OFFER
FREE AIR SHIPPING AND HANDLING

We usually charge $130 for the air delivery shipping and handling. Order before May 1st and we'll provide that for you FREE. Save $130. ORDER TODAY!

—Train with the best—No down payment—NO-COST FINANCING—An UNCONDITIONAL 30 DAY TRIAL—And if you CALL NOW, FREE SHIPPING AND HANDLING WORTH $130.

Life Fitness INC. A Bally Company

* No down payment. 0% APR. $39.95 per month for 42 months = $1677.90 plus applicable sales tax. Offer subject to credit approval. Lifecycle® is a registered trademark of Life Fitness, Inc., A Bally Company.

START REALLY TAKING CARE OF YOURSELF TODAY! CALL 800-877-3867, DEPT. 014 to order, for more information or for the name of your nearest retail dealer.

Maurice Green and Paul M. Lowenstein of the St. Louis University School of Medicine report in *Cell* that free *tat* protein can enter mammalian cells in culture. Moreover, if those cells include an HIV promoter in their nucleus, the protein strongly activates it.

No one knows whether the *tat* protein is active outside its cell of origin in vivo and, if it is, to what extent. Nevertheless, the findings suggest a number of interesting possibilities. For instance, free *tat* protein, perhaps released into the circulation from infected cells that have died, might enter other cells in which HIV is dormant and act as a growth factor to stimulate viral replication. This process could potentially contribute to the marked increase in HIV replication found in the late stages of HIV disease.

There is also a chance, Pabo says, that the *tat* gene descends from a normal cellular gene. If it does, and if the *tat* protein can still influence normal genes, the protein may derange the functioning not only of HIV-infected cells but also of healthy cells.

If strong proof is found that the *tat* protein is peripatetic and has a significant effect on distant cells in the body, the finding could open the way to new research into treatment approaches for HIV. For example, chemical agents might be developed that would prevent the molecule from entering cells. Perhaps a vaccine could stimulate the body to raise its own antibodies against the protein.

The two groups discovered the cellular uptake of free *tat* protein somewhat inadvertently, in the course of investigating the molecular mechanisms by which the protein interacts with the HIV genome. These studies require assays for measuring the activity of synthetic or purified *tat* protein in cultured cells. Green and Lowenstein found that their normal methods for getting the substance into cells were tedious, and so they tried simply mixing the protein with the culture medium. Frankel and Pabo added the *tat* protein to the medium as a control measure, to compare the expected negligible effect of that step with the effect of also scraping the surface of the cells (thereby facilitating the protein's entry).

Various investigators have already shown that the blood of AIDS patients frequently harbors antibodies to the *tat* protein. This suggests that the protein probably does enter the circulation at times. Exactly what it does after that is a question several laboratories expect to be pondering for some time to come. *—Ricki Rusting*

Lukewarm Turkey

Drug firms balk at pursuing a heroin-addiction treatment

The number of heroin addicts in the U.S. has remained roughly constant at about 500,000 for the past decade, according to the National Institute on Drug Abuse (NIDA). Yet during this period the impact addicts have had on society has grown enormously. Intravenous drug abusers, many of whom are heroin addicts, are the fastest-growing population of AIDS victims and the main avenue through which the disease reaches the heterosexual population.

NIDA, a Federal research agency, is trying to develop drugs that can supplement or replace methadone as treatments for heroin addiction, but the institute is having difficulty convincing pharmaceutical companies to support its effort. A drug called buprenorphine provides a case in point. Developed nearly 20 years ago by the British company Reckitt and Coleman Products, Ltd., buprenorphine has been approved by the U.S. Food and Drug Administration as an analgesic for acute, chronic pain. An opioid, it comes in two forms: an injectable liquid and a tablet meant to be dissolved under the tongue.

Methadone, a nonintoxicating (except at high doses) substitute for heroin, has enabled many addicts to rid themselves of the crippling financial and physical burden of heroin addiction. Unfortunately withdrawal from methadone, although not as agonizing as withdrawal from heroin, lasts much longer: about two weeks instead of five or six days. Some investigators think this extended withdrawal prevents many addicts from attaining what should be their ultimate goal: a totally drug-free existence.

Clinical trials at various institutions, including Johns Hopkins University, Yale University and NIDA itself, suggest that buprenorphine may be as effective as methadone as a heroin substitute. According to Jack D. Blaine, chief of treatment research at NIDA, buprenorphine satisfies the craving of addicts; it also dulls the "high" of other opioids and thus discourages backsliding to heroin. Buprenorphine may have even more potential for getting addicts off opioids entirely. Withdrawal from buprenorphine lasts for only a few days and the symptoms are "very minor," Blaine says.

One problem with buprenorphine is that, because the tablets can be dissolved in water and injected to deliver an intoxicating high, it can be illegally diverted to intravenous drug abusers. John W. Lewis of Reckitt and Coleman, who developed buprenorphine and still advises NIDA on the drug, says combining it with naltrexone, an "opioid antagonist" that blocks the effects of opioids, could eliminate that problem. Taken orally, naltrexone has little or no effect; injected, it completely blocks the pleasurable effects of buprenorphine or any other opioid. Blaine says naltrexone may also help to keep addicts who have kicked their opioid habit from relapsing.

Naltrexone has been approved by the FDA as a treatment for heroin addiction, but buprenorphine has not, and it may never fulfill its promise unless NIDA can convince a drug manufacturer to apply for such approval. About a year ago Norwich Eaton Pharmaceuticals, Inc., a subsidiary of Procter & Gamble Co. and the sole company licensed to sell buprenorphine in the U.S., agreed to apply jointly with Reckitt and Coleman for approval of the drug as an addiction treatment; NIDA would have funded the required clinical trials. Last summer, however, Norwich Eaton backed out of the agreement, citing its "limited research resources." Since then Reckitt and Coleman, which has supplied U.S. investigators with buprenorphine and advice for years, has decided it cannot afford to pursue FDA approval alone.

Both Reckitt and Coleman and Norwich Eaton are reluctant to discuss the precise reasons for their decisions. But Mary Jeanne Kreek, who does drug-abuse research at Rockefeller University, says their decisions are understandable, given the history of methadone. Eli Lilly and Company developed methadone as a painkiller, but after it became associated with heroin addiction many doctors stopped prescribing it; the FDA also reclassified the drug, making it more difficult and expensive to dispense. The company's income from addiction-treatment programs, which usually provide methadone free of charge, is minimal. "Eli Lilly has not made a penny off methadone," Kreek says.

Kreek has proposed various measures that might encourage drug companies to develop addiction treatments. For example, a drug that has two indications, one profitable and the other addiction-related, could be marketed under two names. Blaine says such measures will certainly be considered by a task force NIDA has recently established to search for drugs that can help victims of either heroin or that other scourge, cocaine. *—J.H.*

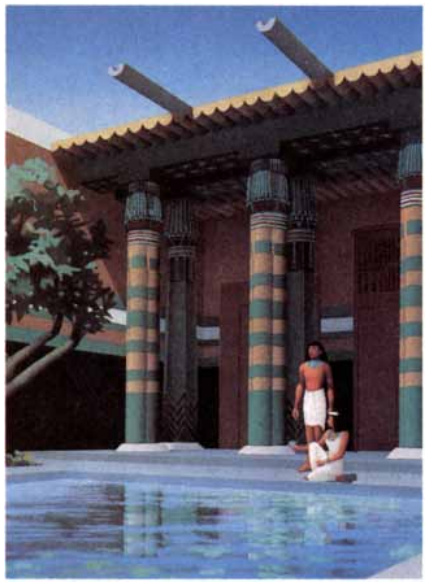
NEW LEADER EMERGES IN CHINA.
WAR RIPS MIDDLE EAST.
EARTHQUAKE SHOCKS EUROPE.
WELCOME TO 3000 B.C.

Much more was going on 5,000 years ago than most history books would lead you to believe.

Not only politics and war, but medical prescriptions, marriage contracts, games, even indoor plumbing were just as much a fact of life as they are today.

Introducing TIME FRAME, a fascinating new series from TIME-LIFE BOOKS that lets you experience the past with all the drama and color you find in today's most revealing news coverage.

Each volume of TIME FRAME gives you a culture-by-culture look at life during a single era. From Africa to Japan, from India to the Americas, you'll see how cultures made contact, influenced each other and changed the course of civilization.



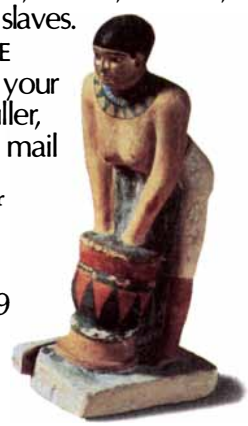
Visit homes you wouldn't mind living in today.

Science, art, religion, politics, fashion—with vivid recreations and little-known details, TIME FRAME brings the past to life as never before.

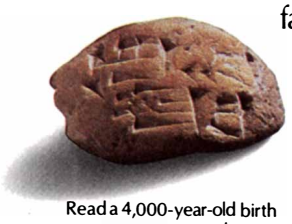
Your journey begins with *The Age of God-Kings 3000–1500 B.C.*—for 10 days free. Explore rarely seen areas of the mysterious pyramids. Uncover the secrets of Stonehenge. Admire stunning jewelry of gold, silver, ivory and lapis lazuli. And meet people like the legendary King Minos of Crete, who was assassinated while taking a bath.

Follow in the footsteps of emperors, pharaohs, warriors, traders and explorers. Share in the daily life of painters, sculptors, scribes, farmers, tax collectors and slaves.

Let TIME FRAME transport you and your family back to a fuller, richer past. Simply mail the reply card to receive *The Age of God-Kings 3000–1500 B.C.*—for 10 days' free examination. Keep it for just $14.99 plus shipping and handling, or return it with no obligation.



Toast the beginnings of beer-making.

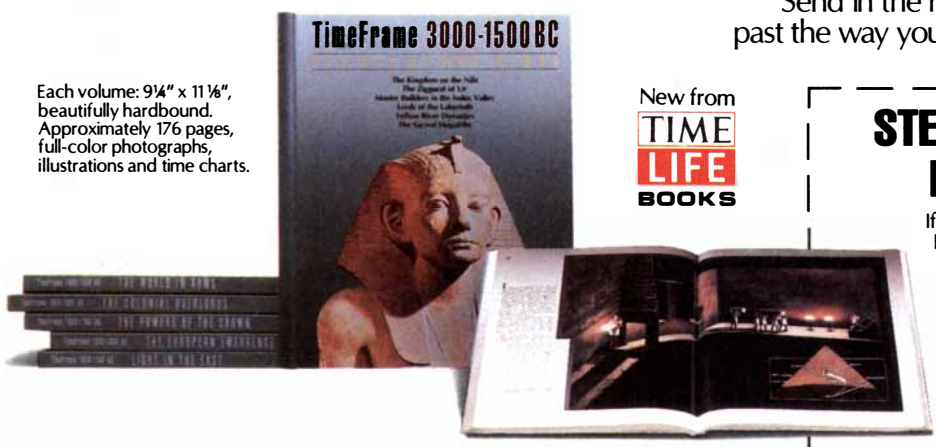


Read a 4,000-year-old birth announcement.

Future volumes, including *Barbarian Tides 1500–600 B.C.*, *A Soaring Spirit 600–400 B.C.*, and *The Colonial Overlords 1850–1900 A.D.*, arrive one about every other month. Same free trial—keep only the ones you want, cancel anytime.

Send in the reply card and start looking at the past the way you see the world today.

Each volume: 9¼" x 11⅛", beautifully hardbound. Approximately 176 pages, full-color photographs, illustrations and time charts.



New from TIME LIFE BOOKS

STEP BACK 5,000 YEARS FOR 10 DAYS FREE.

If card is missing, mail to: TIME-LIFE BOOKS, Box C-32066, Richmond, VA 23261-2066

YES! I would like to have *The Age of God-Kings 3000–1500 B.C.* as my introduction to the TIME FRAME series. Please send it to me for a 10-day free examination. Also send me future volumes under the terms described in this ad.

D9BFM5

Name_____ (Please print)

Address_____ Apt._____

City_____

State or Province_____ Zip or Postal Code_____

All orders subject to approval. Price subject to change.

Time Frame

Traditional history just became a thing of the past.

SCIENTIFIC AMERICAN March 1989 Volume 260 Number 3

U.S. Access to Space

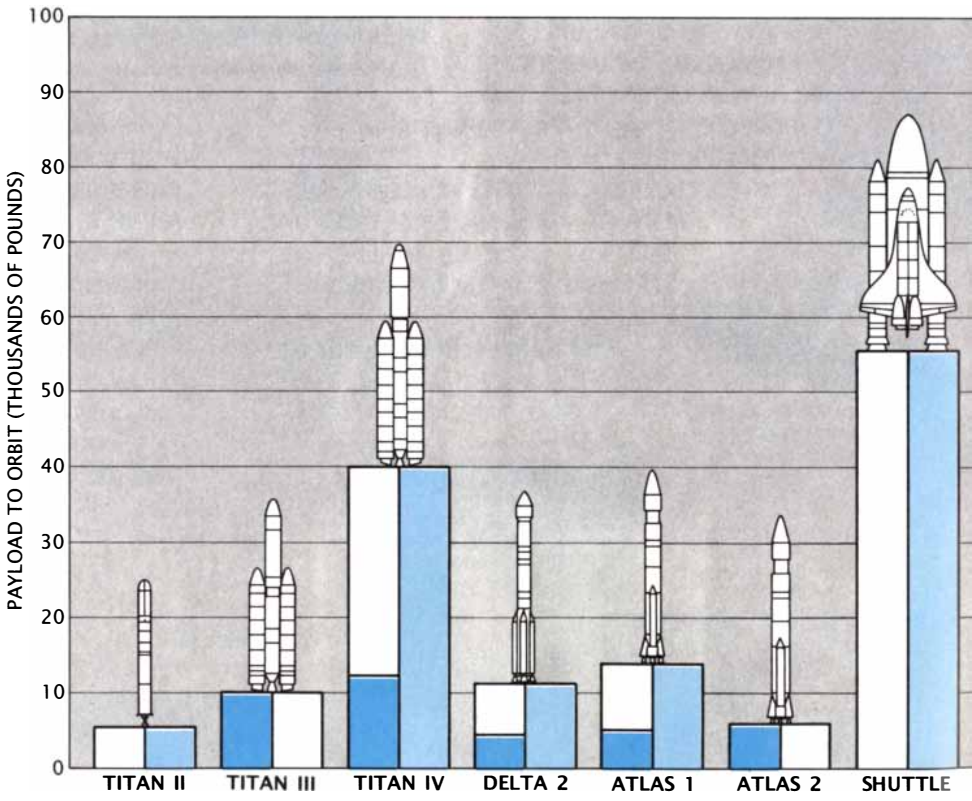
For the next decade the U.S. has to rely on the shuttle and existing expendable launchers to get into space. Before launchers for the 21st century are developed, space-program goals should be established

by John M. Logsdon and Ray A. Williamson

The U.S. is just emerging from the most serious crisis in the history of its space program. The suspension of space-shuttle flights after the 1986 *Challenger* disaster along with concurrent failures of the nation's three workhorse expendable launch vehicles (ELV's) denied the U.S. full access to space for almost three years. As a result many scientific, military and commercial space projects have not moved forward. Expensive spacecraft such as the *Galileo* probe to Jupiter and the Hubble Space Telescope have been aging in storage while their associated scientific teams fume over launch delays of up to five years. Critical national-security satellites could not be put into orbit to replace pre-1986 satellites that had already exceeded their predicted lifetimes. China and the Soviet Union joined the European Arianespace consortium in the global commercial space-launch market while American manufacturers of ELV's could only look on while they waited for the Government to formulate a policy enabling them to enter that very market.

Although the space shuttle has recently resumed flying and a variety of U.S. ELV's are now becoming available, the next decade offers little encouragement for markedly improved capabilities in space transportation. A recent study by the congressional Office of Technology Assessment (OTA)

JOHN M. LOGSDON and RAY A. WILLIAMSON are long-time observers of U.S. space policy. Logsdon is director of the Space Policy Institute at George Washington University, where he is also professor of political science and international affairs. He has a B.S. in physics from Xavier University and a Ph.D. in political science from New York University. Logsdon was a member of a committee recently organized by the National Academy of Sciences and the National Academy of Engineering to examine the country's space policy. Williamson, who holds a Ph.D. in astronomy from the University of Maryland, is a senior associate at the Office of Technology Assessment (OTA) of the U.S. Congress. He is currently directing an OTA assessment of space transportation systems.



SPACE LAUNCHERS available to national governments or commercial customers in the 1990's include two reusable shuttles and an assortment of expendable launch vehicles (ELV's) from the U.S. (*blue*), the U.S.S.R. (*red*), Europe (*purple*), China (*green*) and Japan (*orange*). Some launchers will be used solely to put satellites into a so-

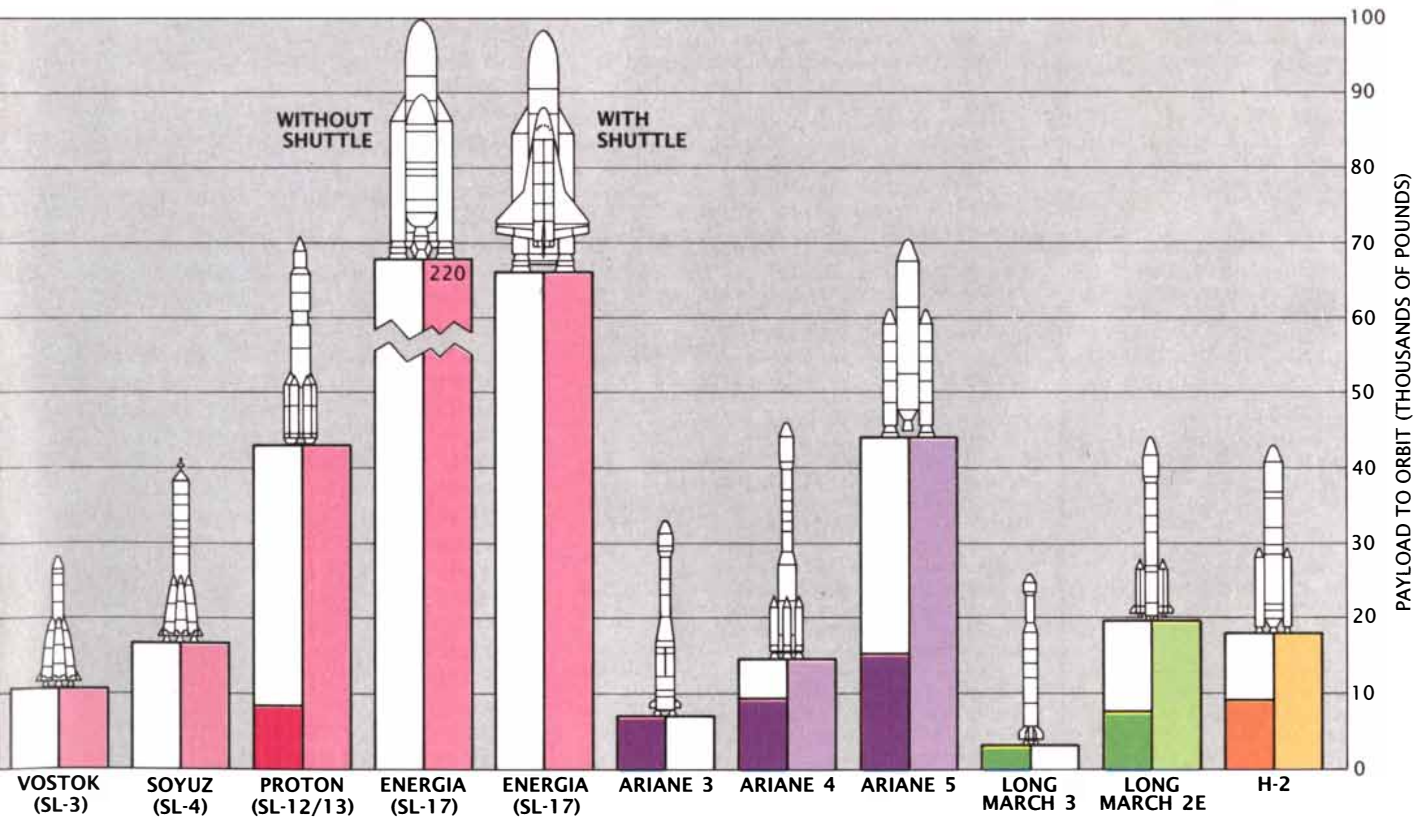
concluded that the late 1990's is the earliest a launch system incorporating significant new technology could enter regular service and that a replacement for the shuttle is not likely until 2005 or later. According to the OTA, if the demand for launches remains near its current level and launch failures are infrequent, existing systems "are capable of meeting U.S. space-transportation requirements." But any plans that call for a significant expansion of U.S. space activities during the 1990's will be limited by the lack of robust, diverse and affordable space-transportation systems.

The U.S. finds itself in this unfavorable situation as a consequence of a massive policy mistake: during the 1970's grossly unrealistic expectations regarding the shuttle's cost and capability led to a national space program that depended solely on the shuttle for access to space. Hence from 1972 until recently virtually no Government funds were invested in other launch-related technologies or in new launch facilities. (The West Coast launch site for the shuttle, which cost more than $3.5 billion, is an embarrassing exception, since it is now mothballed and unlikely ever to be used.)

To make matters worse, in the early 1980's, when the European Ariane ELV's began to be marketed successfully as alternatives to U.S. systems for putting communications satellites into geosynchronous orbit, the Government responded by actively promoting the shuttle as the American entry in that global commercial competition. By 1985 marketeers from the National Aeronautics and Space Administration were roaming the world with glossy brochures stating that the shuttle was "the most reliable, flexible and cost-effective launch system in the world" and asserting to prospective customers that "you can't get a better price." (Actually the price for the shuttle's service was artificially low; it had originally been set as a way to attract U.S. customers.) The Government's mercantilist response effectively blocked attempts by U.S. private companies to commercialize their ELV's and forced them to suspend production of the vehicles altogether: the prospects were dim not only for future orders from the U.S. Government but also for orders from foreign governments and private companies.

Hence when shuttle flights ceased after the *Challenger* exploded, the U.S. had no supply of alternative launchers immediately available. It has taken nearly three years both to get the shuttle flying again and to restore full production of ELV's. And it will take much longer to build the technology base for future launch systems.

Although the U.S. Department of Defense has taken steps in its plans for the 1990's to have a variety of launch systems at its disposal, NASA will continue to rely primarily on the shuttle. Indeed, as a result of its limited budget and its desire to keep the space-station program on schedule, the agency has decided that the shuttle will be the sole means for launching space-station components into orbit and the primary vehicle for ferrying crews and supplies once the station begins operation. The fact that the major U.S. civilian space project for the 1990's is so dependent on reliable and regular shuttle operations evokes a discomforting sense of déjà vu, particularly since the probability is high that the shuttle will experience serious problems in the future—including some that could once again ground the entire fleet.



called geostationary transfer orbit (*dark color*), which brings the satellite to a point 22,300 miles above the earth, where it can then be injected into a circular orbit that keeps it above the same spot on the earth at all times. Other launchers will be used mainly to put payloads into lower orbits (*light color*). (The low-earth-orbit payloads shown are calculated for a circular orbit that lies between 115 and 150 miles above the earth.) Certain launch vehicles can be employed for both purposes.

Currently the U.S. has a mixed fleet of launchers [*see illustration on preceding two pages*] for putting payloads of between 5,000 and 55,500 pounds into low earth orbit (ranging in altitude from 100 to 500 miles above the earth's surface) and payloads of between 1,000 and 12,000 pounds into geosynchronous orbit (22,300 miles above the earth). Such payloads have constituted the bulk of Government and commercial space cargo to date. (A few entrepreneurial companies are developing new launchers specifically for lighter payloads in the belief that low-weight satellites will become commoner in the future. Such smaller launch vehicles will not be discussed in this article.) At planned production rates, by 1990 the U.S. fleet will be capable of launching a total of 860,000 pounds into orbit per year. For comparison, a total of approximately 600,000 pounds was launched into space by the U.S. in 1985. At least in overall terms, then, there is no shortage of launch capacity for the time being.

IF THE U.S. WANTS TO	THEN THE U.S. SHOULD
limit growth of NASA and DOD space programs,	maintain existing launch systems and limit expenditures for future systems. Current capabilities are adequate to meet present levels of U.S. space activity.
deploy a space station by the mid-1990's while maintaining an ambitious NASA science program,	fund the development of improvements to the space shuttle, such as advanced solid-fuel and liquid-fuel boosters, or of the shuttle-C, an unmanned cargo vehicle. The current shuttle can transport space-station components into orbit, but it will do so more effectively if its design is improved or if it operates in conjunction with a heavy-lift cargo vehicle.
send humans to Mars or establish a moon base,	develop an unmanned cargo vehicle (either NASA's shuttle-C or the Air Force's Advanced Launch System) and continue funding the Advanced Manned Launch System and the National Aerospace Plane, since a commitment to piloted space flight will require a shuttle replacement by the end of the century.
continue launching heavier satellites or pursue an aggressive SDI test program,	develop an unmanned cargo vehicle based on existing technology by the mid- to late 1990's. In theory current launch systems could be expanded to meet future needs, but new systems are likely to be more reliable and less expensive to operate.
deploy a space-based strategic defense system,	develop an unmanned cargo vehicle, such as the Air Force's Advanced Launch System. Current launch systems are neither economical nor reliable enough for the task.
maintain U.S. leadership in launch-system technology,	increase funding for space-transportation studies. Maintaining leadership will require an integrated NASA/DOD development program that includes a range of technologies. Focused development efforts should be balanced with basic research.
improve its ability to recover quickly from a launch-system failure,	improve the reliability of current launch vehicles or develop a new high-capacity, high-reliability launch vehicle and expand current ground facilities. If high launch rates are planned, the new vehicle is probably the most economical option.
increase launch-vehicle reliability and safety,	aggressively fund improvements in subsystem reliability, redundancy and fault tolerance and the development of on-pad abort capabilities and of rocket engines that can be shut off in flight.
reduce the environmental impact of high launch rates,	limit use of toxic liquid fuels and replace shuttle and Titan solid-fuel boosters with new liquid-fuel boosters or clean-burning solid-fuel boosters.

POSSIBLE GOALS for a U.S. space-transportation system and the steps necessary to achieve them were examined in a recent report by the congressional Office of Technology Assessment (OTA). The report's recommendations, which are summarized in the table, depend on what the Government would like to accomplish in space.

All three U.S. ELV types—Delta, Atlas and Titan—trace their heritage to ballistic missiles that were developed in the late 1950's and early 1960's. These rockets have been modified considerably since their first incarnation, and the current versions have nearly reached the limits of the capability imposed by their basic design. The principal U.S. user of ELV's during the next decade will be the Defense Department, which currently has on order 11 Atlas 2's, 14 Titan II's, 20 Delta 2's and 23 Titan IV's. The total value of these orders is some $6.3 billion. Additional orders are expected, since the department intends to establish a stockpile of launch vehicles.

That was not the way it was originally planned. The department actually committed itself in the 1970's to purchasing its launch services exclusively from NASA, using the space shuttle for the delivery of all its payloads into orbit. (In 1985, however, the Air Force was able to persuade the White House that it needed a backup system for launching the most critical national-security payloads, and it contracted for the development of the Titan IV even before the *Challenger* explosion.)

This is not to say that the Defense Department will not make substantial use of the shuttle over the next few years. Many national-security payloads designed before 1986, including some that have been grounded since the *Challenger* disaster, can be put into orbit only from the shuttle. In the longer run, however, the department is determined not to depend on any launch system it does not control. Indeed, during the mid- and late 1990's the military will rely on the shuttle for an average of only two missions per year—those that specifically require the shuttle's unique capabilities. In an ironic reversal of roles, the shuttle will also serve as a backup to the Titan IV, which has been designated the main Defense Department launch vehicle for heavy payloads in the 1990's. (That decision carries its own risks, however, since the Titan IV is an unproved system. Its first launch is scheduled for March of this year.)

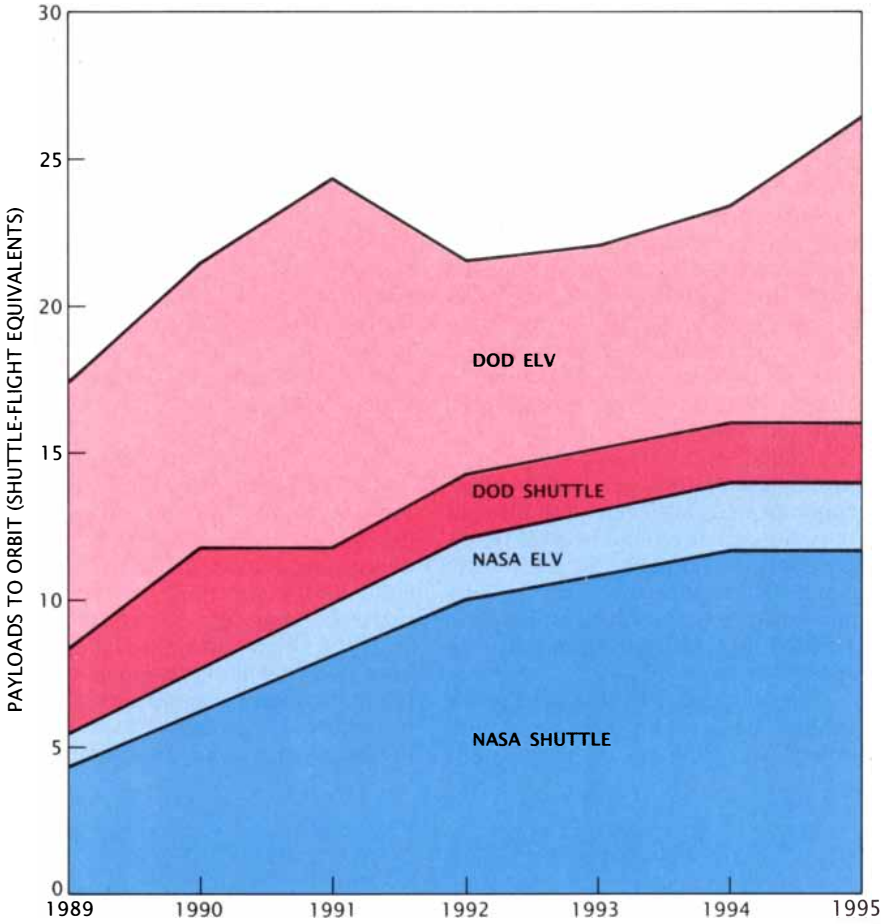
What, then, is the future of the U.S. space shuttle in the civilian space program for the 1990's? Although the successful per-

formance of the *Discovery* orbiter last September removed a tremendous psychological burden, the more fundamental questions of how frequently and reliably the current three-orbiter shuttle fleet can be launched—and at what cost—will not be answered for several years. The fate of the Government's civilian space program depends greatly on those answers. Until the U.S. has them, the rate at which the country's space science, space technology and space exploration can advance will remain uncertain.

In January, 1989, NASA issued an optimistic schedule that called for 16 shuttle and seven ELV launches during the 1989-90 period (not counting launches of the small Scout ELV) and 60 shuttle launches between 1989 and 1994. The ability to meet that schedule will be a fair test of NASA's policy of continued reliance on the shuttle. After a fourth orbiter enters the U.S. shuttle fleet in 1992, NASA projects a launch rate of between 12 and 14 flights a year. Some observers, however, suggest that between eight and 10 flights per year is a more realistic estimate. At least the illusion that the shuttle will be inexpensive to operate appears to have vanished. Although precise costs per launch are almost impossible to establish, they are likely to range between $250 and $500 million, depending on the actual number of flights per year.

The shuttle's value to potential users can be increased if certain crucial capabilities are augmented by means of technological improvements. Currently the shuttle can remain in orbit for a maximum of about nine days. Because it usually takes a shuttle crew a day to become acclimated to the weightlessness of space and a day to prepare for the return to the earth, the crew has really only a few days to complete its major experimental tasks in orbit. Modifying one or two shuttle orbiters for extended stays of up to 16 days or more would greatly enhance the orbiter's capability to function as a laboratory in space for research in areas such as microgravity science.

The shuttle's weight-lifting capability, a space launcher's main figure of merit, could also be increased. In its original design the shuttle had a maximum payload of 65,000 pounds, but modifications made in the aftermath of the *Challenger* accident have reduced the maximum payload to 55,500 pounds. Replacing the current solid-fuel boosters with more powerful ones may increase the shuttle's maximum lift capacity to 67,500 pounds.



SPACE FLIGHTS planned for the near future by the Department of Defense (*red*) and the National Aeronautics and Space Administration (*blue*) differ markedly with respect to launch-vehicle type. Whereas the DOD has shifted the bulk of its flights to ELV's (*light color*), NASA has opted to continue relying heavily on the shuttle (*dark color*).

NASA believes such an enhanced shuttle will provide an adequate means of transporting the various components of the space station into orbit. Indeed, 20 shuttle flights between 1995 and 1998 are currently set aside for that purpose.

Such continued reliance by NASA on the shuttle as its primary launch vehicle has been criticized by many observers of the civilian space program. In general the scientific community is disturbed by the fact that it will have to depend on the shuttle for most of its space missions, because the presence of astronauts adds costs and complicates mission planning without any compensating advantages in performance or reliability. Indeed, space scientists have long argued that NASA should use more ELV's—particularly for planetary missions, which must lift off within a few days of schedule or wait a year or more for the next launch window. (Although NASA recently shifted the 1992 Mars Observer mission from the shuttle to a Titan III ELV in response to such concerns, the shuttle will be used for launching the *Magellan* space probe to Venus and the *Galileo* space probe to Jupiter respectively in April and October of this year.)

Some observers are also disturbed by NASA's decision to rely on the shuttle as the sole launch system for the space station. What happens, they ask, if another shuttle failure grounds the entire fleet while the station is under assembly or in operation? Even if the shuttle is 99 percent reliable, they point out that a simple statistical analysis shows there is an even chance of one major failure in the next 75 missions. In order to keep the space station continually operating after it is assembled and commissioned, it may be necessary to develop alternatives to the space shuttle for taking crews and supplies to and from the station. In fact, ELV's are already being con-

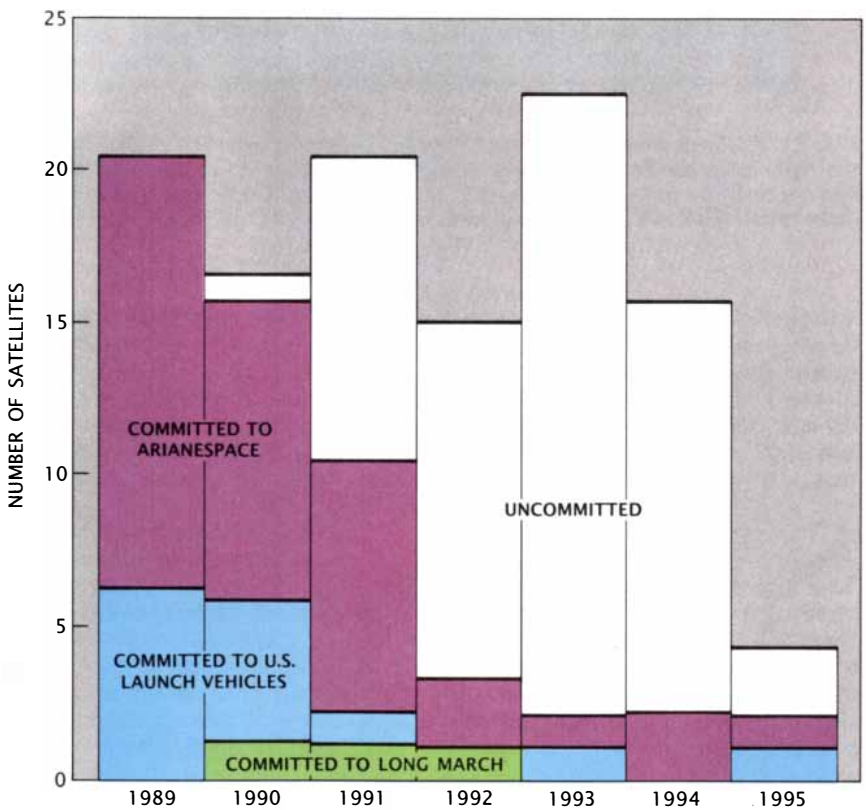
sidered for station-resupply missions.

In addition, a heavy-lift ELV could prove invaluable for the space station's assembly. Such an ELV could boost into orbit fully outfitted space-station modules that would later be assembled by shuttle astronauts. Unfortunately no U.S. ELV with the necessary payload capacity for such a mission exists. To provide the requisite heavy-lift capability, NASA engineers at the agency's Marshall Space Flight Center have suggested developing a large unmanned vehicle based on the shuttle's main engines, external fuel tank and solid-fuel boosters. The vehicle, called the shuttle-C (for cargo), would essentially have a relatively unsophisticated cargo carrier in place of the shuttle's orbiter. The shuttle-C's proponents maintain that the vehicle could be operational as early as 1994 and estimate that it could lift between 100,000 and 150,000 pounds to the space station's orbit.

They also estimate that developing the shuttle-C would cost about $1.5 billion. Nevertheless, if the vehicle were used just four times during the construction of the space station, the number of piloted shuttle missions could be cut by more than half. In addition, since the shuttle-C could carry preassembled modules into space, it would reduce the amount of risky extravehicular activity by astronauts that is needed for the station's assembly. There are also other potential missions for a vehicle like the shuttle-C, such as launching large solar-system probes and carrying into orbit additions to the space station and heavy experimental platforms for the Strategic Defense Initiative.

In spite of the fact that NASA could lessen its reliance on the shuttle by including additional ELV's in its plans (as the Defense Department has) or by developing a heavy-lift launch system (such as the shuttle-C), neither alternative is being aggressively pursued. Both the White House and Congress have been unwilling to provide funds for substantial ELV purchases, and the top levels of NASA as well as the Office of Management and Budget have resisted adding the cost of developing an interim heavy-lift launch system to an already tight NASA budget.

Experts disagree vigorously, then, about current U.S. space-transportation policy as it applies to the civilian space program. Substituting ELV's for the shuttle whenever possible, while feasible in principle, would be costly, and it might actually be superfluous if the post-*Challenger* shuttle system proves capable of operating regularly and reliably. On the other hand, if the shuttle again fails to meet expectations, the nation's claim to be a leader in space exploration and exploitation will be rightfully questioned. Revising U.S. space-transportation plans for the next decade to place added emphasis on the use of ELV's not only in military but also in civilian Government projects may well be necessary to ensure continued leadership in space.



COMMERCIAL LAUNCH SERVICES are likely to be rather competitive in the early 1990's. Most private organizations or countries that want to put payloads into orbit within the next two years have already contracted with the European Arianespace consortium, but as of January, 1989, the launchers for a total of 58 projected commercial payloads had yet to be selected. Such an open market offers a business opportunity for U.S. ELV manufacturers, although they will have to vie for customers not only with Arianespace but also with China and perhaps with the U.S.S.R. and Japan.

Although the U.S. Government will be the major user of launch services in the next decade, there will also be a significant global demand for commercial launch services. Estimates of the size of the commercial launch market vary considerably but average close to $1 billion per year. That market is now open to U.S. manufacturers of ELV's, namely McDonnell Douglas, General Dynamics and Martin Marietta, since the principal impediment has been removed: after the *Challenger* accident the Reagan Administration reversed its earlier policy and barred NASA from using the shuttle to compete for commercial launch contracts.

The economic implications of a U.S. commercial launch industry are not insignificant. Each launch contract is worth between $40 and $60 million, sometimes more. According to one authoritative market projection, there are 58 future launch contracts (worth a total of some $3 billion) that have not yet been committed to a specific launch-services supplier.

Yet before U.S. corporations could sell ELV space-transportation services many important questions had to be resolved. The list included the terms of access to U.S. Government launch, range-safety and tracking facilities and the provision of insurance against damage to those facilities as well as third-party liabilities. In fact, it took more than two years for the Government to lay out a policy framework the manufacturers were willing to accept. Nevertheless, U.S. launch-services providers have been making significant inroads in the global launch market;

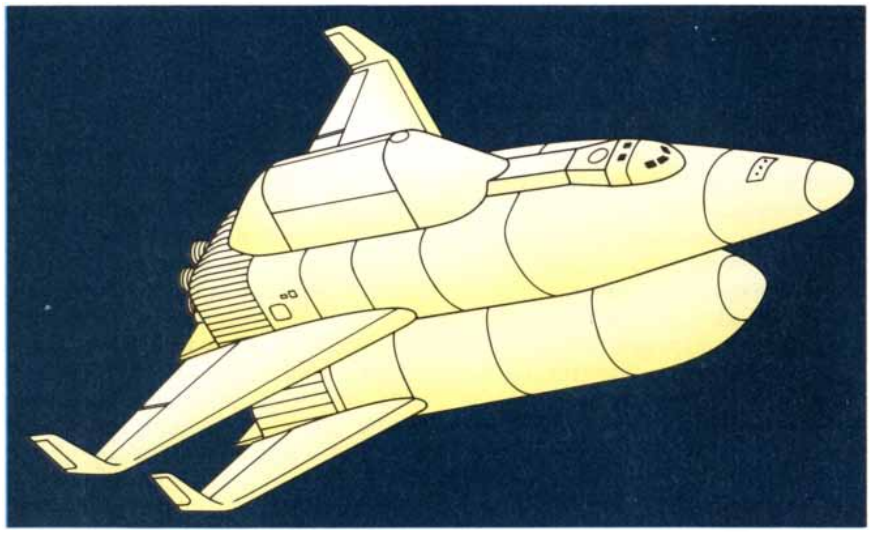
all three have scheduled their first commercial launch for later this year.

The main competitor to U.S. companies for launch contracts is now Arianespace, a partnership among various European government, industrial and financial institutions to market the family of Ariane ELV's developed under the auspices of the European Space Agency. At the time of the *Challenger* accident commercial launches were nearly evenly divided between Ariane and the shuttle, but a number of new customers signed up with Arianespace as the U.S. struggled to develop an acceptable policy framework for its commercial launch industry. Yet Arianespace has had its own set of difficulties: a failure during a launch in May, 1987, grounded the Ariane launchers for more than a year.

Companies that want to put satellites in orbit now find themselves in a buyer's market for launch services. The three U.S. companies and Arianespace combined can launch between 30 and 40 spacecraft per year, but the foreseeable annual demand is for only between 15 and 25 launches. Such a situation makes for lively competition, and the launch-services market is likely to become even more competitive in the near future. Both the Soviet Union and the People's Republic of China are attempting to enter the market by offering to potential customers respectively the Proton and Long March launchers. Japan is also developing a launch vehicle, known as the H-2, that could offer significant competition by the mid-1990's.

Yet the primary payloads for commercial launch services are communications satellites, almost all of which employ technologies that fall under restrictions limiting their export to the Soviet Union. As a result the Proton is not likely to be a significant competitor to Western ELV's unless improved Soviet-American relations lead to a softening of technology-export restrictions. Such softened restrictions already apply to China.

Actually in China's case the relaxation of export restrictions has provoked less complaint from U.S. launch-vehicle companies and Arianespace than the fact that the Chinese are offering to launch satellites at what they admit are artificially low introductory prices. In the interests of both national security and fair trade, the U.S. recently struck a deal with China. The U.S. Government agreed to license several satellites built by the Hughes Aircraft Company, an American enterprise, for launch on Long March vehicles on condition that the Chinese alter their pricing policy after a few discount-price launches and adopt launch-preparation practices that reduce the chance of their surreptitiously gaining technical knowledge from Western payloads.



ADVANCED MANNED LAUNCH SYSTEM being studied by NASA is meant to transport passengers to and from orbit. The vehicle would be completely reusable and would consist of a piloted orbiter and an unmanned booster that lands on a runway.

Although it appears that there is enough business to support several commercial launch firms for the next few years, the long-term outlook for the industry is worrisome. Communications satellites already in orbit are operating below capacity, and fiber-optic data-transmission systems are proving effective competitors to satellites for high-volume point-to-point communications. The demand for launching satellites into geosynchronous orbit will grow only if satellites are utilized for such new functions as mobile communications, direct television broadcasting and navigational position fixing.

In sum, Government contracts will keep U.S. ELV production lines open for the foreseeable future and thereby ensure the availability of a variety of space launchers, but it does not appear that there will be a large enough market to sustain indefinitely all current sellers of launch services. For this reason some people have suggested that U.S. companies should consolidate their launch activities and form a single quasi-commercial entity like Arianespace in order to compete effectively in the future. Inasmuch as success in the global launch market would be an important symbol of economic payoffs from space and would help the nation's overall balance of trade, the Government has made commercializing space launches an important policy objective.

Three main technical objectives dominate current U.S. thinking about unmanned launch systems for the late 1990's and beyond: reducing costs, improving readiness and reliability and creating the capability to launch payloads heavier than the space shuttle can accommodate. Those objectives have been important for the future of the U.S. space program ever since the shuttle's high operating costs and performance limitations became apparent. But there is another reason for their significance: any major space initiative, such as deploying a sizable space-based strategic defense system, establishing a lunar base or sending human beings to Mars, would require more capable, reliable and affordable launch systems.

In 1985 a joint Defense Department-NASA study was initiated to examine how to meet the launch requirements of possible future U.S. space activities, civilian as well as military. Within a year this so-called Space Transportation Architecture Study uncovered a problem that was a product of the U.S.'s single-minded pursuit of the space shuttle: the existing technology base for space transportation was simply too thin to make an intelligent recommendation about what technologies should be employed in future launch systems. Over the past two years both NASA and the Defense Department have in response substantially increased funding for research

and development in areas such as liquid- and solid-fuel propulsion, automation, robotics and lightweight materials. Nevertheless, significant levels of continued investment will be needed to build an adequate technology base for making choices about future space-transportation systems.

The main Government framework for such investment is the Advanced Launch System (ALS) program, which is being funded by the Defense Department and managed jointly by the Air Force and NASA. The fact that an Air Force officer is in charge of the program reflects a new reality: future unmanned launch systems will be developed under Defense Department direction primarily to meet department requirements.

The ALS program has a legislative mandate to create the capability to launch payloads at a cost of $300 per pound or less into low earth orbit (current costs range between $3,000 and $5,000 per pound) while also providing space vehicles that have a wide range of payload capacity, are highly reliable and can quickly resume operation after a failure. As currently constituted, the ALS program has as its primary focus developing generic space-transportation technology and a new family of U.S. launch systems embodying that technology, rather than developing a specific vehicle.

Strong emphasis is being given to managing the program as a complete transportation system, since it is unlikely that drastic reductions in launch costs will ever be achieved without substantial changes in the way launch operations are conducted—on the ground as in space. Indeed, operations costs may account for as much as 45 percent of the total tab for a shuttle launch. The OTA recently reported that such costs can be significantly reduced by applying new technologies to ground-based space facilities and by streamlining the documentation and oversight processes involved in launch preparations.

By the early 21st century not only the U.S. and the U.S.S.R. but also Europe, Japan and perhaps China expect to have the capability to fly human beings to and from space. Yet only the U.S. will be using for that purpose a vehicle initially designed with the technology of the late 1960's in mind, namely the space shuttle. NASA is beginning to consider possible evolutionary improvements to the basic shuttle design that could be incorporated either into the existing shuttle-orbiter fleet or into new orbiters meant to replace the aging original ones. (Such replacement shuttles are not yet part of NASA's official plans, but some observers believe they will be required.) Still, there are limits to the improvements that can be made to old designs. Hence sometime after the turn of the century the U.S. will have to replace the shuttle with a new vehicle for putting people into space, at least if it wants to remain a leader in this most visible aspect of space activity.

To that end, NASA has emphasized the development of what the agency calls the Advanced Manned Launch System (AMLS). As currently conceived, the AMLS would have a limited cargo-carrying capacity (20,000 pounds or so) and orbital-stay time (between two and five days). It is intended mainly as a passenger vehicle, carrying six or more people to and from the space station or low-earth-orbit satellites that need servicing; it would not launch large payloads. NASA believes it could develop by 2005 a fully reusable system that includes not only a piloted shuttlelike orbiter but also an unmanned winged booster capable of gliding back to the earth for a landing on a conventional runway.

The U.S. is examining other alternatives for bringing humans into space. One proposed vehicle, the National Aerospace Plane, would make use of hydrogen-burning engines that are capable of operating both in the atmosphere and in the vacuum of space to fly directly into orbit after lifting off from a runway. The development of this piloted launch system is, like the ALS, being managed jointly by the Defense Department and NASA with the department in charge. The department provides more than 80 percent of the budget for this highly classified program and sets priorities for its capabilities.

The notion of a vehicle that could take off from a conventional airfield and either ascend directly into orbit at 25 times the speed of sound or fly at hypersonic speeds of 3,500 miles per hour from the U.S. to Japan was briefly aired in public by President Reagan as the "Orient Express" in his 1986 State of the Union address. Since then the aerospace plane has been kept under tight wraps. The X-30 aircraft, a test-bed craft for various engines and material technologies that could make the concept feasible, is scheduled to fly in the mid-1990's, but little information is available on plans for incorporating these technologies into piloted space vehicles for civilian or military applications.

If the ALS, AMLS or National Aerospace Plane program is successful, new launch vehicles could be available beginning at the turn of the century. Yet any decision about which systems to develop as the next generation of U.S. space launchers should be made only in the context of decisions on future U.S. goals in space. If the U.S. intends to pursue vigorous programs in space science and space exploration, enhance the effectiveness of future national-security space systems and open up space at an affordable price to those who want to investigate its commercial promise, future space-transportation systems must operate routinely and at low cost. That was the promise of the space shuttle in 1972. Perhaps by early in the 21st century, nearly 30 years later, that promise will finally be realized.

As the U.S. learned in the months following the recent spate of launch failures, there can be no space program if there is no access to space. On the other hand, the development of a space-transportation system should not be mistaken for a national space program. A space program pursues established goals, and a space-transportation system is merely the means by which to achieve them. One can hope that in the coming decade the U.S. will not succumb once again to its propensity for substituting a decision about a means—whether it is a launch system such as the shuttle or an orbital facility such as the space station—for a decision about the purposes to be served by U.S. activities in space.

FURTHER READING

SETTING SPACE TRANSPORTATION POLICY FOR THE 1990's. Congressional Budget Office. U.S. Government Printing Office, October, 1986.

THE PRIVATE SOLUTION TO THE SPACE TRANSPORTATION CRISIS. James Bennet and Philip Sadin in *Space Policy,* Vol. 3, No. 3, pages 181–205; August, 1987.

LAUNCH OPTIONS FOR THE FUTURE: A BUYER'S GUIDE. Office of Technology Assessment. U.S. Government Printing Office, July, 1988.

REDUCING LAUNCH OPERATIONS COSTS: NEW TECHNOLOGIES AND PRACTICES. Office of Technology Assessment. U.S. Government Printing Office, September, 1988.

THE SPACE SHUTTLE: THE JOURNEY CONTINUES. Richard Truly. NP-117, National Aeronautics and Space Administration, September, 1988.

U.S. SPACE TRANSPORTATION POLICY: HISTORY AND ISSUES FOR A NEW ADMINISTRATION. Scott Pace in *Space Policy,* Vol. 4, No. 4, pages 307–318; November, 1988.

INTRODUCING THE NEW DODGE SPIRIT.

THE ALL-NEW DODGE SPIRIT ES. *Dodge Spirit ES is a true driver's sedan. You can tell by its aggressive lines. Monochromatic look. And by how quickly it moves when you put your foot down. A new 2.5L fuel-injected turbo provides the power. Or choose 3.0 liters of fuel-injected V-6 performance matched to a new electronic Ultradrive automatic. You also get responsive power steering. A sport suspension. And our exclusive 7 year or 70,000 mile Protection Plan.* Best of all, Spirit prices start at just $9,995, while the premium ES (shown) is only $12,495.† So see your Dodge dealer. Then let your spirit soar.* 7/70

Dodge **THE NEW SPIRIT OF DODGE**
THE PERFORMANCE DIVISION OF CHRYSLER MOTORS

*See this powertrain limited warranty & its restrictions at dealer.
†Sticker prices exclude tax, destination charge & options.

BUCKLE UP FOR SAFETY.

Multidrug Resistance in Cancer

An ancient pump protein that flushes toxins out of cells may be to blame when cancer chemotherapy fails. Its identification offers hope that multidrug-resistant cancers might be made vulnerable again

by Norbert Kartner and Victor Ling

Ms. Smith had complained of cramping abdominal pain to her family doctor. After a thorough examination she was referred to a local cancer clinic, where she was diagnosed as having an abdominal tumor and was immediately booked for surgery. Removal of the primary tumor was successful but, as is often the case, the cancer had already spread widely to other tissues. The patient underwent chemotherapy with a combination of anticancer drugs, the method of choice for delocalized tumors that are untreatable by surgery or radiation, and the response was miraculous. By all available diagnostic methods she was free of disease. Three months later, during a routine follow-up, Ms. Smith was found to have relapsed: tumors had emerged in several organs. She underwent a second course of chemotherapy but responded poorly. A month later a third course of chemotherapy, with different drugs, had no effect on the growing tumors. Three weeks after that Ms. Smith died.

Why did chemotherapy eventually fail Ms. Smith when at first it had appeared to be highly successful? Why are some cancers curable by chemotherapy alone, whereas others are unaffected by drugs and are apparently incurable? These are not new questions. Indeed, the resistance of parasites and infectious-disease organisms to antibiotics is as old as chemotherapy itself. The German chemist Paul Ehrlich, the father of chemotherapy, had envisioned "magic bullets," drugs that would cure many of the diseases that plague mankind, but after decades of experience with antimicrobial drugs he lamented that drug resistance had followed the development of new drugs "like a faithful shadow."

Because cancer chemotherapy has its roots in the antimicrobial chemotherapy that has been under development since the turn of the century, clinical resistance to anticancer drugs was not entirely unexpected. A decade before cancer chemotherapy came to the clinic soon after World War II, experiments with tumors transplanted into mice had already demonstrated the development of progressive resistance to experimental drugs. Since then experimental tumors resistant to every class of anticancer drugs have been isolated. All organisms, including the cells within a cancer patient's tumor, seem to have the capacity to become resistant to drugs that would otherwise kill them.

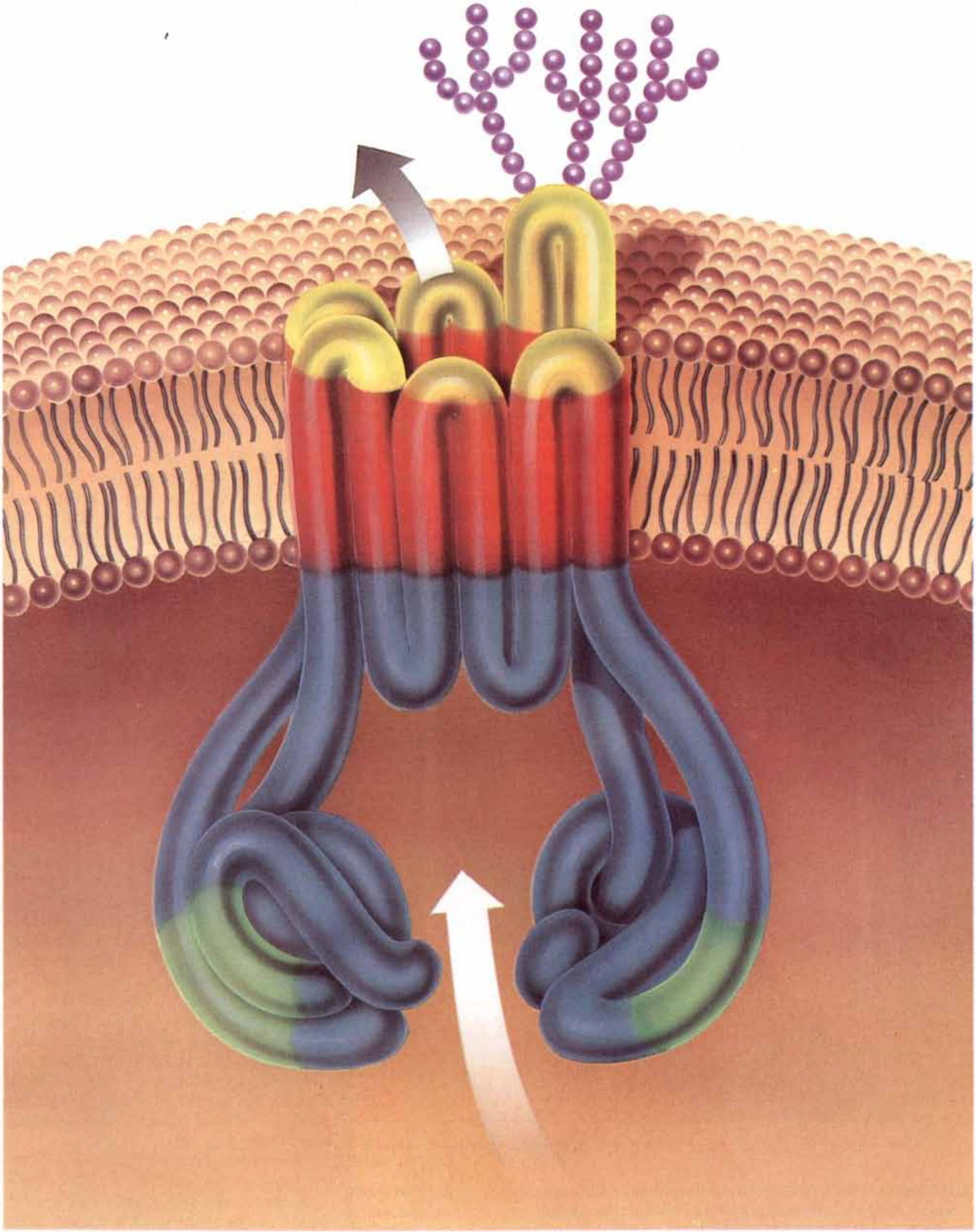
The underlying cause of progressive drug resistance, whether in infectious diseases or in cancer, is the same. Spontaneous genetic mutations occur in all living cells, giving rise to heritable traits that are passed on to succeeding generations. In any cell population, mutants that are resistant to a given drug occur at a frequency of somewhere between one in 10^5 and one in 10^8 cells. How can such a rare event have an impact on the outcome of chemotherapy? To answer this question, one needs first to consider the microscopic scale of the cell and the limitations of methods for the early detection of cancer.

A tumor of average detectable size, say one centimeter in diameter, already contains hundreds of millions of cells, some of which are likely to be drug-resistant. Hence, in spite of the rarity of mutations that produce drug resistance, tumors containing some drug-resistant cells at the time of diagnosis seem likely to be the norm. The outcome of treating such a tumor with a single drug can be predicted. At first a remission will be seen, in which the tumor shrinks to an undetectable size as a result of the killing of the predominant drug-sensitive cells. Yet the remaining drug-resistant cells, all of whose progeny are also drug-resistant, continue to multiply; they eventually dominate the cell population of the tumor, which grows to a size that results in the death of the patient. It has been confirmed that even a single drug-resistant cell introduced into an otherwise curable tumor transplanted into a mouse will eventually multiply in the course of chemotherapy and dominate the tumor-cell population, resulting in an incurable and ultimately fatal disease.

The solution to the problem would seem to be rather simple. In theory, treatment at the outset with a combination of drugs that act differently should preclude the outgrowth of a drug-resistant tumor, because the probability that two or more different drug resistances would arise spontaneously in the same cell is very small. Combination chemotherapy seemed to be the answer.

Recognizing the need for an arsenal of drugs and the futility of the single magic bullet for the treatment of advanced infectious disease was arguably one of the conceptual landmarks of pharmaceutical research in the first half of this century. Most of the tenets of antimicrobial chemotherapy were adopted wholeheartedly by cancer clinicians, who demanded research into the production of novel drugs and developed protocols for administering drugs in combination. Newly developed drugs and combination chem-

NORBERT KARTNER and VICTOR LING worked together on P-glycoprotein and multidrug resistance at the University of Toronto. Kartner recently earned his Ph.D. from the department of medical biophysics there. Ling is professor in that department and senior scientist at the Ontario Cancer Institute. He got his Ph.D. in biochemistry from the University of British Columbia in 1969 and did postdoctoral work with Frederick Sanger in Cambridge. He recently received the C. Chester Stock Award and, for his work on the genetics of drug resistance in mammalian cells, a Milken Cancer Research Award.



P-GLYCOPROTEIN resides in the cell membrane, where it may act to pump toxins out of the cell. The painting shows a model of the protein's structure that is based on its known sequence of amino acids. The protein chain is thought to snake back and forth 12 times across the lipid bilayer of the membrane, forming a 12-sided pore. The part of the protein outside the cell bears sugar chains (*purple*); two large and nearly identical domains protrude into the cell. They include regions (*green*) that bind the cellular energy-carrying compound ATP, which probably provides the energy that drives the efflux (*arrows*).

otherapy produced real victories a couple of decades ago in the high cure rates for some childhood leukemias and for Hodgkin's disease, but the big killers—lung cancer, breast cancer and cancers of the gastrointestinal tract—remained refractory to chemotherapy.

The perplexing failures, particularly those of combination chemotherapy, seemed to defy understanding. Many theories were proposed to explain the observations, but few could be adequately tested. Early in the development of experimental chemotherapy in mice it was recognized that simultaneous resistance to a number of drugs was an unexpectedly common occurrence. Yet research focused on the more easily understood phenomenon of resistance to single agents. It was not until the late 1960's, when investigators began to do experiments with drug-resistant tumor cells in vitro, that the issue of multiple drug resistance resurfaced and the first insights into what is now known as the multidrug-resistance phenotype were gained.

The observations defined the fundamental properties of multidrug resistance. Although drug-resistant mutants were selected by means of a single anticancer drug, they were often simultaneously resistant—that is, cross-resistant—to completely unrelated drugs. Most important was an observation arising from a number of independent genetic experiments: multidrug resistance appeared to result from a single mutation. In other words, a single gene could account for the multiple cross-resistance to unrelated drugs.

That concept had three important consequences. It spurred research to find the multidrug-resistance gene in experimental tumors, it stimulated inquiry into the gene's effect and it provided a rational explanation for failures of combination chemotherapy. Since a single drug-resistance mutation is a rare event, the acquisition of multiple mutations in the same cell, yielding resistance to unrelated drugs, is an occurrence that would be highly improbable. The multidrug-resistance phenotype that resulted from a single mutation explained how resistance to combination chemotherapy could be a common occurrence. But how could a single gene have such a broad effect?

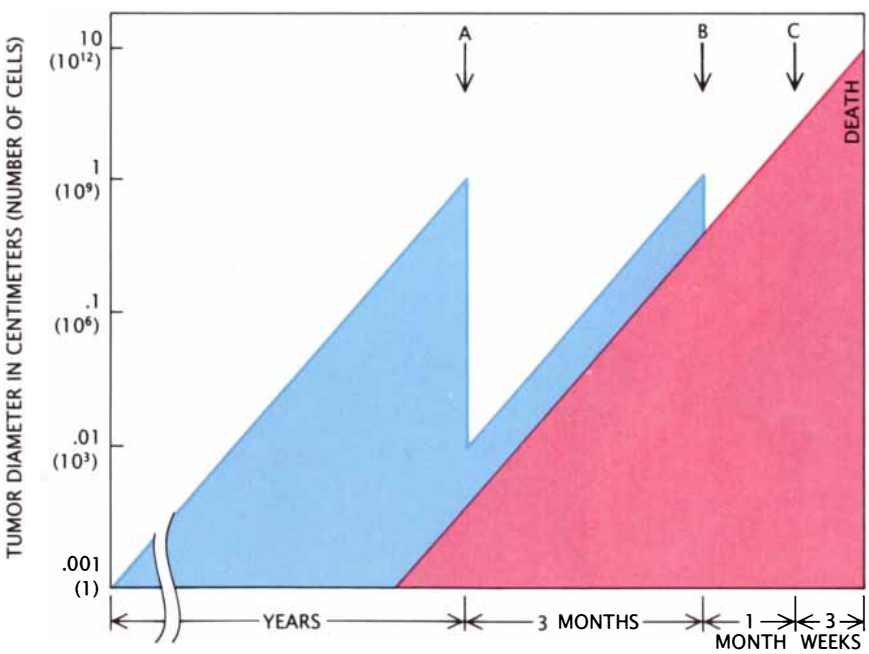
Working with various systems, investigators had found that cells that were resistant to a drug somehow excluded it. This observation suggested a mechanism for the drug resistance: there appeared to be some barrier that kept the drug from reaching the interior of the cell, where it would have its lethal effect. Two possible theories were put forward to account for the evidence.

One theory proposed that a permeability barrier prevented drug entry into the cells. The other suggested that an efflux pump, a mechanism that actively pumped drug out of the cell once it had got inside, was at work in the resistant cells. The latter model was based on observations of the kinetics of drug flow into and out of the cells. It was found that when a resistant cell was temporarily poisoned with cyanide to inhibit energy production, the cell behaved like a drug-sensitive one: it could not keep out the drug. When the cyanide was washed out and normal metabolism was restored, the cell could once again exclude the drug. Furthermore, the cell was then able to pump out the drug that had accumulated while it was poisoned. Hence an energy-dependent drug-efflux pump seemed to be the simplest explanation.

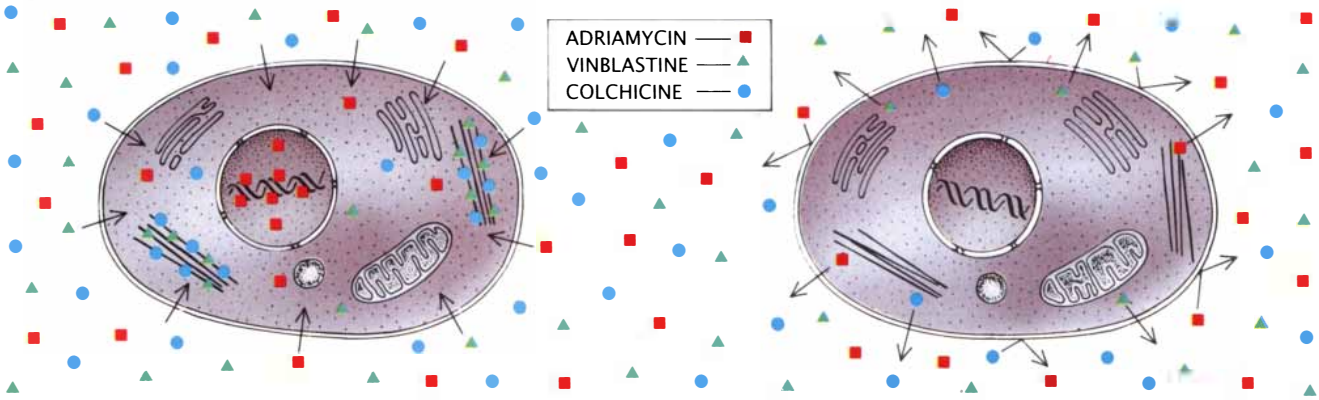
Whatever the actual mechanism, two points seemed clear. One was that the process of keeping drugs out of the cell would need to be rather nonspecific, that is, able to cope with drugs of diverse molecular structure. The other was that because the cell's surface membrane (the plasma membrane) is the first line of defense against the entry of drugs, the difference between drug-sensitive and drug-resistant cells would probably be found there.

Perhaps the first direct evidence of a specific alteration of the plasma membrane in multidrug resistance came from our studies of Chinese hamster cells that were resistant to the drug colchicine, done in collaboration with Rudolph L. Juliano and later with John R. Riordan, both of the Hospital for Sick Children in Toronto. We separated components of the plasma membranes of Chinese hamster cells by gel electrophoresis, a technique in which molecules are drawn through a gel by an electric field and are thereby sorted according to size.

The process revealed that there was a unique glycoprotein in the drug-resistant cells that seemed to be absent in the drug-sensitive ones. Glycoproteins are complex molecules, made up of protein and carbohydrate, that are usually associated with the plasma membrane. This glycoprotein was rather large in size (its molecular weight was roughly 170,000), and it



FAILURE OF CHEMOTHERAPY often follows initial success. In this simplified model of cancer progression the first course of chemotherapy (*A*) appears to be successful: it reduces the tumor to an undetectable size by killing most of the drug-sensitive cells (*blue*). It has no effect, however, on a tiny population of drug-resistant mutant cells (*red*), which grows exponentially as the drug-sensitive population recovers. Detection of the recurrence leads to a second course of chemotherapy (*B*). This course again reduces the drug-sensitive population, but it has no effect on the fraction—by now sizable—of drug-resistant cells. A final course of chemotherapy (*C*) has no apparent effect, and soon unchecked tumor growth kills the patient.



MULTIDRUG RESISTANCE enables a cell to withstand the effects of toxic molecules that vary in size, structure and site of action in the cell. A common anticancer drug, adriamycin, acts in the nucleus of a drug-sensitive cell (*left*), interfering with the transcription of DNA and its synthesis during cell division. Two other toxic compounds used in chemotherapy and studies of drug-resistant cells, vinblastine and colchicine, affect microtubules, which play an important role in cell division. Either a passive barrier or an active pump in the cell membrane could explain simultaneous resistance to such diverse agents (*right*).

turned out to be associated specifically with the plasma membrane. We named it P-glycoprotein for its association with the apparent permeability barrier to drugs that accompanied multidrug resistance.

Not long after the first report of P-glycoprotein in Chinese hamster cells, similar findings were reported by other groups. Each group was working with a different tissue-culture system. A variety of mouse, hamster and human cells were selected for resistance to any one of a variety of drugs: adriamycin, colchicine, daunomycin, vinblastine, vincristine and so on. All these systems showed extensive cross-resistance to unrelated drugs, reduced intracellular accumulation of the drugs involved and alterations in the cell's surface membrane. The most consistent of the observed alterations was the appearance of a high-molecular-weight cell-surface glycoprotein similar in size to the P-glycoprotein. These findings prompted us to ask whether the various phenomena were in some way related.

To answer that question more specific tools were required, and so we decided on an immunochemical approach. The technique involves creating traceable antibodies that adhere to a specific molecule, such as P-glycoprotein, so that it can be isolated and studied. To develop the highly specific antibodies, called monoclonal antibodies, we first injected purified plasma membranes from multidrug-resistant cells into mice. Then, by fusing spleen cells from those immunized mice with immortal tumor cells, we produced clones of identical antibody-secreting hybrid cells. We isolated the clones that secreted monoclonal antibodies to P-glycoprotein.

We then sought to determine whether an increased amount of P-glycoprotein was correlated with a high degree of cross-resistance. To do so we employed a technique known as immunoblotting, in which the antibodies served to identify P-glycoprotein in a complex mixture of proteins and glycoproteins that had been separated by gel electrophoresis. Not surprisingly, the immunoblots showed that there was very little P-glycoprotein in the drug-sensitive Chinese hamster cells; progressively more P-glycoprotein appeared in cell lines that were found to be increasingly resistant to the drug colchicine.

We had foreseen that result. When we applied the same method to a variety of cell lines supplied by other groups, however, we were astonished to see that Chinese hamster, Syrian hamster, mouse and human cell lines selected for resistance to a variety of drugs all showed components like P-glycoprotein in their plasma membranes. These components not only were indistinguishable in size but also reacted with the same antibodies that were highly specific for P-glycoprotein in Chinese hamster cells.

It had become clear that P-glycoprotein was a conserved molecule: a molecule that has retained its structural identity across different mammalian species. Moreover, regardless of the species of origin or the drug of selection, the drug-resistant cells all exhibited a large elevation in P-glycoprotein expression in concert with the development of drug resistance. Conservation of structure in biological molecules is usually indicative of an important functional role; this premise, and the concept of the universality of P-glycoprotein expression in concert with multidrug resistance, was of key importance in establishing the direction we would take later.

Clearly P-glycoprotein seemed to play an important role in multidrug resistance. We therefore turned to tools of molecular biology that could give us a very close look at the structure and ultimately the function of the molecule itself. The most powerful way to determine the predicted structure of a protein is to find the genetic sequence that encodes it. The appropriate tool is complementary DNA, or cDNA, which corresponds to the DNA that encodes the protein whose properties are being studied.

Our cDNA was developed in collaboration with Riordan, who had established a cDNA library from highly colchicine-resistant Chinese hamster cells. A cDNA library is a living repository for genetic material. This one consisted of a mixture of bacteria, each infected by a virus called a bacteriophage. These particular bacteriophages were recombinant, that is, their genetic material contained an inserted fragment of foreign DNA. The foreign DNA was cDNA derived from messenger RNA that was being actively translated into various proteins by the drug-resistant cells. (Messenger RNA, or mRNA, is the intermediary in the transfer of information from DNA to proteins.) The cDNA fragments were inserted into a viral gene coding for the enzyme beta-galactosidase. When the viral gene infiltrated the bacteria's genetic material, the bacteria expressed an altered beta-galactosidase that contained an additional protein fragment, one encoded by the cDNA.

To find the cDNA that corresponded to P-glycoprotein in dishes containing tens of thousands of bacterial colonies, each one producing a different protein fragment, would seem equivalent to finding a book in the stacks of the University of Toronto library by looking at each volume. But with a radioactively labeled monoclonal antibody acting as a "magnet" specific for P-glycoprotein, we quickly identified the right cells. Growing the selected colony, in which all bacteria originated from a single cell that carried a single cDNA fragment, resulted in the isolation of a cloned fragment of P-glycoprotein cDNA. This cloned fragment could then serve as a probe with which to perform blot hybridization, one of the powerful analytical methods of modern molecular biology.

Blot hybridization, in which a cDNA probe serves to pick out corresponding sequences in DNA or RNA that has been separated by electrophoresis, provided insight into both the nature of the P-glycoprotein molecule and how it functions. Two types of blot hybridization were done. In the first one, known as Northern blotting, the isolated cDNA was used to probe mRNA derived from different cell lines. It became clear that an mRNA species about 4.5 kilobases long (an mRNA including about 4,500 of the chemical subunits known as bases) was associated with various multidrug-resistant cells and not with their drug-sensitive counterparts; the more resistant the cells were, the larger the amount of the specific mRNA species was. It was reasonable to suppose this RNA was associated with the production of P-glycoprotein.

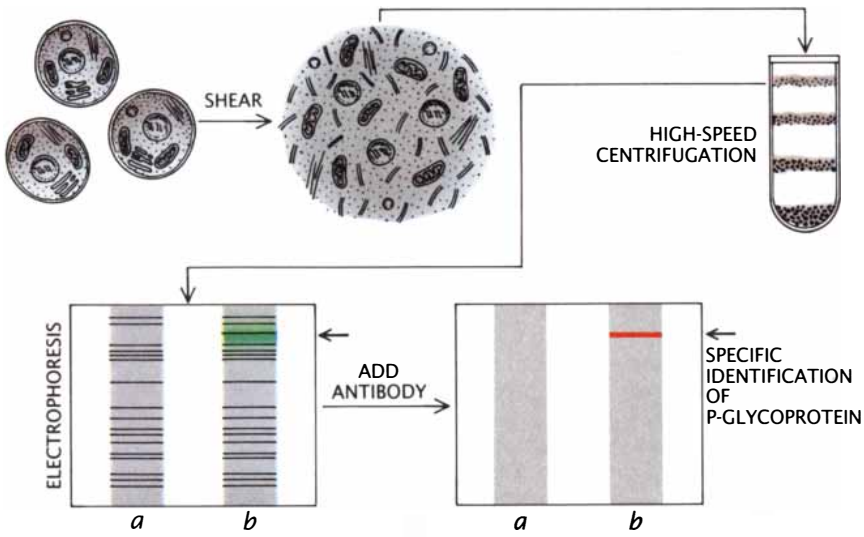
A second method, known as Southern blotting, employed the cDNA to probe genomic DNA, the DNA of the cell nucleus. From the results of this work it became apparent that the increased amount of P-glycoprotein seen in multidrug-resistant cells arose through a process of gene amplification: an increase in the number of copies of a gene. As many as 60 copies of the same P-glycoprotein gene could be seen in resistant cells. This observation confirmed independent evidence that multidrug resistance resulted from gene amplification. In our system Southern blots revealed multiple bands in both drug-sensitive and drug-resistant cells, instead of the one or two bands expected for a single gene. The simplest explanation was that there was more than one multidrug-resistance gene in the normal genetic makeup of the cell—not identical genes, but very closely related ones, constituting what is known as a multigene family.

While that work was in progress, other groups were looking into multidrug resistance. Different approaches were being taken by Igor B. Roninson of the Massachusetts Institute of Technology and Piet Borst of the Netherlands Cancer Institute in Amsterdam. They cloned DNA fragments associated with multidrug resistance by independent means. Regardless of the methodology, their work confirmed our results: gene amplification and the overexpression of a 4.5-kilobase mRNA were seen in each case. The size of the mRNA yielded a clue to the protein it encoded; it was the length expected of an mRNA coding for a protein the size of P-glycoprotein.

Monoclonal antibodies enabled us to identify our cDNA probe with P-glycoprotein beyond any doubt. The partial gene product made by the bacteria fitted three independent monoclonal antibodies that recognized different sites on P-glycoprotein. An exchange of partial DNA sequence information made it clear that all three groups had independently, and by quite different approaches and rationales, cloned P-glycoprotein genes. A functional, causative role for P-glycoprotein in the multidrug-resistance phenotype seemed assured.

From a rigorous scientific point of view such circumstantial evidence was not entirely satisfying; direct proof was needed, and it was sought. This direct proof came from the laboratory of Philippe Gros at McGill University. Gros took a piece of cDNA that contained the full length of the coding region for P-glycoprotein from a drug-sensitive mouse cell and inserted it into a normal, drug-sensitive hamster cell by a process known as gene transfection. When the progeny of the transfected cell grew in the presence of a selecting drug, Gros could conclude that they were drug-resistant. He isolated DNA and mRNA from the cells and probed them by means of blot hybridization. He found that the hamster cells contained multiple copies of the mouse P-glycoprotein gene and were expressing the gene. Since no other changes had occurred in what was originally a drug-sensitive cell, it appeared that the elevated levels of P-glycoprotein expression alone could account for the drug resistance.

When the hamster cells were tested for resistance to unrelated drugs, it was found that they displayed the same extensive cross-resistance seen in spontaneously arising multidrug-resistant cells. Since only a single gene, a single member of the P-glycoprotein multigene family, had been inserted, it became clear that a single type of P-glycoprotein molecule could account for the extensive cross-resistance to



CELL-SURFACE COMPOSITION distinguishes drug-sensitive and drug-resistant cells. The surface proteins are extracted by shearing the cells and placing their components in a sucrose solution whose density increases from top to bottom. When the solution is spun in a centrifuge, components of the cell membrane, which have a low density, form a band near the top. Gel electrophoresis then separates the membrane proteins by size. Multidrug-resistant cells (*b*) yield a protein (*green band*) that is apparently absent in drug-sensitive cells (*a*). The development of a specific antibody that labels the protein—P-glycoprotein—was crucial in the authors' work.

unrelated drugs that is characteristic of multidrug resistance.

How could a single molecular species such as P-glycoprotein accomplish such an apparently complicated task? The first steps toward understanding the details of how P-glycoprotein works were taken by deducing the full primary amino acid sequence of the protein. Sequencing DNA has become a routine practice, and so the base sequence of the cDNA that represents the full-length mRNA coding for P-glycoprotein was soon determined. Once the coding sequence was known it could be translated into an amino acid sequence. That is possible because the genetic code is known and every three-base codon of the DNA sequence codes for a specific amino acid. The so-called primary sequence turned out to be about 1,280 amino acids long. With the primary sequence available, one could search the sequence for certain structural features known to have specific functions. Reliance on computers and an extensive protein-sequence data base makes this a relatively simple task.

Certain features of P-glycoprotein were immediately apparent. Short sequences were found that constitute sites where sugar molecules are added to produce a glycoprotein. Another crucial clue emerged from the fact that different amino acids have different affinities for lipid or water. A so-called hydropathy plot can identify regions of a primary amino acid sequence that would be associated with the lipid bilayer of the plasma membrane. When these regions have a continuous length of about 21 amino acids, they are said to be transmembrane regions: they can span the membrane from the inside to the outside of the cell or vice versa. A number of such transmembrane regions were identified in the P-glycoprotein sequence.

We began to understand the significance of these structural features by comparing their sequences with those of other regions within the same protein molecule and with sequences of other known proteins. In addition to showing that the P-glycoprotein sequences characterized by the different laboratories were very similar, these comparisons revealed some intriguing features of the P-glycoprotein structure. The P-glycoprotein molecule is internally duplicated: the first half of the protein sequence is very similar to the second half. This suggests that a simple ancestral gene was duplicated to give a tandem repeat yielding the more complex protein seen today.

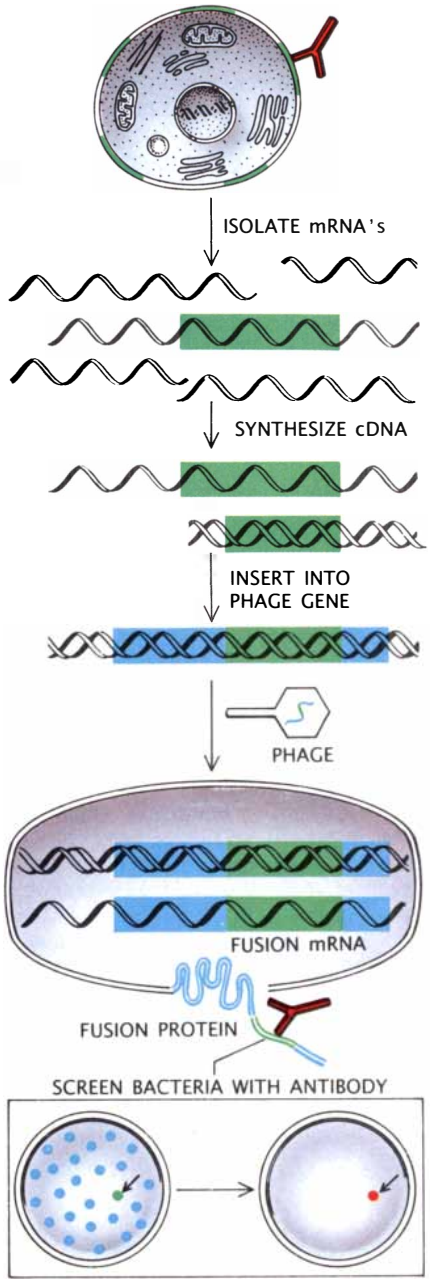
Each half-sequence has six putative transmembrane regions, meaning that the P-glycoprotein molecule can snake back and forth across the membrane 12 times. Such a complex transmembrane configuration is characteristic of channel-forming, or pore-forming, proteins involved in the transport of nutrients, ions and cellular metabolites across the cell's surface membrane, either into or out of the cell.

Comparison with a protein data base confirmed the suspicion that P-glycoprotein resembles a membrane transport protein. Regions of similarity with known transport proteins in organisms ranging from bacteria to insects were found. The primary region that showed a high degree of conservation among such widely divergent species turned out to be an ATP-binding region. ATP, or adenosine triphosphate, is a molecule that provides energy for biochemical activity. Both halves of P-glycoprotein contain a lengthy hydrophilic region, a region that is more likely to be in contact with an environment of water than one of lipid. This region was known to be on the inner side of the surface membrane. It was there that the ATP-binding site was situated.

The most surprising discovery on searching through the existing protein data base was that the homologous halves of P-glycoprotein greatly resembled a previously described protein known as hemolysin B. Hemolysin B is present in the surface membrane of certain bacteria and is responsible for transporting a protein called alpha-hemolysin out of the cells.

These amino acid sequence studies and comparisons with other proteins have led to a proposed model for P-glycoprotein structure, one that suggests possible ways the protein might provide multidrug resistance. It is likely that the 12 transmembrane regions of P-glycoprotein converge to form a 12-sided pore. On the outside of the cell there is little exposure of protein; this is the site where the sugar chains that make it a glycoprotein are attached. On the inside of the cell there are two large, homologous domains projecting into the cytoplasm, which bear the ATP-binding sites. The sites that accept ATP on the P-glycoprotein molecule suggest that the protein has an energy-transducing function—such as the energy-dependent extrusion of toxic drugs from the cell.

It is likely that the P-glycoprotein pumps drugs out of the cell in one of two ways. Either it binds a variety



CLONING of P-glycoprotein DNA begins with the identification, with a monoclonal antibody (*red*), of a cell producing P-glycoprotein (*green*) and the isolation of messenger RNA's (mRNA's) coding for the cell's proteins. Double-strand complementary DNA (cDNA) is then synthesized for a portion of each mRNA; each cDNA is inserted into a gene (*blue*) of phage lambda, a virus that infects bacteria. An infected bacterium transcribes the resulting "fusion gene" into mRNA and translates the mRNA into a fusion protein that includes part of a protein from the original cell. Each bacterium carrying a fusion gene multiplies into a clone of genetically identical bacteria expressing the same fusion gene. The antibody identifies the clone bearing the fusion protein incorporating a portion of P-glycoprotein. That clone can then serve as a source of P-glycoprotein DNA.

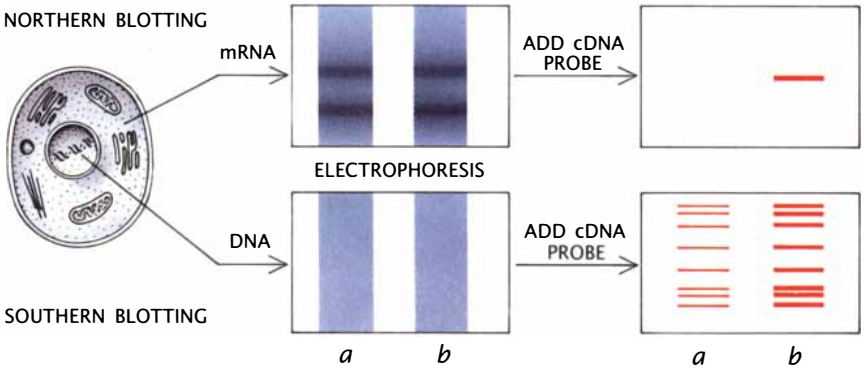
of drugs and extrudes them directly through the membrane by way of its putative transmembrane pore, or a second molecule (a carrier protein) binds to the drug and the drug-carrier complex is extruded across the membrane. The latter possibility is based on the observation that P-glycoprotein resembles hemolysin B, which extrudes alpha-hemolysin across the bacterial cell membrane. No direct evidence of an ancillary carrier protein for P-glycoprotein has yet been found. There is evidence that some drugs may bind directly to P-glycoprotein, possibly as a first step in their ultimate transport across the surface membrane.

Further sequencing of various P-glycoprotein genes in various species by the increasing number of laboratories working in the field has allowed some comparisons to be made between genes, both within a species and among different species. Such comparisons have shed some light on the evolution of P-glycoprotein and the organization of its genes. The similar organization of coding sequences and intervening sequences in different P-glycoprotein genes from the same species suggests that the internal duplication of the ancestral gene occurred prior to the formation of a multigene family. Similarities in the organization of homologous members of the multigene family in different mammalian species suggest that the formation of a multigene family preceded the divergence of species, at the dawn of the evolution of mammals.

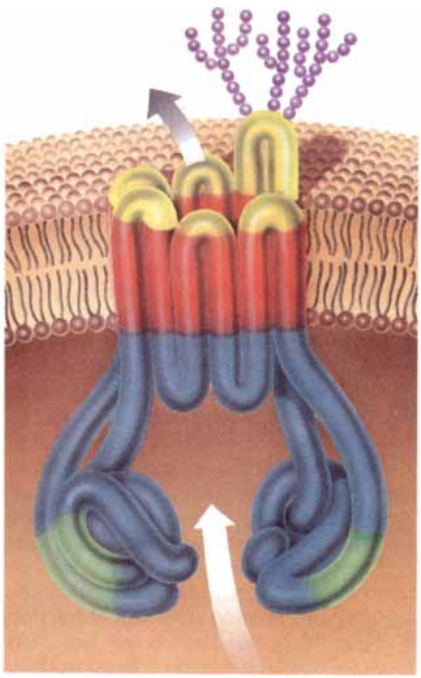
The apparently long evolutionary history and conserved structure of P-glycoprotein beg two questions: What is the molecule's normal function and what are the details of its activity? As yet the answers are merely speculative, but two theories have been proposed. One is that P-glycoprotein does the same job in normal cells that it does in drug-resistant cells: it removes toxins from within the cell. A survival strategy that goes far back into evolutionary history is the secretion by an organism of toxic compounds to which it is immune in order to kill nearby competitive organisms. Some of the anticancer drugs in use today—and many of the antibiotics—are, in fact, toxins produced by lower organisms for precisely that purpose.

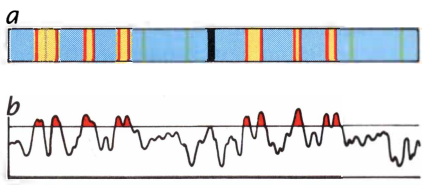
The evolution of a gene that protects an organism from such toxins would have provided a tremendous survival advantage. The P-glycoprotein gene could be a highly evolved descendant of such a primordial gene, protecting higher organisms from the natural toxins to which they are normally exposed through ingestion of food tainted by spoilage or contaminated by one or another of a myriad of toxic plants.

A second possibility is that P-glycoprotein is involved in some transport process critical to the physiology or development of a complex organism such as a mammal. With cDNA probes and monoclonal antibodies it has been determined that P-glycoprotein is normally expressed in the kidneys, adrenal glands, liver and parts of the gastrointestinal tract of the normal adult. These tissues are involved in the transport of nutrients and solutes and in secreting a variety of protein and steroid substances; perhaps P-glycoprotein plays a role in some of the processes.

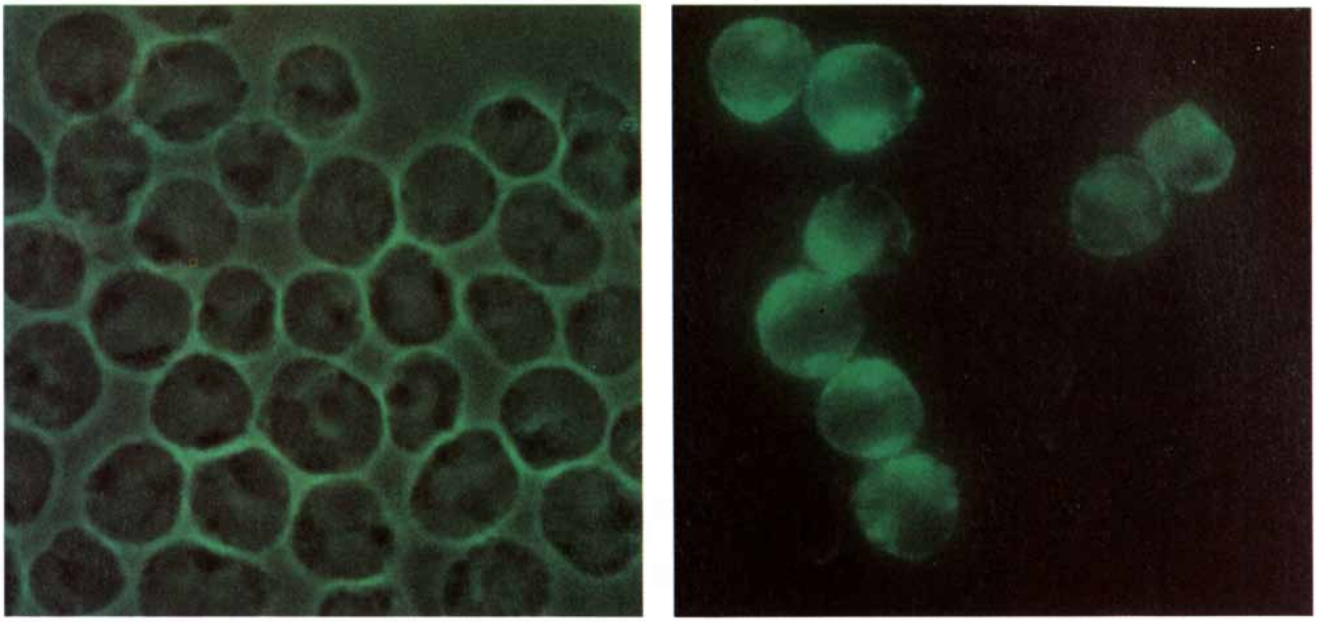


GENETIC BASIS of multidrug resistance can be probed with cloned P-glycoprotein cDNA. In Northern blotting, mRNA is extracted from cells, separated by electrophoresis and transferred to filter paper. When radioactively labeled cDNA is applied to the filter, it binds to, and thereby labels, the corresponding mRNA—P-glycoprotein mRNA in this case. The procedure shows that whereas drug-sensitive cells (*a*) produce little mRNA for P-glycoprotein, drug-resistant cells (*b*) produce an amount corresponding to their level of resistance. The source of the increased mRNA is revealed by Southern blotting, which probes DNA rather than RNA. The DNA is cut into fragments, separated by electrophoresis, transferred to a filter and exposed to radioactive cDNA. In both sensitive and resistant cells the cDNA probe identifies eight DNA fragments, suggesting that P-glycoprotein is encoded by a family of genes. The fragments stain much more intensely in resistant cells (*b*), suggesting that resistance develops when the genes are amplified (copied many times).





STRUCTURE OF P-GLYCOPROTEIN was inferred from its sequence of amino acids, which revealed that the protein chain (*a*) has two similar halves. Each amino acid was assigned a value for hydropathy: affinity for a fatty environment such as the cell membrane rather than an aqueous environment such as the interior or exterior of the cell. A graph of hydropathy (*b*) for the chain suggested that 12 separate segments (*red*) are embedded in the cell membrane. The sequence information also revealed some segments (*green*) likely to bind the energy-carrying molecule ATP and a region (*purple*) where sugar chains are likely to be attached. The information about the segments of the chain suggested the model for the protein's structure (*left*).



LEUKEMIC CELLS, some of which are drug-sensitive and others of which are drug-resistant, are indistinguishable in a photomicrograph (*left*). When the same cells are exposed to a fluorescent antibody that binds specifically to P-glycoprotein, a fluorescence micrograph made under ultraviolet light reveals only the drug-resistant cells, which are rendered fluorescent because they bear high levels of P-glycoprotein (*right*). Such antibody testing can serve to identify drug-resistant cells in tumor biopsies, and the antibodies may someday serve as vehicles for delivering toxins capable of destroying drug-resistant cells. The micrographs were provided by Grace Bradley of the Ontario Cancer Institute and the University of Toronto.

The occurrence of P-glycoprotein in the tissues cited above does not preclude the first theory, since some of the organs are also involved in detoxification processes. It is interesting, as an aside, that such organs often give rise to tumors that are innately drug-resistant; that is, they are unresponsive to combination chemotherapy from the start. Possibly the normal expression of P-glycoprotein in these tissues is preserved in the cancerous cells that originate there. Whatever the normal function of P-glycoprotein may be, it seems likely that it plays some kind of membrane transport role, whether it is the extrusion of exogenous toxic substances or of endogenous, physiologically important cellular products.

Another important question is one that brings us full circle to the rationale for studying experimental drug resistance in the first place: Is P-glycoprotein relevant to the failure of chemotherapy in cancer patients? It has been established that in ovarian carcinomas, leukemia and a variety of sarcomas some of the tumors contain elevated levels of P-glycoprotein. In the small number of cases in which patient follow-ups have been possible, increased amounts of P-glycoprotein have been seen in concert with increasing unresponsiveness to chemotherapy; in perhaps 10 or 20 percent of the tumors tested, clear evidence of raised levels of P-glycoprotein has been obtained. Based on these preliminary studies, one can conclude that a significant fraction of treatment failures might be attributable to P-glycoprotein-mediated multidrug resistance, although much further work needs to be done to substantiate this conclusion.

As more becomes known about multidrug resistance and the function of P-glycoprotein, ways to improve the effectiveness of drugs administered in chemotherapy will become clearer. Recently it has been found that a variety of compounds inhibit the function of P-glycoprotein, rendering multidrug-resistant tumor cells sensitive to drugs that would otherwise be ineffective. These compounds have been referred to as "chemosensitizers." Preliminary research suggests that some of them act by interfering with the binding of drugs to P-glycoprotein. Presumably the binding of a drug molecule to P-glycoprotein is an important first step in its transport out of the cell, so that blockade of the binding allows the drug to accumulate in the cell and kill it as intended. Subtler means of manipulating P-glycoprotein function may eventually come to light, making it possible to use anticancer drugs—which are often extremely effective in the absence of drug resistance—to their full potential.

Another possible way to defeat multidrug-resistant tumor cells may be to exploit the very fact that they contain P-glycoprotein. Perhaps monoclonal antibodies bearing a radioactive compound or a toxic drug could be targeted to P-glycoprotein in order to kill tumor cells that are untreatable by conventional means. Paul Ehrlich's vision of a magic bullet may yet be substantiated.

FURTHER READING

THE RELATIONSHIP BETWEEN TUMOR MASS AND RESISTANCE TO CHEMOTHERAPY: IMPLICATIONS FOR SURGICAL ADJUVANT TREATMENT OF CANCER. Vincent T. DeVita, Jr., in *Cancer,* Vol. 51, No. 7, pages 1209–1220; April 1, 1983.

THE GENETIC ORIGIN OF DRUG RESISTANCE IN NEOPLASMS: IMPLICATIONS FOR SYSTEMIC THERAPY. James H. Goldie and Andrew J. Coldman in *Cancer Research,* Vol. 44, No. 9, pages 3643–3653; September, 1984.

MULTIDRUG RESISTANCE. James H. Gerlach, Norbert Kartner, David R. Bell and Victor Ling in *Cancer Surveys,* Vol. 5, No. 1, pages 25–46; 1986.

RESISTANCE TO MULTIPLE CHEMOTHERAPEUTIC AGENTS IN HUMAN CANCER CELLS. Michael M. Gottesman and Ira Pastan in *Trends in Pharmacological Science,* Vol. 9, pages 54–58; February, 1988.

WHY YOU SHOULD CON
386 SYSTEMS, DESPITE THEIR

Our new 386^-based systems are priced about 35% less than comparable systems–like Compaq's^. Which may make you wonder if we've left something important out. Like high performance.

Well we haven't.

In fact, these are among the fastest 386-based systems available. With more advanced features than you'd get in systems that list for up to $3000 more.

Like Compaq's.

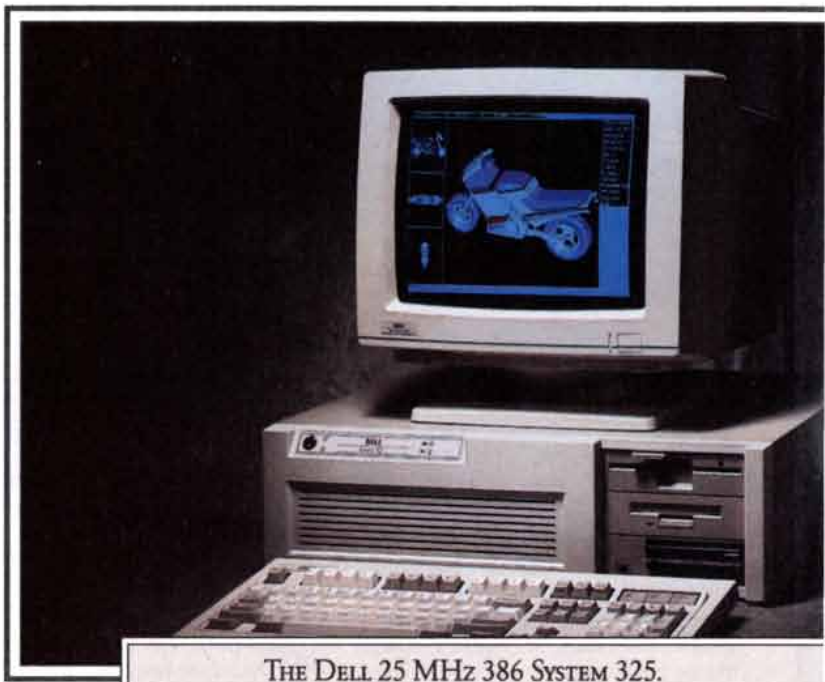
For instance, our 20 MHz System 310 offers you the most extraordinary value available in any 386-based system. It's the machine that PC Magazine (6/14/88) described as "fast enough to burn the sand off a desert floor."

AND IF THAT SOUNDS FAST, WAIT TILL YOU SEE OUR NEW 25 MHz 386-BASED SYSTEM.

At 25 MHz, our new System 325 offers you the highest possible performance in a 386.

Like the System 310, it utilizes the very latest technology, including the Intel^ 82385 Cache Memory Controller, advanced 32-bit architecture and high performance drives. And of course, both systems are fully IBM^ PC compatible.

But speed isn't the only reason to buy from Dell. Or even the best.



THE DELL 25 MHz 386 SYSTEM 325.

STANDARD FEATURES: • Intel^ 80386 microprocessor running at 25 MHz. • 1 MB of RAM* expandable to 16 MB using a dedicated high speed 32-bit memory slot. • Advanced Intel^ 82385 Cache Memory Controller with 32 KB of high speed static RAM cache. • Page mode interleaved memory architecture. • VGA systems include a high performance 16-bit video adapter. • Socket for 25 MHz Intel 80387 or 25 MHz WEITEK 3167 math coprocessor. • 5.25" 1.2 MB or 3.5" 1.44 MB diskette drive. • Enhanced 101-key keyboard. • 1 parallel and 2 serial ports. • 200-watt power supply. • 8 industry standard expansion slots.

**Lease for as low as $245/Month.*

The Dell System 325 is an FCC Class A device, intended for business use only.

SYSTEM 325	WITH MONITOR & ADAPTER	
Hard Disk Drives	VGA Mono	VGA Color Plus
150 MB-18 ms ESDI	$6,799	$7,099
322 MB-18 ms ESDI	$8,799	$9,099

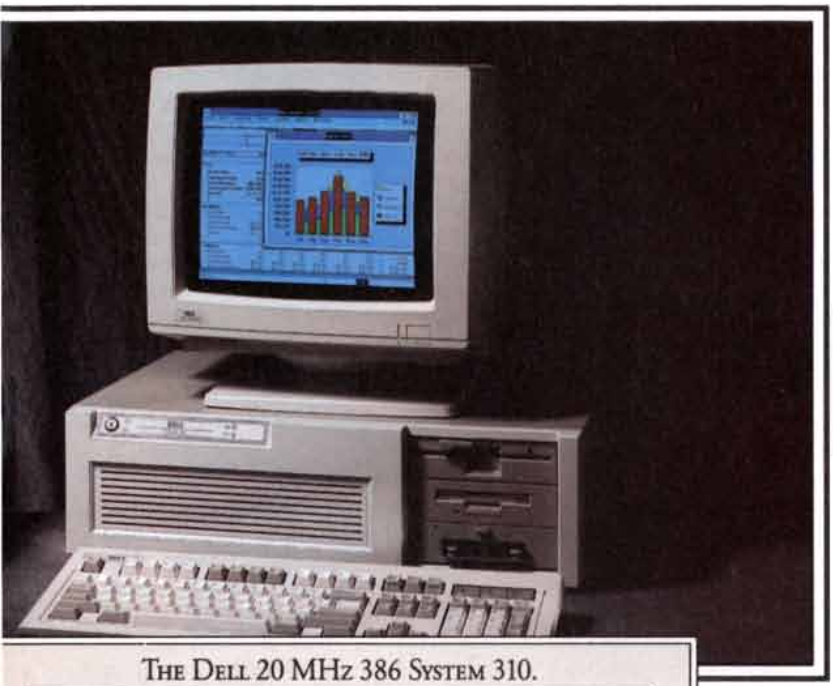
SYSTEM 325 AND 310 OPTIONS: • Intel 80387 math coprocessor: 25 MHz for 325; 20 MHz for 310. • 1 MB or 4 MB memory upgrade kit. • 2 MB or 8 MB memory expansion board kit. • Dell Enhanced Microsoft* MS-DOS* 3.3. • Dell Enhanced Microsoft MS-DOS 4.0. • Both

THE FIRST PERSONAL COMPUTER THAT'S TRULY PERSONAL.

When you order from Dell, we custom configure a system to your exact personal specifications. After evaluating your busi-

All prices and specifications are subject to change without notice. Dell cannot be responsible for errors in typography or photography. **Payments based on a 36-month open-end lease. ◊Service in remote locations will incur additional travel charges. △For the complete terms of our On-Site Service Contract, Satisfaction Guarantee, Warranty and leasing plans, write: Dell Computer

SIDER THE NEW DELL SUSPICIOUSLY LOW PRICES.



THE DELL 20 MHz 386 SYSTEM 310.

STANDARD FEATURES: • Intel 80386 microprocessor running at 20 MHz. • 1 MB of RAM* expandable to 16 MB using a dedicated high speed 32-bit memory expansion slot. • Advanced Intel 82385 Cache Memory Controller with 32 KB of high speed static RAM cache. • Page mode interleaved memory architecture. • VGA systems include a high performance 16-bit video adapter. • Socket for 20 MHz Intel 80387 or 20 MHz WEITEK 3167 math coprocessor. • 5.25" 1.2 MB or 3.5" 1.44 MB diskette drive. • Enhanced 101-key keyboard. • 1 parallel and 2 serial ports. • 200-watt power supply. • 8 industry standard expansion slots.

****Lease for as low as $140/Month.**

SYSTEM 310	WITH MONITOR & ADAPTER	
Hard Disk Drives	VGA Mono	VGA Color Plus
40 MB-28 ms	$3,899	$4,199
90 MB-18 ms ESDI	$4,699	$4,999
150 MB-18 ms ESDI	$5,199	$5,499
322 MB-18 ms ESDI	$7,199	$7,499

MS-DOS versions with disk cache and other utilities. • Dell Enhanced MS* OS/2. *640 KB is available for programs and data. The remaining 384 KB is reserved for use by the system to enhance performance.

ness needs, we will help you select the features that are right for you. After your system unit is custom built, we burn-in everything to make sure the entire system works perfectly.

TOLL-FREE SUPPORT AND ON-SITE SERVICE INCLUDED IN THE PRICE.

Every Dell system includes a complete set of diagnostic tools. So troubleshooting is easy. In fact, most problems can be resolved over our toll-free support line. It's staffed by Dell's own expert technicians from 7 AM to 7 PM (CT) every business day.

TO ORDER, PLEASE CALL
800-426-5150
IN CANADA, CALL 800-387-5752

And if your system requires hands-on service, a technician will be at your location the next business day. At no cost to you.◊

Included in the price of your system is a full year of on-site service.

But that's not all. You're also protected by our 30-day money-back guarantee. And our one-year limited warranty on parts and workmanship.△

AND IF YOU STILL THINK YOU GET WHAT YOU PAY FOR, CONSIDER THIS.

When you buy from Dell, you buy directly from our manufacturing facility in Austin, Texas. Which means we eliminate dealer mark-ups, allowing us to give you a lot more 386 for less. We can even design a custom lease plan for your business, which gives you another way to save.

So go beyond your suspicions. Call us at (800) 426-5150 and order the system that's right for you.

DELL COMPUTER CORPORATION

Corporation, 9505 Arboretum Blvd., Austin, Texas 78759-7299. In Canada, configurations and prices will vary. Microsoft, MS and MS-DOS are registered trademarks owned by Microsoft Corp. Signifies trademarks of entities other than Dell Computer Corporation. ©1988 DELL COMPUTER CORPORATION.

AD CODE NO. **11PC9**

Plasma Particle Accelerators

Electric fields generated in plasma can propel electrons to high energies. The process promises fields 10,000 times stronger than those of the most powerful conventional accelerators

by John M. Dawson

Smashing charged particles together at nearly the speed of light, high-energy accelerators have advanced understanding of the structure of matter, the fundamental forces of nature and the origin of the universe. In the 1930's cyclotron accelerators generating energies of a million electron volts (MeV) simulated conditions in the cores of giant stars, providing a laboratory environment for studying nuclear reactions. More recently synchrotrons and linear accelerators, at a billion electron volts (GeV), have explored the environment in the interior of neutron stars and proved the existence of antimatter. Today proton synchrotrons, at a trillion electron volts (TeV), are probing the conditions of the universe to within a billionth of a second of its birth. As plans are laid for the world's largest accelerator, the Superconducting Supercollider, accelerator technology is approaching practical limits. Fortunately a new technology, plasma particle acceleration, has emerged that could constitute a promising way to achieve still higher energies.

The $4.4-billion Superconducting Supercollider (SSC) will require an 87-kilometer accelerator ring to boost particles to 40 TeV [see "The Superconducting Supercollider," by J. David Jackson, Maury Tigner and Stanley Wojcicki; SCIENTIFIC AMERICAN, March, 1986]. The SSC's tremendous size is due in part to the fact that its operating principle is the same one that has dominated accelerator design for 50 years: it guides particles by means of magnetic fields and propels them by strong electric fields. If one were to build an equally powerful but smaller accelerator than the SSC, one would need to increase the strength of the guiding and propelling fields. Actually, however, conventional technology may not be able to provide significant increases in field strength. There are two reasons. First, the forces from magnetic fields are becoming greater than the structural forces that hold a magnetic material together; the magnets that produce these fields would themselves be torn apart. Second, the energy from electric fields is reaching the energies that bind electrons to atoms; it would tear electrons from nuclei in the accelerator's support structures.

It is the electric field problem that plasma particle accelerators can overcome. Plasma particle accelerators are based on the principle that particles can be accelerated by the electric fields generated within a plasma: a state of matter heated to a temperature at which electrons are stripped from their atomic hosts. Because the plasma has already been ionized, plasma particle accelerators are not susceptible to electron dissociation. They can in theory sustain accelerating fields thousands of times stronger than conventional technologies. If accelerating fields of such magnitude could be produced over extended distances, a plasma particle accelerator a few hundred meters long could match the projected energy of the 87-kilometer Superconducting Supercollider.

Electric fields that accelerate particles can be created in a plasma because of the medium's remarkable properties. A plasma as a whole is electrically neutral, but because the electrons and the positively charged ions are separated, a disturbance can create regions of negative charge (high concentrations of electrons) and regions of positive charge (high concentrations of positive ions). Such an uneven distribution of charge sets up an electric field, which runs from positive to negative regions. The electric field pulls the electrons and ions together with equal force. Since an electron's mass is much smaller than that of an ion, the electrons move toward the positive regions, whereas the ions remain essentially stationary.

As the electrons from negative regions are drawn to positive regions, they steadily gain velocity and momentum. The momentum does more than carry the electrons to a positive region: it causes them to overshoot it, whereupon the electric field reverses direction, first opposing the electrons' motion and slowing them down and then pulling them back again. The process repeats itself, establishing an electron "pendulum."

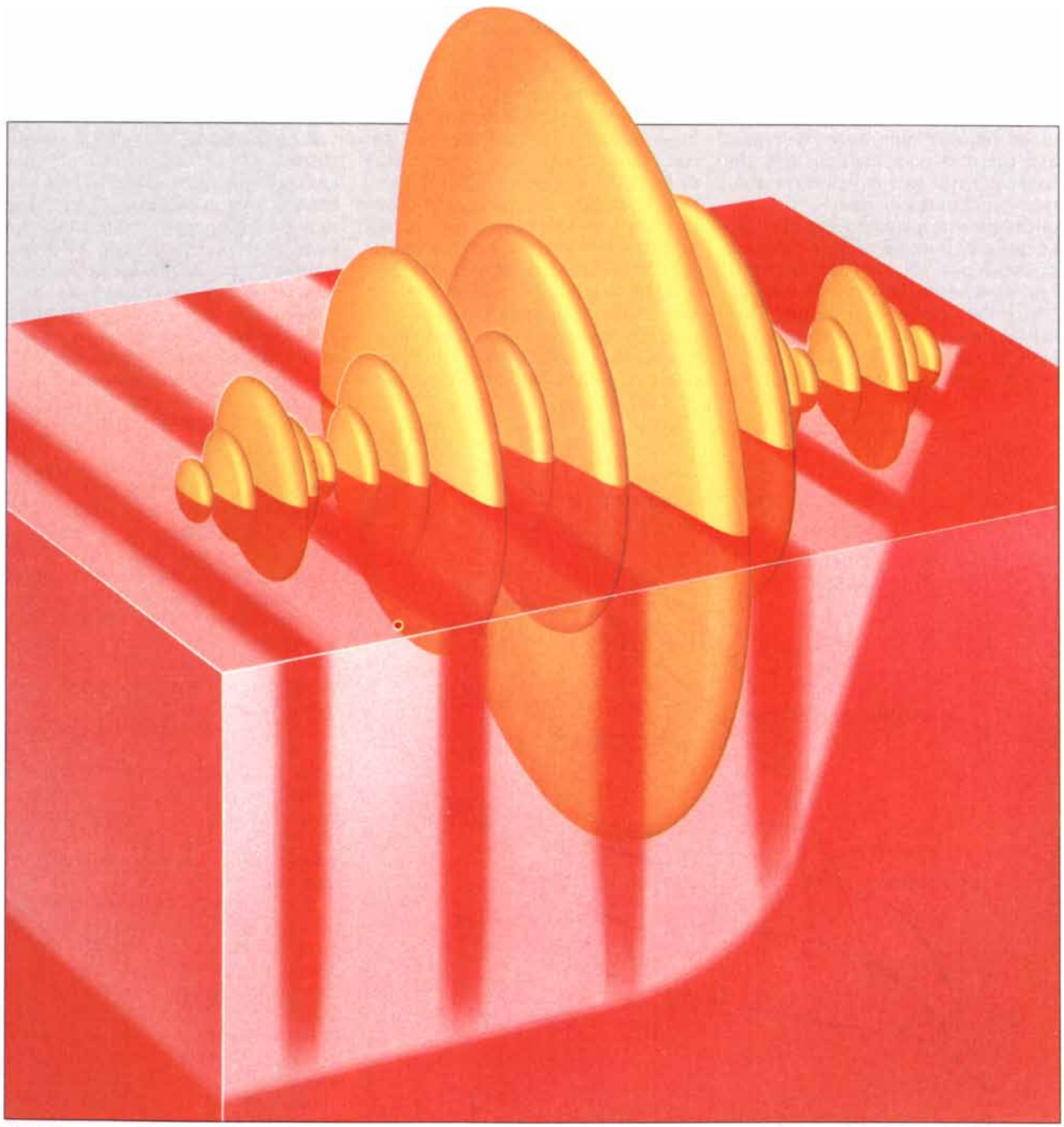
An array of such electron pendulums, created by disturbances in a plasma, can create an electric field that accelerates charged particles. Imagine a thin cylinder of plasma, its long axis horizontal. Every electron pendulum is centered at a different point in the cylinder and oscillates to the left and right, parallel to the axis. Suppose the pendulums are made to oscillate in sequence. Say a pendulum at the left end of the plasma cylinder occupies its leftmost position. A small distance along the plasma cylinder another electron pendulum is in its center position, and it is followed by a pendulum in its rightmost position. Next comes a pendulum at its center position and then one at its leftmost position [*see top illustration on page 57*]. If all these electron pendulums were to start oscillating at the same frequency and achieve the same maximum amplitude, regions of negative (or positive) charge would appear to move along the cylinder in a traveling wave. The effect would resemble the

JOHN M. DAWSON is professor of physics and a member (former director) of the Center for Plasma Physics and Fusion Energy at the University of California, Los Angeles. He received his B.S., M.S. and Ph.D. from the University of Maryland. He has won the James Clerk Maxwell Prize in Plasma Physics given by the American Physical Society. His professional interests include thermonuclear fusion, plasma theory and computer simulations of plasma. Dawson is also an amateur astronomer.

lights on a theater marquee: although each bulb switches on and off in sequence, waves of light seem to move across the sign.

In this way the combined motion of the electron pendulums forms a longitudinal wave of positive and negative regions traveling through the plasma—a plasma wave. In turn the positive and negative regions establish an electric field that travels along with the plasma wave. If a charged particle is injected into the plasma at approximately the same velocity as the plasma wave, it will stay in phase with the field, absorb energy from the field and accelerate steadily. This phenomenon is the basis for particle acceleration in plasmas.

The acceleration of the particle does not follow the intuitive notion of acceleration as being an increase in velocity over time. Charged particles in present-day high-energy accelerators travel at nearly the speed of light. An electron from a 50-GeV accelerator falls short of the speed of light by only five parts in 10^{11}, that is, if an electron raced a light pulse around the earth,



PLASMA, a state of matter in which electrons dissociate from ions, supports waves of charge (*red*). The waves are generated by two beams of laser light, which create an interference pattern called the beat wave (*yellow*). As the beat wave travels into undisturbed plasma, radiation pressure from the wave causes electrons to oscillate in such a manner that regions of charge move along the plasma. This traveling wave of charge creates an electric field that can accelerate charged particles.

the electron would cross the finish line only a tenth of a millimeter behind the light. When particles traveling at these speeds absorb energy from a field, they are accelerated in the sense that their mass increases in accordance with Einstein's theory of relativity. The particles' velocity, however, increases very little. In order to accelerate (to add mass to) a beam of charged particles already moving at very high speed, a plasma accelerator must create a plasma wave traveling at nearly the speed of light so that charged particles do not outrun the electric field wave.

The velocity of a plasma wave is the product of its wavelength (the distance between one negative region and the next one) multiplied by the frequency (the rate at which the electron pendulums oscillate). The frequency of the plasma wave in a given plasma is fixed; it is a function of the density of available electrons. The wavelength, however, is variable. Hence the velocity of a plasma wave can be made equal to the speed of light by giving it the proper wavelength. Given a plasma wave at the plasma's natural frequency and the proper wavelength, charged particles can be injected into the plasma and remain at a constant phase of this wave, gradually gaining energy from the electric field.

There is a constraint on the amount of energy a plasma wave can impart to an electron. The strongest electric field that can be produced in a plasma is set by the maximum amplitude of the oscillations, which in turn is determined by wave breaking. Just as an ocean wave breaks when its crest falls into its trough, so a plasma wave can break when one set of electron pendulums swings into a region before electron pendulums from another region have vacated it. For a plasma wave that has a phase velocity equal to the speed of light, theory predicts that the limiting amplitude due to wave breaking is related to the square root of the undisturbed electron density. Plasmas that have densities of between 10^{16} and 10^{21} electrons per cubic centimeter are routinely produced in laboratories, and so the strongest accelerating fields are in theory between 100 million and 30 billion volts per centimeter. In comparison, accelerating fields of 200,000 volts per centimeter are generated in existing accelerators, and fields of one or two million volts per centimeter are projected for future conventional accelerators.

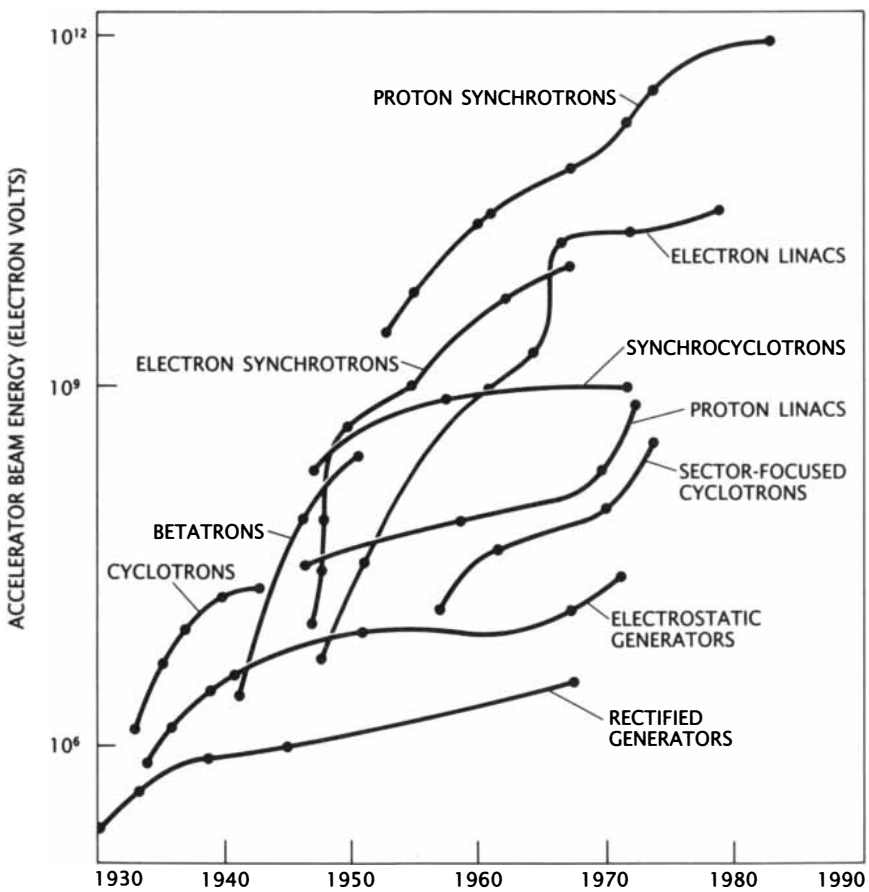
Since plasma waves promise stronger electric fields than conventional technologies, they are good prospects for accelerating particles to high energy. How, then, are plasma waves generated? So far two methods for creating plasma waves for accelerators have been proposed and tested: the wake field and the beat wave.

The wake-field method exploits a "bunch" of many electrons to generate waves in a plasma, thereby accelerating a group of a few electrons to higher energies. Just as a boat moving through the water pushes some of the water aside, making a wake, so a bunch of many electrons traveling through plasma generates a wake of plasma waves.

As an electron bunch enters a region, the plasma electrons move out of the way so that the combination of the plasma and the bunch remains electrically neutral. As the electron bunch exits, leaving a deficit of electrons in the region, the plasma electrons rush back to reestablish equilibrium. This movement of plasma electrons initiates the oscillation of the electron pendulums and results in a plasma wave traveling at a velocity equal to that of the electron bunch. The plasma wave establishes an electric field: the so-called wake field. A group of a few electrons, suitably placed in the wake field, can then be accelerated to energies higher than those of the driving bunch.

To some extent, however, the plasma electrons resist being pushed aside by the driving bunch, thereby reducing the energy available to produce plasma waves. This resistance can be minimized by shaping the electron bunch so that the density of electrons first increases slowly and then is cut off rapidly. In this manner the plasma's electrons adjust easily to the incoming bunch but are then driven into large oscillations. By devising a computer model for this process, Jao-Jang Su of the University of California at Los Angeles has shown that an appropriately shaped 75-MeV bunch can accelerate electrons to 1,000 MeV.

Turbulence can break up plasma waves and also can lead to significant energy loss by producing instabilities in both the driving bunch and the accelerated group of electrons. The driving bunch is the more susceptible



ENERGIES generated in high-energy particle accelerators have increased more than one million times since 1930. Each line represents the progress of a major type of particle-accelerator technology, and each dot marks the energy they have achieved.

to instabilities, because it has more electrons and less energy than the accelerated group. There are two critical types of instability in the driving bunch: the two-stream instability and the Weibel instability.

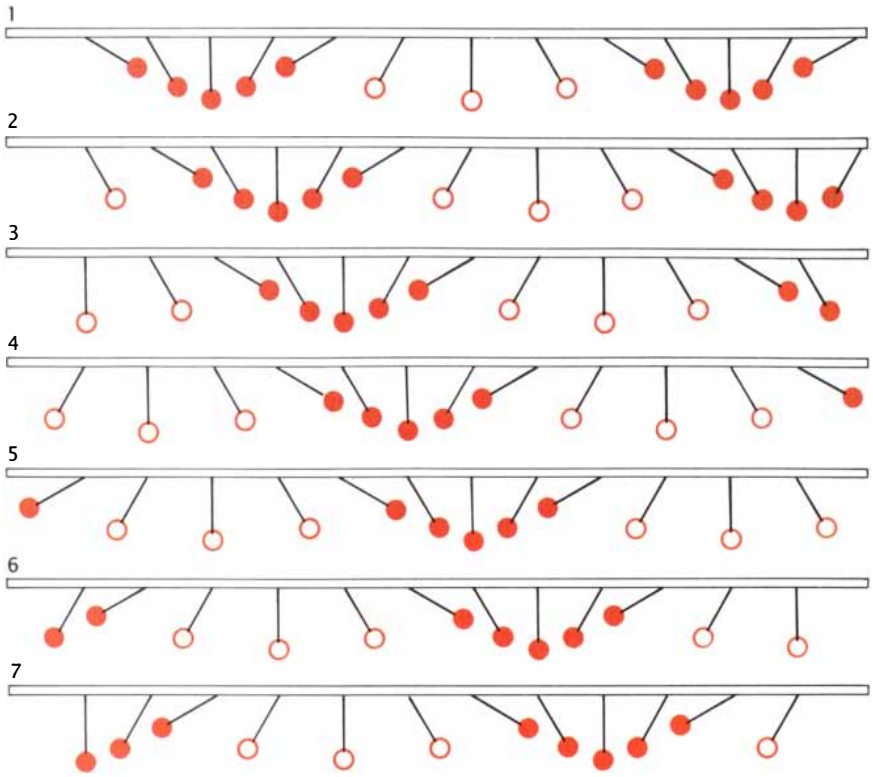
The two-stream instability develops when some electrons in the bunch move faster than others. As the bunch moves through the plasma this difference in velocity becomes more pronounced, decreasing the efficiency with which plasma waves can be produced. Simulations by Su indicate that, in spite of the two-stream instability, a plasma wave can absorb approximately 80 percent of the energy of the driving bunch before it breaks up.

In the Weibel instability the driving bunch subdivides and contracts into several concentrated "filaments" within the plasma. Each electron of the driving bunch runs parallel to the others and constitutes an independent electric current. Because parallel currents attract one another, the electrons of the bunch tend to move closer together. The bunch as a whole does not contract uniformly; rather, groups of electrons move together to form filaments.

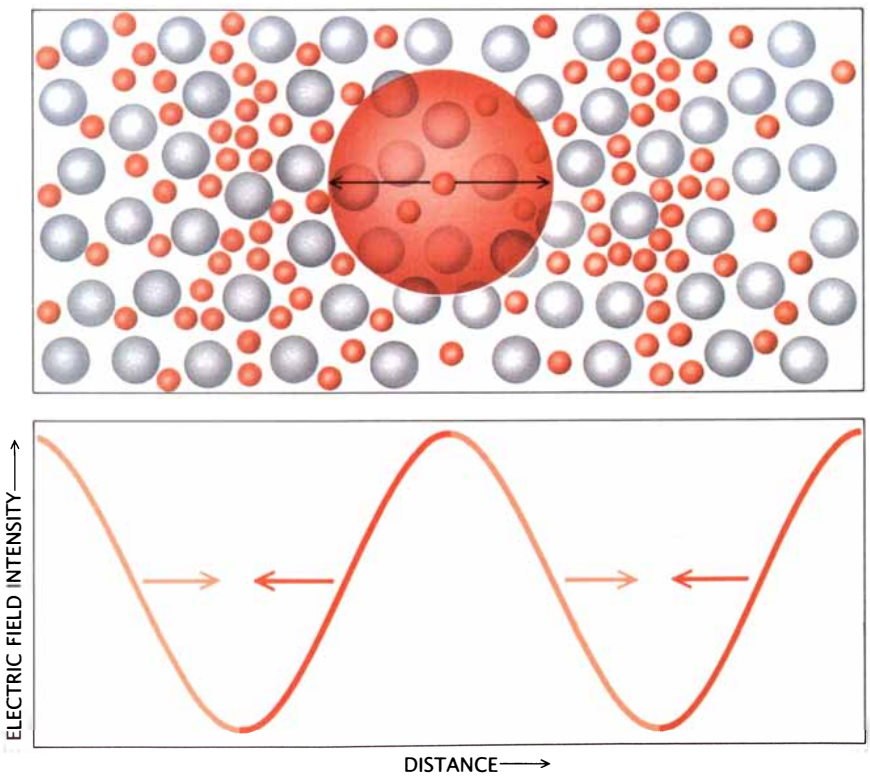
It appears that one can control Weibel instability by giving the electrons of the driving bunch a little random motion perpendicular to the direction of the bunch. The random motion tends to disperse any filaments that form. A transverse energy of about 50,000 eV, which is easily generated, suffices to stabilize bunches that have energies of 100 MeV. Although this added transverse energy will tend to disperse the bunch too far, the tendency can be countered by introducing a guiding magnetic field. If the Weibel instability is kept under control, Thomas C. Katsouleas and Scott C. Wilks of U.C.L.A. have found, the efficiency with which energy can be transferred from the wake field to the accelerated particles is 30 percent.

These results have been partially confirmed by the first experimental test of a wake-field accelerator. The experiment was started by David B. Cline of U.C.L.A. and the University of Wisconsin at Madison in collaboration with workers from the Argonne National Laboratory. James B. Rosenzweig of Wisconsin built the plasma source, and James D. Simpson of Argonne directed the conversion of a 20-MeV linear accelerator into a source of electron bunches for producing the wake field.

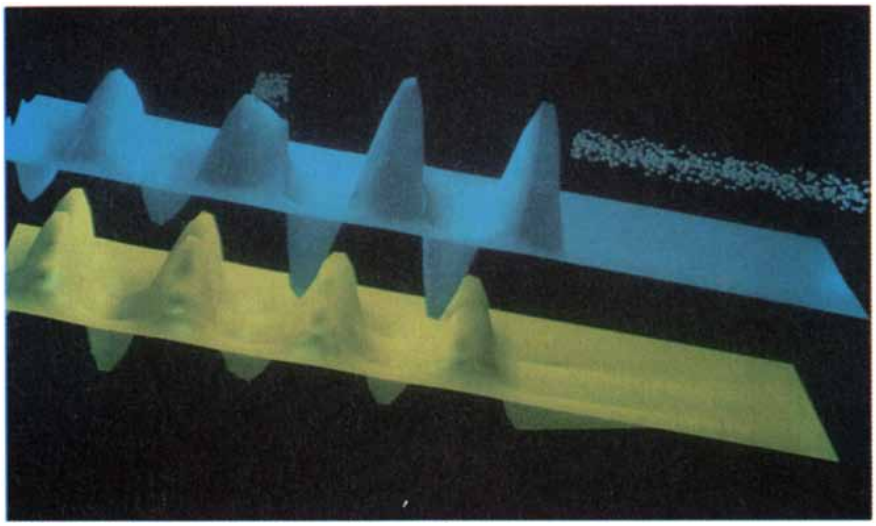
A small fraction of the electrons generated by the linear accelerator provided the electrons to be accelerat-



TRAVELING WAVE develops in a row of pendulums oscillating left and right about a stationary point. Each pendulum along the row is in a different phase of its oscillation, but each has the same frequency and amplitude. As the row evolves in the course of time, a cluster of pendulums (*solid red*) appears to move from left to right.



JUST AS PENDULUMS oscillating can form traveling clusters, so electrons oscillating in a plasma can be arranged to form a traveling plasma wave. A cross section of a plasma wave (*top*) shows clusters of mobile electrons (*red*) among stationary ions (*gray*). The colored area indicates the region in which an electron typically oscillates. The graph (*bottom*) shows the intensity and direction of the electric field.



WAKE from a bunch of many electrons (*right*) traveling through plasma generates plasma waves and thereby creates an electric field. In this computer simulation the field is shown in two components: axial (*blue*) and radial (*yellow*). The axial component of the electric field accelerates a group of a few electrons (*left*). The intensity of the axial field is low near this group, indicating that the energy from the field is being absorbed efficiently. The radial field focuses the group toward the axis.

ed in the wake field. These electrons were passed through a thin foil to slow them down to 15 MeV and were magnetically deflected from the main bunch. The 15-MeV group then entered a plasma behind the driving bunch and acted as a test pulse. Computer modeling of this experiment by Rhonald K. Keinigs and Michael E. Jones at Los Alamos and by Su indicate that at peak performance the wake can accelerate electrons 300,000 volts per centimeter. So far the actual experiment has achieved only 50,000 volts per centimeter, because there were fewer electrons in the driving bunch than expected. The experiment, however, demonstrated acceleration by a wake field.

The beat-wave method has been studied more intensively in both theory and experiment than the wake-field method. Beat-wave accelerators have already accelerated particles in the laboratory from .5 to 2 MeV and have generated electric fields approximately 10 times stronger than those produced in the most advanced conventional accelerators.

Rather than employing an electron bunch, the beat-wave method generates plasma waves from two intense laser beams whose light has different frequencies. The beams are combined so that the light waves interfere, forming alternating regions where the two waves are in phase (and reinforce) and regions where they are out of phase (and cancel). The result is a beat wave in the composite wave, oscillating at a frequency equal to the difference in the frequencies of the parent beams. If the composite beam is then focused into a plasma, the beam creates regions of high and low radiation pressure. If the frequency of the beats is equal to the plasma's natural frequency of oscillation, the plasma electrons respond resonantly and produce powerful plasma waves.

In the beat-wave process both the evolution of the plasma wave and the propagation of the two laser beams through the plasma are highly complex. Computer models have been constructed that follow the motion of several million charged particles through the electromagnetic fields generated by the particles and the lasers. Each model particle actually represents a large number of plasma electrons, but the model closely emulates nature—including relativistic effects—and has been quite successful in predicting the complex behavior of beat waves interacting with electrons and ions in plasma.

Joseph M. Kindel and David W. Forslund of the Los Alamos National Laboratory and Warren B. Mori of U.C.L.A. have modeled the interaction of a plasma that has a density of 10^{17} electrons per cubic centimeter and two laser beams that have wavelengths of 9.6 and 10.6 microns and intensities of about 10^{16} watts per square centimeter. These values are similar to those already achieved in experiments. In their simulation Kindel, Forslund and Mori achieved an accelerating field along the axis of the laser beam of about 50 percent of the theoretical maximum possible value, which suggests that the beat-wave method puts extremely intense electric fields within reach.

Computer models have also been helpful in diagnosing problems that must be solved for a stable and efficient accelerator. If the composite laser beam does not remain focused as it travels through the plasma, for instance, the beam will be ineffective at generating beat waves over large distances. The width of a laser beam ordinarily grows because of diffraction. The unusual optical properties of intense light in a plasma tend to focus laser beams. Both theory and computer modeling indicate that the width of a laser beam of sufficient intensity will remain stable (at about .2 millimeter in the Forslund, Kindel and Mori model). Such focusing is essential to the beat-wave strategy, since a plasma accelerator would require laser beams to propagate over large distances.

Like the beam, the electrons that are accelerated by the plasma wave must also stay focused in a coherent bunch at the center of the plasma wave. Fortunately the plasma waves generated by laser beams not only create accelerating electric fields along the axis of the laser beam but also produce radial electric fields. These radial fields can focus the electrons to the axis. Two additional effects help the focusing. First, the plasma channel tends to become positively charged, because the radiation pressure from the light exerts an outward push on the plasma electrons. Second, strong confining magnetic fields are produced by the current owing to the accelerated electrons. If the accelerated electrons are focused to the axis, their progress is unimpeded and they do not lose a significant amount of energy.

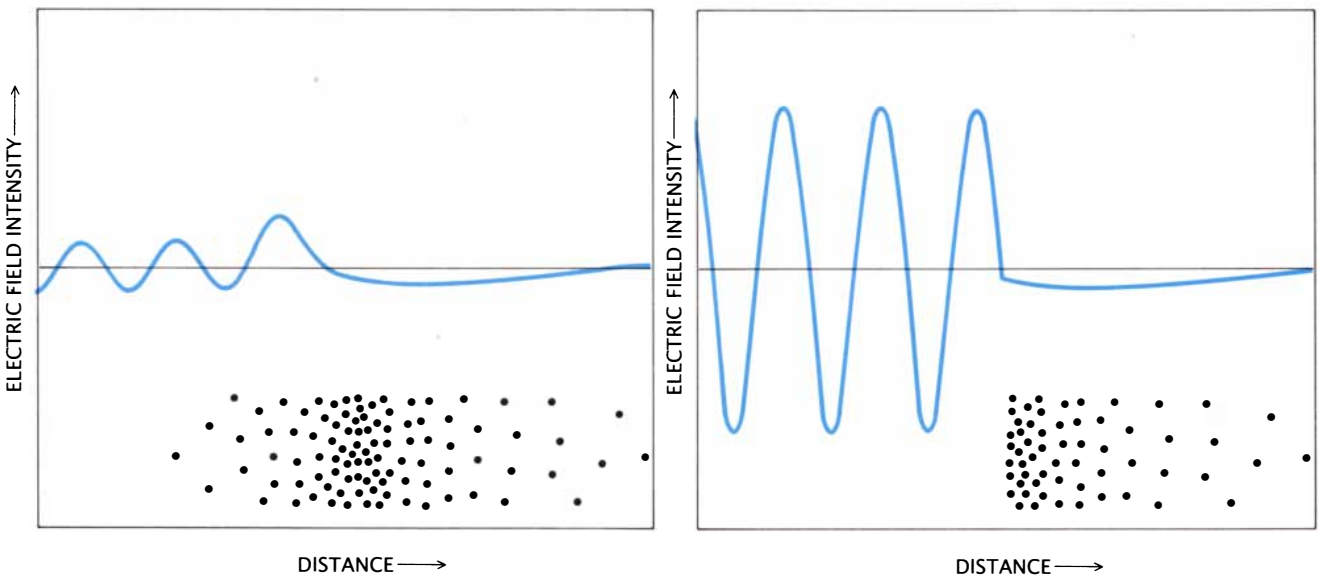
The laser beam can occasionally produce turbulence that disrupts the formation of plasma waves, however. Such turbulence arises when not only the electrons but also the positive ions in a plasma react to the laser beam; the ions must be essentially stationary to produce effective plasma waves. The ions will move, though if the laser beam is focused on a plasma region for more than 10^{-10} second, or .1 nanosecond. That is enough time for radiation pressure from the laser beam to push the ions out of the region of the beam. Therefore if a laser pulse shorter than .1 nanosecond

travels through undisturbed plasma, the pulse can propagate for many meters and generate a nonturbulent plasma wave.

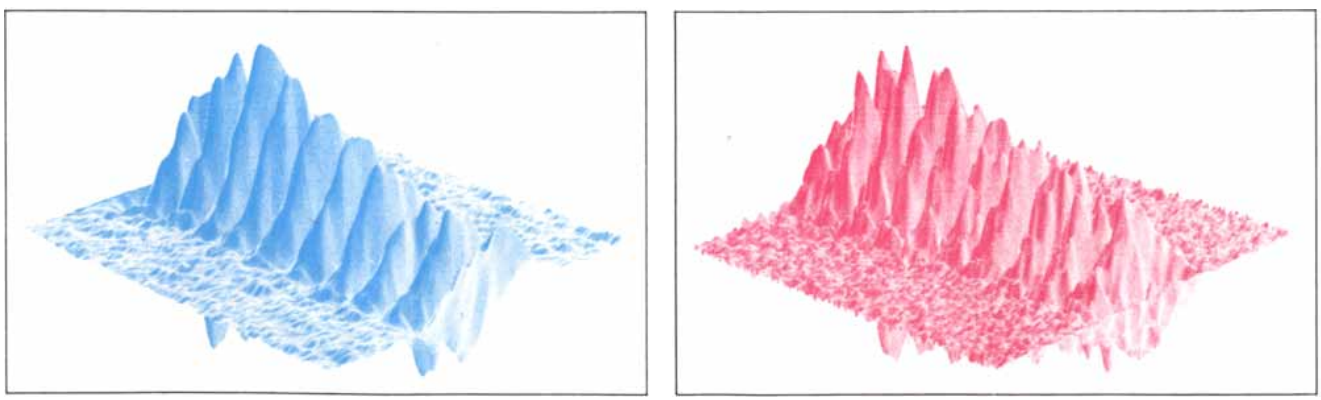
The creation of plasma waves that have enough velocity to accelerate particles efficiently presents another technical difficulty. The velocity with which a beat wave and the plasma wave propagate through the plasma is dependent on the natural frequency of the plasma and the frequency of the lasers that produced the beat wave. This puts a limit on the energy that accelerated electrons can ultimately attain. For instance, lasers that radiate at a frequency of 10^{15} hertz and a plasma that has a natural frequency of 10^{12} hertz could yield electron accelerations as high as 100 million volts per centimeter and maximum energies approaching 1 TeV.

The energy limit can be overcome, however. If the accelerated electrons move at an angle to the direction of wave propagation, the electrons must move faster than the wave itself to keep up with it. The accelerated electrons can be forced to move at the proper angle by applying a magnetic field perpendicular to the plasma wave's direction. Riding waves at an angle is a strategy favored by surfboarders in order to move faster and get a longer ride. Hence the hypothetical accelerator based on this strategy is called the Surfatron.

Soon after Toshiki Tajima of the University of Texas at Austin and I published the first theoretical paper on the beat-wave method in 1979, the first experiments were begun by Chandrashekhar J. Joshi and his collaborators at the National Research Council of Canada in Ottawa. To create a plasma they focused a powerful carbon dioxide laser beam on a thin carbon foil. Instead of combining two additional laser beams to create a beat wave in the plasma, however, they took advantage of an interaction between the first laser and the plasma. The interaction generated secondary light waves whose frequency differed from the frequency of the original beam by the plasma frequency. The interference of the secondary light and the original laser beam formed plasma waves by a mechanism much like the beat-wave process except that the plasma waves propagated in every direction. This experiment gave crude but encouraging results: intense plasma waves were generated and electric fields of about 10 million



STRENGTH of the wake field depends on the shape of the driving electron bunch. A bunch whose density rises and falls slowly forms a weak field (*left*). A bunch whose density rises slowly and becomes zero abruptly forms a strong field (*right*).

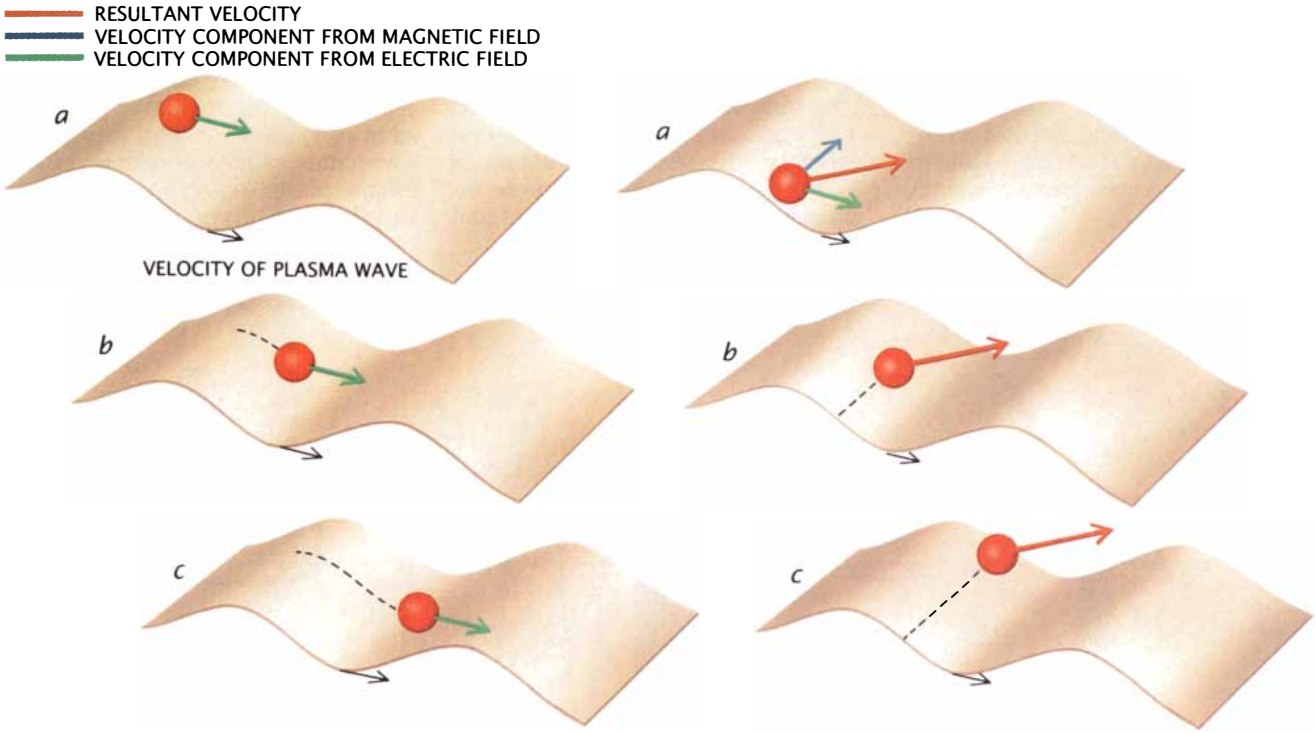


PLASMA WAVE (*left*) from a computer simulation of the beat-wave method generates an electric field (*right*). If a group of charged particles was injected into a plasma and moved in phase with the electric field, the group would accelerate.

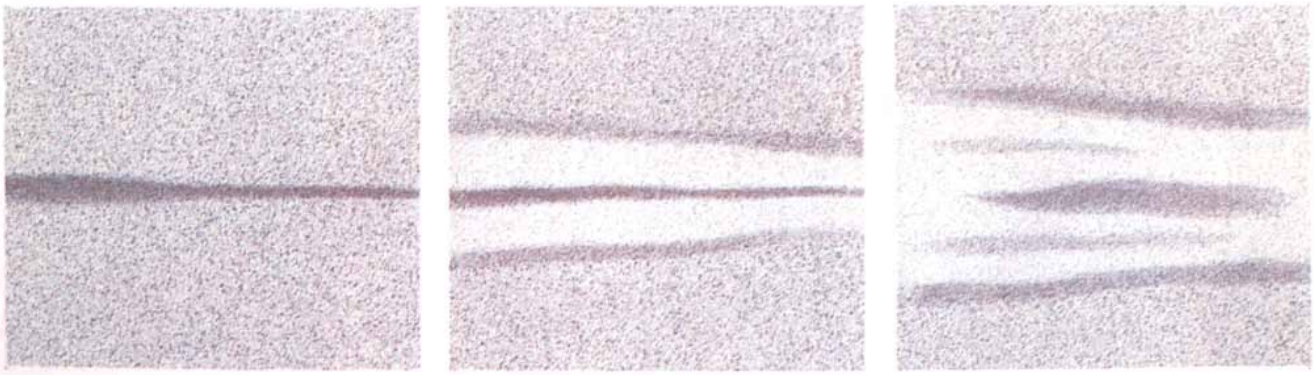
volts per centimeter were produced.

Joshi went on with this research at U.C.L.A. He constructed an experimental device that would drive the plasma waves directly and measure them quantitatively. This time he generated a beat wave with synchronized pulses one nanosecond long from two carbon dioxide lasers that had wavelengths of 9.6 and 10.6 microns. The frequency of the beat wave matched the plasma frequency. Joshi directed the beat wave into a plasma formed by an electric spark. He detected plasma waves by scattering a third laser off the plasma wave, and he observed electric fields whose strengths ranged between three million and 10 million volts per centimeter.

The experiment was deficient in two respects, however. First, as predicted by theory, the nanosecond laser pulses were long enough to disturb the plasma ions, and a large number of turbulent waves developed among the ions, hampering the operation of the accelerator. Second, no electrons were accelerated. The velocities of the plasma electrons were much too low to be accelerated by the plasma wave. Just as surfers must paddle in order to gain enough speed to catch an ocean wave, so electrons must have enough

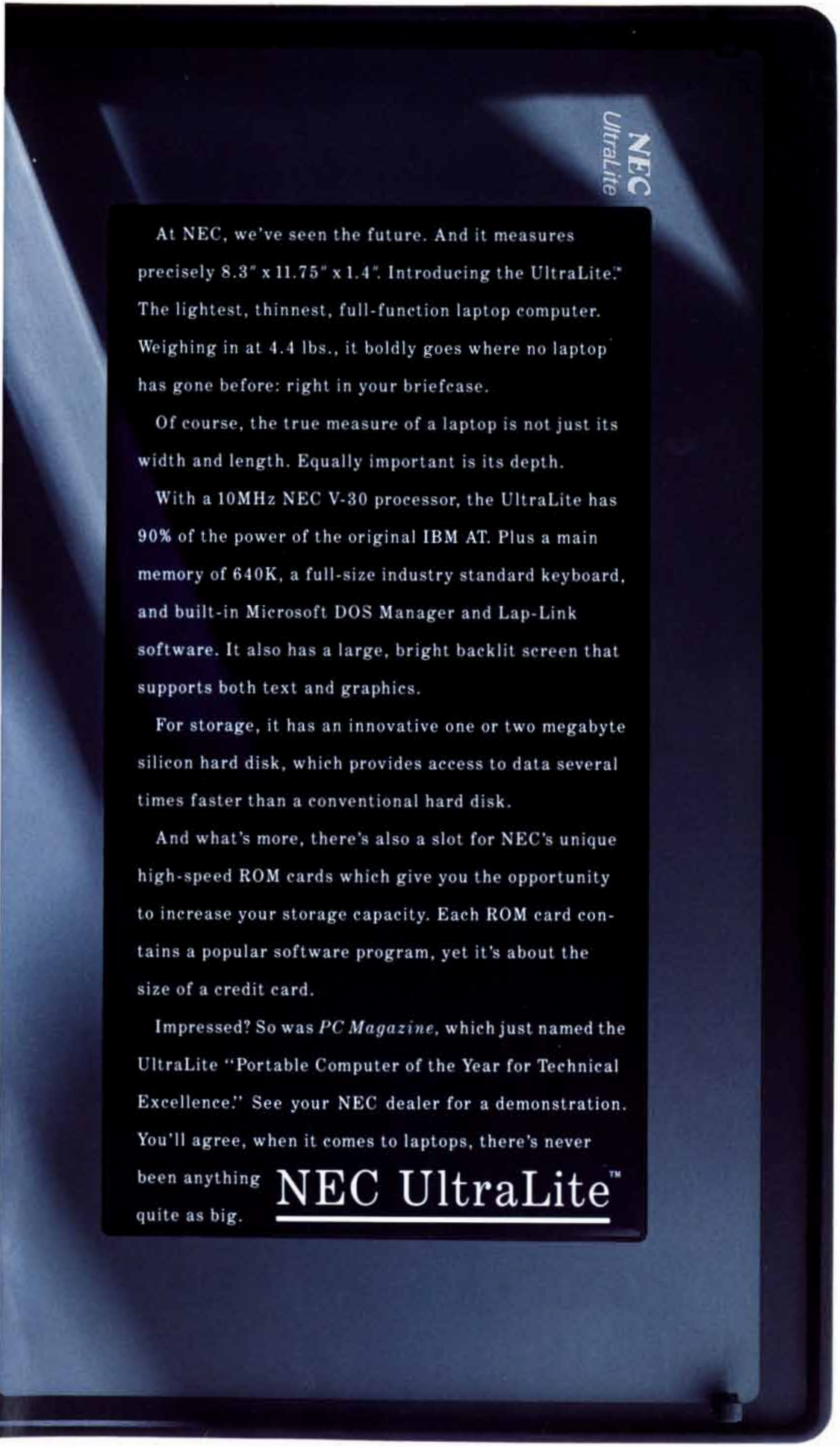


STEADY ACCELERATION of an electron requires that the electron remain in phase with the electric field (*tan*) of the plasma wave. If an electron (*left*) travels slightly faster than the plasma wave and outruns it, the electron will not accelerate and can even decelerate. If an electron (*right*) travels slightly faster than the wave and a magnetic field is applied to move the electron at an angle to the wave, the electron remains in phase and is continuously accelerated to ever higher energies.



PLASMA WAVES break down when radiation pressure from laser light pushes ions (*color*) out of the laser channel. In this case a laser is focused into the plasma continuously for 15 picoseconds. At eight picoseconds the ions are distributed uniformly. At 10 picoseconds the ions begin to leave the laser channel. By 15 picoseconds the ions have been blown out of the laser channel, creating such turbulence in the plasma waves that they are no longer effective in accelerating particles.

Once again,
we're about to place
the future right
in your lap.



At NEC, we've seen the future. And it measures precisely 8.3" x 11.75" x 1.4". Introducing the UltraLite.™ The lightest, thinnest, full-function laptop computer. Weighing in at 4.4 lbs., it boldly goes where no laptop has gone before: right in your briefcase.

Of course, the true measure of a laptop is not just its width and length. Equally important is its depth.

With a 10MHz NEC V-30 processor, the UltraLite has 90% of the power of the original IBM AT. Plus a main memory of 640K, a full-size industry standard keyboard, and built-in Microsoft DOS Manager and Lap-Link software. It also has a large, bright backlit screen that supports both text and graphics.

For storage, it has an innovative one or two megabyte silicon hard disk, which provides access to data several times faster than a conventional hard disk.

And what's more, there's also a slot for NEC's unique high-speed ROM cards which give you the opportunity to increase your storage capacity. Each ROM card contains a popular software program, yet it's about the size of a credit card.

Impressed? So was *PC Magazine*, which just named the UltraLite "Portable Computer of the Year for Technical Excellence." See your NEC dealer for a demonstration. You'll agree, when it comes to laptops, there's never been anything quite as big. NEC UltraLite™

Wherever you can put this magazine, you can now put a computer.

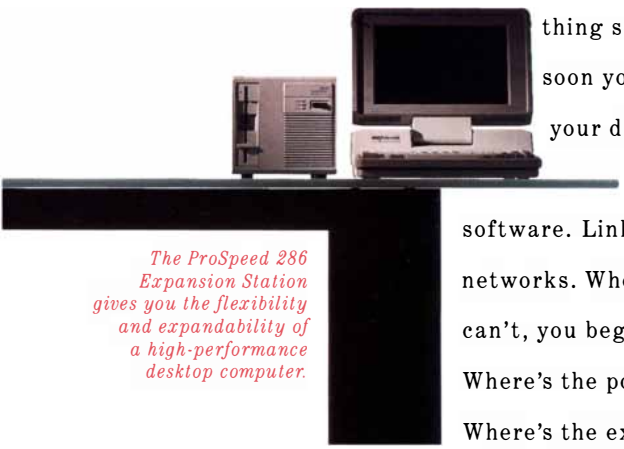
Microsoft and DOS Manager are registered trademarks of the Microsoft Corp. Lap-Link

Introducing the laptop computer

Problems, problems, problems.

Computers were designed to solve them.

But sometimes they've been known to cause a few of their own. Example: You just purchased a laptop. At first, everything seems fine. But soon you want to expand your data base. Use more sophisticated software. Link-up to local area networks. When you realize you can't, you begin to ask yourself: Where's the power? Where's the expandability? Where's the receipt?

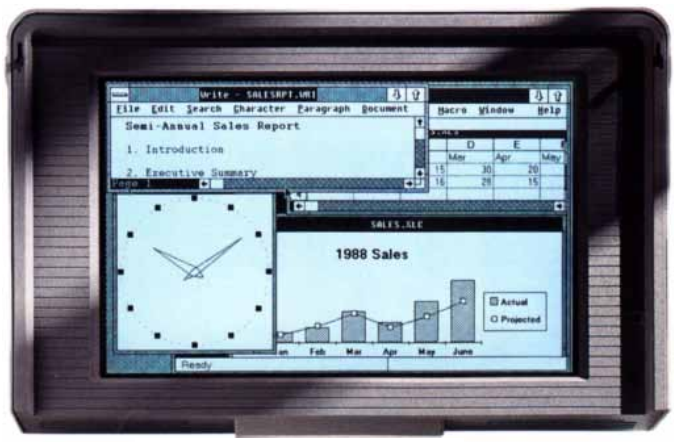


The ProSpeed 286 Expansion Station gives you the flexibility and expandability of a high-performance desktop computer.

Introducing an end to all your problems. The ProSpeed™ 286, from NEC. Weighing under 15 lbs., the ProSpeed 286 is surprisingly light. But in other areas, it's a real heavyweight. Like expandability, connectivity and speed. In fact, it offers the full functionality of a high-performance desktop. The ProSpeed comes equipped with one megabyte of memory that's expandable to five. And, it comes with either 20 or 40 megabyte hard disk drives. There's even an advanced model that provides a high-speed 100 megabyte drive for special applications.

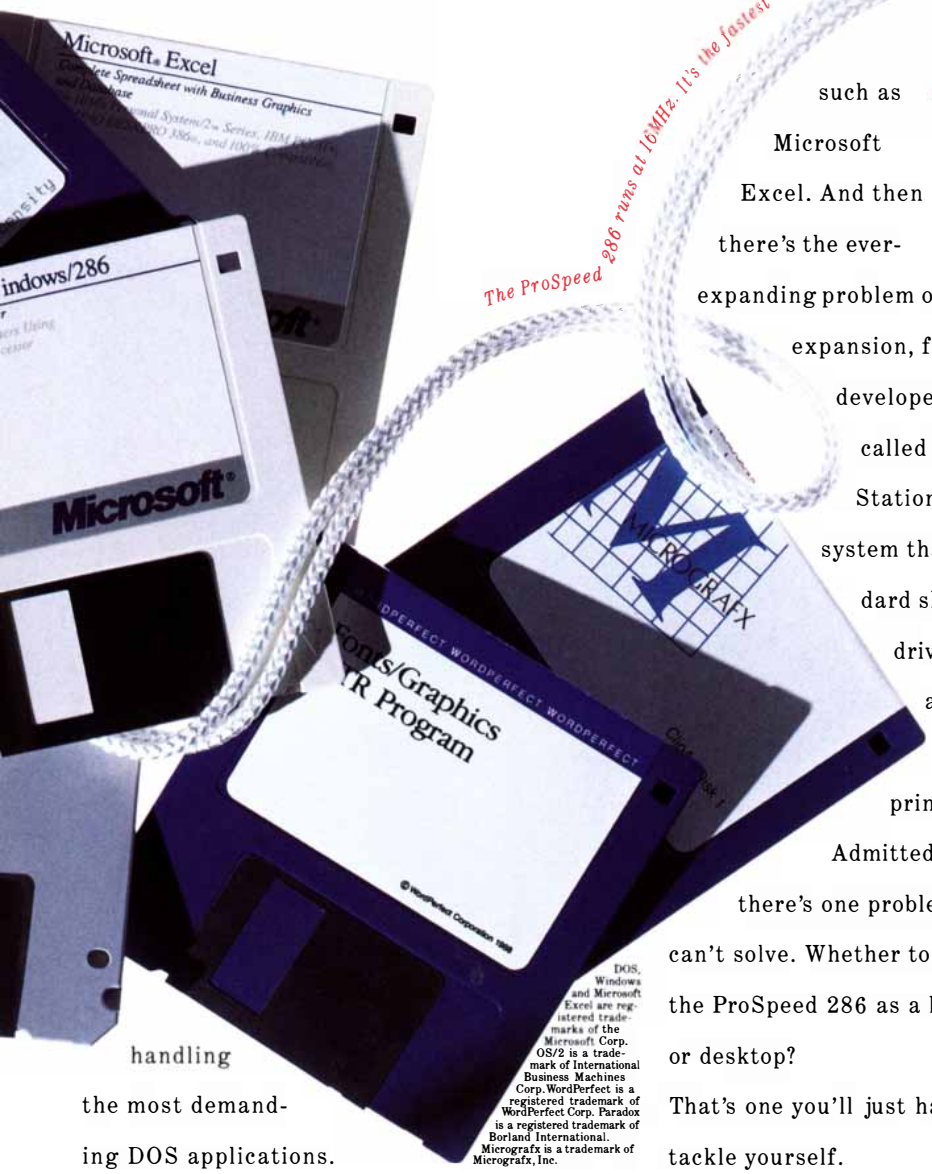
As for its display, it clears up a very serious problem: clarity. NEC's Monograph™ CTN screen gives you crisp, backlit images that provide CRT quality with VGA resolution.

For power-hungry executives, the ProSpeed 286 is powered by a CMOS 80286 processor running at 16MHz. Meaning it's fully capable of



The ProSpeed 286 screen is so clear and crisp you can even display multiple windows.

that solves problems others can't.



The ProSpeed 286 runs at 16MHz. It's the fastest 80286 laptop around.

such as Microsoft Excel. And then there's the ever-expanding problem of expansion, for which NEC has developed a unique solution called an Expansion Station.™ It's an innovative system that includes four standard slots, one standard drive bay, a power supply and permanent connections for both printer and serial devices.

The ProSpeed 286 features a full-function keyboard with standard spacing.

Admittedly, there's one problem we can't solve. Whether to use the ProSpeed 286 as a laptop or desktop?

That's one you'll just have to tackle yourself.

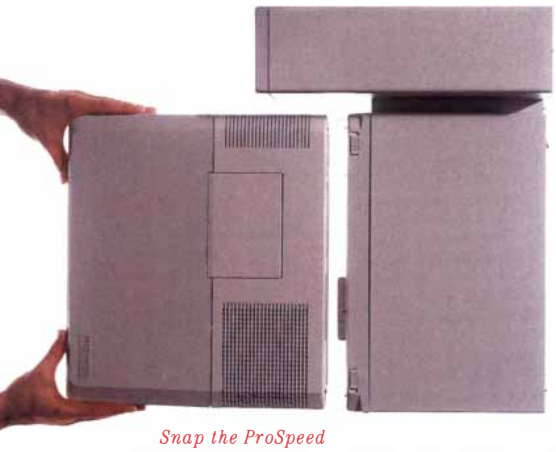
handling the most demanding DOS applications. Not to mention OS/2 Presentation Manager and Windows applications

DOS, Windows and Microsoft Excel are registered trademarks of the Microsoft Corp. OS/2 is a trademark of International Business Machines Corp. WordPerfect is a registered trademark of WordPerfect Corp. Paradox is a registered trademark of Borland International. Micrografx is a trademark of Micrografx, Inc.

NEC ProSpeed™ 286

ProSpeed, Monograph and Expansion Station are trademarks of NEC Home Electronics (USA) Inc.

In order to make the ultimate desktop comp



Snap the ProSpeed 386 into the Docking Station, and it's a desktop. Snap it out and it's a laptop.

Our sincere apologies to desk makers everywhere.

But when we set out to make the ultimate desktop computer, we knew it had to have three critical components. Awesome power. Unequaled expansion capabilities. And most importantly—a handle.

Introducing the ProSpeed™ 386 from NEC. The first personal computer to offer the portability of a laptop and the power of an 80386 desktop.

With the ProSpeed 386, the designers at NEC did more than create a new computer. They established a new class of computers—the first modular workstation.

To realize how this can benefit you in the future, let us remind you how it was in the past.

With ordinary laptops you had to hook and unhook peripherals, phone lines and all sorts of cables every time you left the office.

But thanks to a revolutionary Docking Station™ design, the ProSpeed 386 leaves all that behind.

You simply slide the laptop out of the Docking Station, and then slide out of your office. Leaving all your connections connected for when you return.

And saving yourself

You've never had this much power sitting in your lap.

uter, we had to remove one major obstacle.

one of the most precious commodities of all—time.

Most importantly, the Docking Station can accommodate a full range of expansion options; it has four full-size card slots, and two standard drive bays for tape back-up, CD-ROM or 5¼″ floppies.

Take away the Docking Station and what have you got? One of the world's most powerful laptops.

Running at 16MHz, it offers two megabytes of 32-bit memory that's expandable to 10.

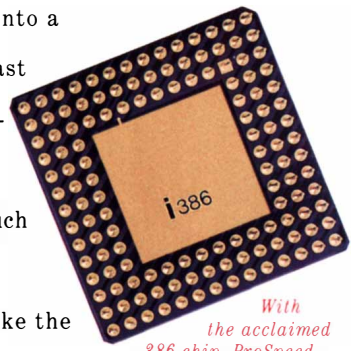
It's equipped with a hard disk and is available with either 40 megabytes or 100 megabytes of storage capacity.

As for its 10½″ diagonal black-on-white display, it's positively brilliant.

Advanced screen technology delivers a paper-white image and EGA resolution.

In more ways than one. NEC's advanced screen technology, called Monograph™ CTN, provides CRT-quality video with a paper-white image and EGA resolution.

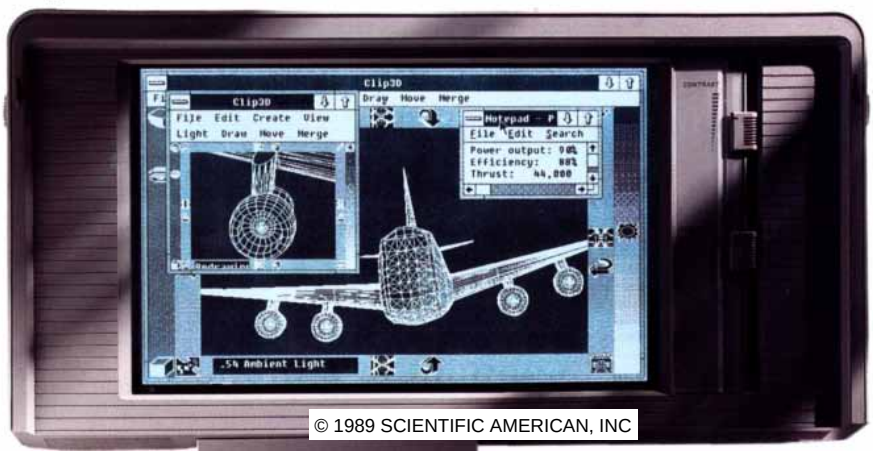
Which translates into a higher level of contrast and increased resolution for graphics-oriented programs, such as Windows.



With the acclaimed 386 chip, ProSpeed delivers exceptional speed and full multi-tasking capabilities.

Which helps to make the ProSpeed 386 not just the ultimate desktop, but the ultimate value. Whether your desk has four legs or just two.

NEC ProSpeed™ 386



For a closer look at the future, see your NEC dealer. For product literature, call 1-800-826-2255.

For technical details, call NEC Home Electronics (USA) Inc. at 1-800-FONE-NEC.

C&C
Computers and Communications

NEC

initial energy to be accelerated by the beat wave. Joshi is hoping to eliminate turbulence by shortening the laser-pulse lengths, and he is injecting some electrons at a high initial energy (1.5 MeV). He expects that the high-energy electrons will be accelerated by a plasma wave to about 10 MeV.

A group under the direction of A. E. Dangor from the Imperial College of Science and Technology in London and Roger G. Evans of the Rutherford Appleton Laboratory near Oxford is currently working on a beat-wave accelerator that relies on neodymium glass lasers producing light at wavelengths of roughly one and 1.1 microns. The group has encountered a number of experimental difficulties but has recently seen evidence of the generation of plasma waves.

Those workers also created a reproducible and uniform plasma—one that is perfect within the limits of their measurements. The plasma was formed by rapidly bombarding every atom in a gas with photons of light and thereby removing the electrons from virtually every atom. Such a uniform, quiescent plasma can support high-quality plasma waves; it is a necessity if plasma waves more than several meters long are to be supported in both beat-wave and wake-field accelerators.

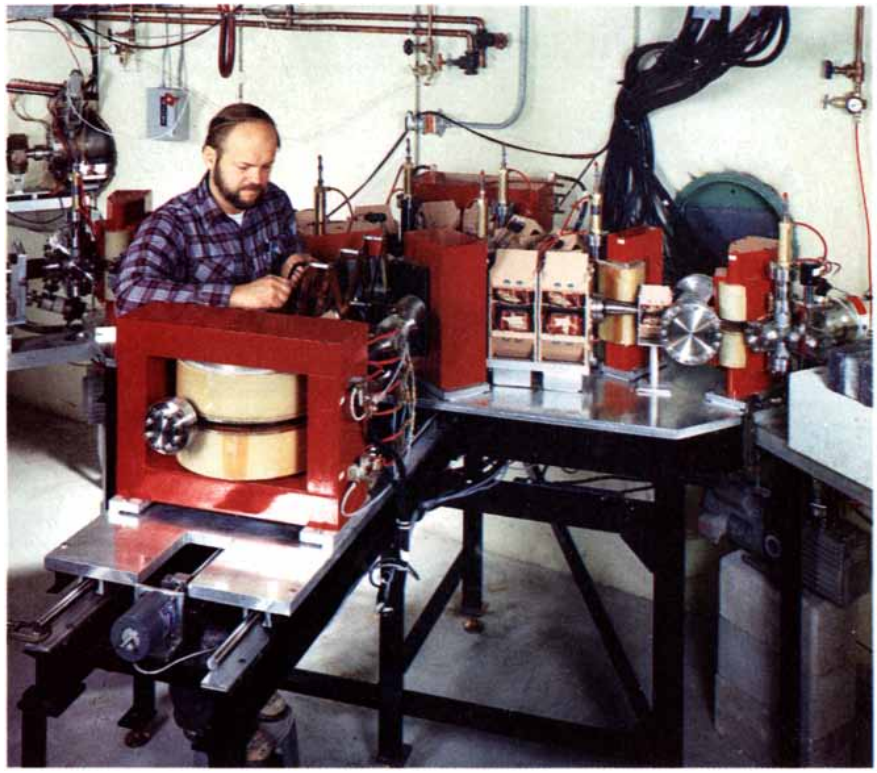
An experiment much like Joshi's was conducted by François Martin and his group at the University of Quebec. The experiment involved two plasmas, one to provide a pulse of accelerated electrons and the other to accelerate them further by the beat-wave process. Electrons were initially accelerated to more than .5 MeV by a process similar to Joshi's first experiment. A laser beam struck a solid target, creating a plasma of high electron density and temperature. The most energetic electrons were extracted and injected into the second plasma, where they were traveling fast enough to be further accelerated by the beat-wave method. When the workers analyzed the electrons coming from the beat-wave region, they recorded energies as high as 2 MeV. Since the beat-wave region was only 1.5 millimeters long, accelerating fields of approximately 10 million volts per centimeter must have been produced.

A number of additional investigations are now under way in the U.S., Canada, England, France and Japan. Although promising electric fields have been produced, more research is necessary to determine whether plasma particle accelerators can compete with existing accelerators in energy, intensity and quality. Surpassing the awesome power of the Superconducting Supercollider and the impressive record of accelerator technologies developed over the past 50 years remains a major challenge for plasma physicists.

In addition to the original goal of devising new means for accelerating particles, the ongoing research holds promise for several other areas of study. It may contribute to an understanding of particle-acceleration processes in nature. Pulsars, for example, are thought to generate intense waves in plasmas that accelerate particles to extremely high energies in a relatively short distance.

Further research into beat-wave accelerators could lead to new sources of X rays and gamma rays. Electrons accelerated by a beat wave could be directed into tungsten foils, which would then emit X rays. Electrons could also be injected into a plasma wave perpendicular to the direction of the electric field. These electrons would then be accelerated transversely to their direction of motion, causing them to radiate X rays and gamma rays.

Plasma-wave technologies will also provide new opportunities to study the interaction of electromagnetic fields with plasmas. Plasma waves can generate tremendous electric fields, and similar techniques may yield magnetic fields of similar scale. As in many scientific endeavors, the most important results of studies on plasma-acceleration may be totally unexpected.



WAKE-FIELD ACCELERATOR at the Argonne National Laboratory focused a bunch of many electrons into a plasma to generate plasma waves. Immediately after the bunch a group of a few electrons was injected into the plasma to determine the apparatus's accelerating capabilities: about 50,000 volts per centimeter of plasma.

FURTHER READING

LASER ELECTRON ACCELERATOR. T. Tajima and J. M. Dawson in *Physical Review Letters,* Vol. 43, No. 4, pages 267–270; July 23, 1979.

ULTRAHIGH GRADIENT PARTICLE ACCELERATION BY INTENSE LASER-DRIVEN PLASMA DENSITY WAVES. C. Joshi, W. B. Mori, T. Katsouleas, J. M. Dawson, J. M. Kindel and D. W. Forslund in *Nature,* Vol. 311, No. 5986, pages 525–529; October 11, 1984.

PLASMA-BASED HIGH-ENERGY ACCELERATORS. Edited by T. Katsouleas in *IEEE Transactions on Plasma Science,* Vol. PS-15, No. 2; February, 1987.

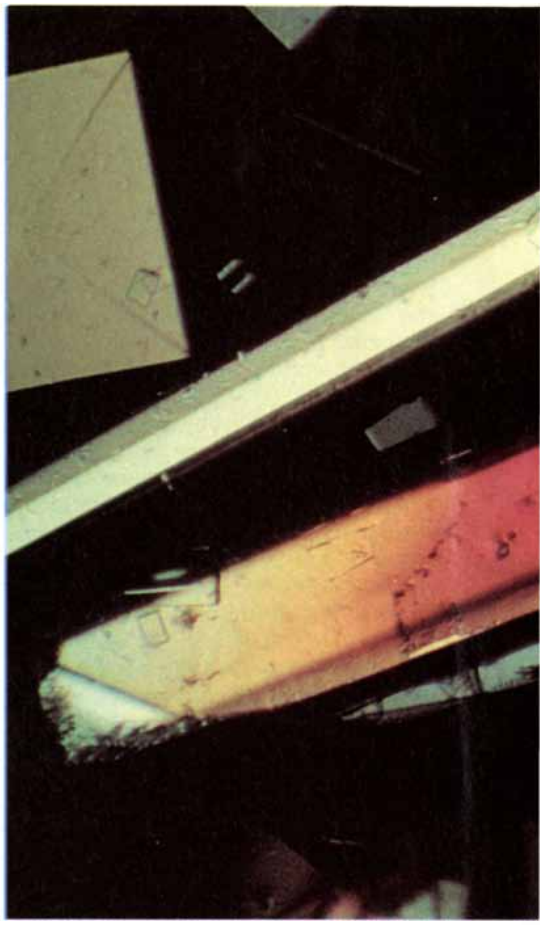
LASER ACCELERATORS. F. F. Chen in *Physics of Laser Plasma,* edited by A. Rubenchik and S. Witkowski. Elsevier Science Publishing Co., 1989.

PLASMA ACCELERATION OF PARTICLE BEAMS. T. Katsouleas and J. M. Dawson in *Physics of Particle Accelerators,* in press.

Macromolecular Crystals

The growth of crystals is now the key to deducing the structure of large molecules

by Alexander McPherson



With the advent of genetic engineering and recombinant-DNA technology, biochemists and molecular biologists have achieved new control over the synthesis of molecules that mediate and catalyze the chemical reactions that are life. In parallel with this novel synthetic capability, powerful analytic methods have been developed that can probe complicated macromolecules—large molecules such as proteins and nucleic acids—as well as more conventional chemical compounds. Principal among these methods is X-ray diffraction analysis, which extracts molecular images from the patterns created when crystals scatter X rays.

X-ray crystallography is one of the best methods for analyzing molecular structure in three dimensions at the atomic level [see "The Three-dimensional Structure of a Protein Molecule," by John C. Kendrew, SCIENTIFIC AMERICAN, December, 1961, and "The Hemoglobin Molecule," by M. F. Perutz, SCIENTIFIC AMERICAN, November, 1964]. The technique can also visualize macromolecular interactions with other molecules. With X-ray crystallography it is possible to explore, for example, what takes place when an enzyme binds a substrate molecule to catalyze a chemical reaction or when an antibody forms a complex with an antigen molecule.

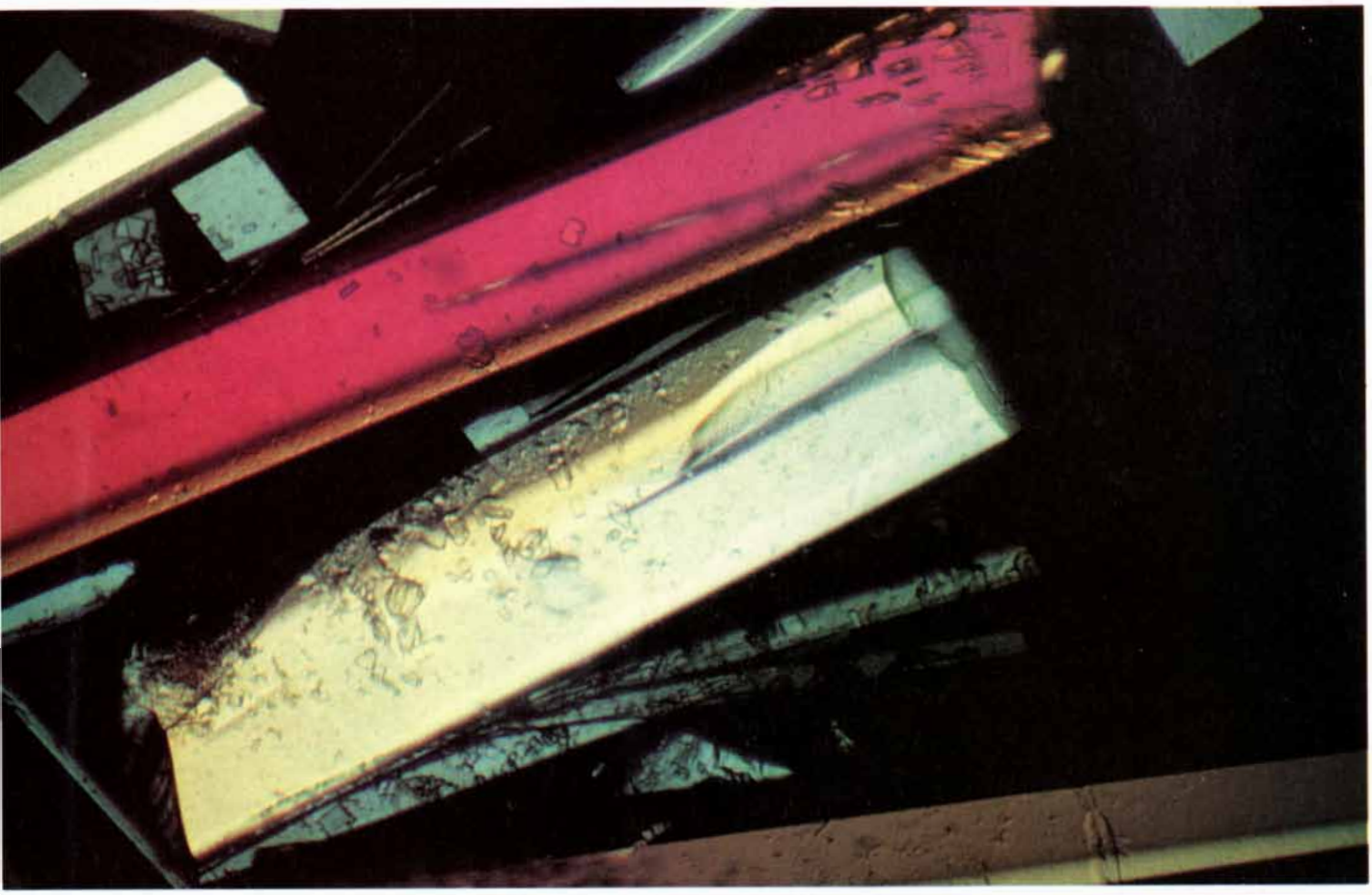
Since X-ray crystallography was introduced in 1913, advances in X-ray sources, detection devices and computing speed and precision have drastically reduced the amount of time it takes to obtain a crystallographic image. Structures of conventional molecules, which once took years to construct, can now be determined in one or two weeks. The structure of a macromolecule can take no more than a few months' time to solve, refine and display in three dimensions. If a protein's structure has already been solved, the details of its complexes with other molecules can often be determined in only a few days.

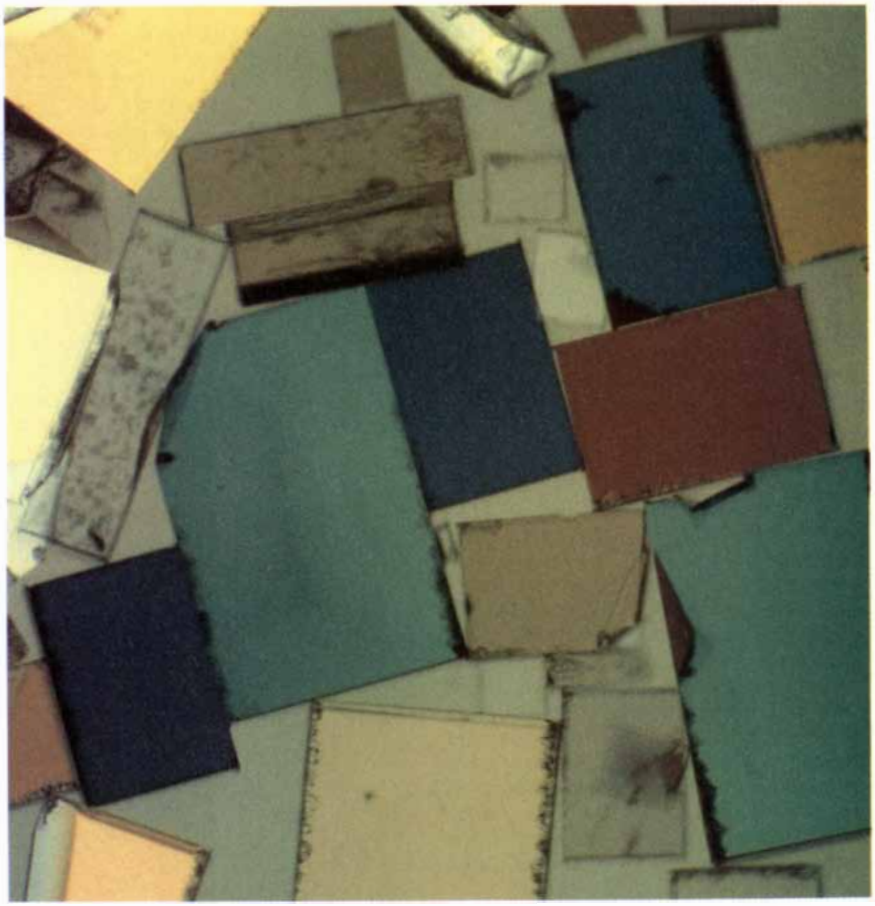
Indeed, the limiting factor in X-ray crystallography is no longer the technique itself but the availability of macromolecular crystals. Only a small fraction of the thousands of known proteins have ever been crystallized. Most proteins and nucleic acids do not crystallize readily, and the crystals subjected to X-ray diffraction analysis must be virtually flawless. Consequently biochemists have had to confine their studies to the molecules they are able to crystallize, enzymes exemplified by lysozyme from chicken eggs, pepsin from stomach juices and hemoglobin from blood.

Although the first protein crystals were grown more than a century ago, the process of crystal growth has been largely ignored. Before the invention of X-ray diffraction, crystal growth (which occurs with the exclusion of contaminants) served to demonstrate the purity of a preparation. The practice was considered to be as much an art as a science. Even today there are not many people trained in the procedures and methods of growing macromolecular crystals, and there are still only a few empirical approaches.

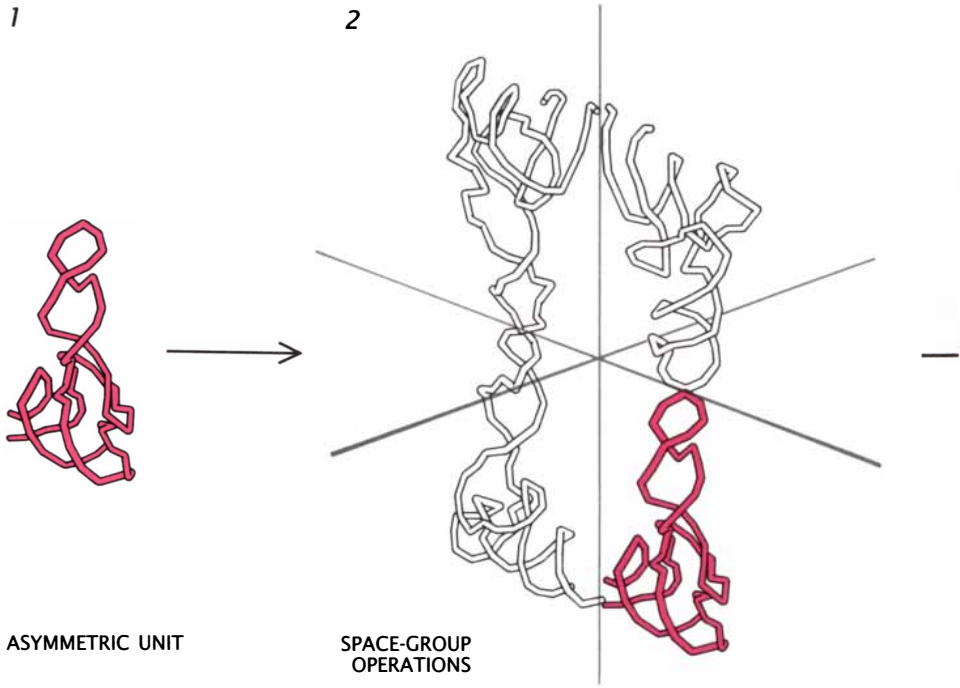
Failure to obtain crystals of certain proteins and nucleic acids for visualization hinders basic understanding of natural macromolecular interactions. It also frustrates the applied efforts of the biotechnology industry. The potential and scope of that industry, equipped with the tools to modify native molecules and create new ones, is staggering. The ability to procure a broader spectrum of macromolecular images would benefit the industry tremendously. Enzymes used for baking, brewing and food processing could be engineered to alter solubility, heat stability or acid tolerance if the critical parts of the enzymes could be identified; the nutritional value of cereals

ALEXANDER MCPHERSON is chairman of the department of biochemistry at the University of California, Riverside, and founder of Cryschem Inc., a biotechnology company that specializes in the crystallization and structural analysis of proteins. His interest in the structure of macromolecules dates to his graduate work at Purdue University, where he helped to solve the structure of the enzyme lactate dehydrogenase. After receiving his Ph.D. there in 1970, he joined a research team at the Massachusetts Institute of Technology that defined the structure of the nucleic acid transfer RNA. McPherson has written a book on protein crystals and coorganized two international conferences on the growth of macromolecular crystals.





PLANT PROTEINS concanavalin B **(top)** ***and vicilin*** **(right)** ***are two of the few proteins known to form crystals readily. The crystals are viewed under polarized light to reveal any inhomogeneities that would disrupt an X-ray analysis. Protein crystals and crystals of nucleic acids such as DNA are much more difficult to grow than crystals of simple compounds such as table salt. But macromolecules must be crystallized if they are to be amenable to X-ray crystallography, one of the most powerful methods available for solving molecular structure.***



CRYSTALLINE STRUCTURE can be considered a unique distribution of an asymmetric unit* (1)*, in this case represented by the nucleic acid transfer RNA. Several asymmetric units are arranged in a symmetrical configuration defined by a set of geometric operations referred to as a space group* (2)*. The arrangement of molecules specified by the space group can be enclosed in an imaginary container of minimal volume called a unit cell* (3)*; repetition of the unit cell at regular intervals gives rise to a crystal* (4)*.

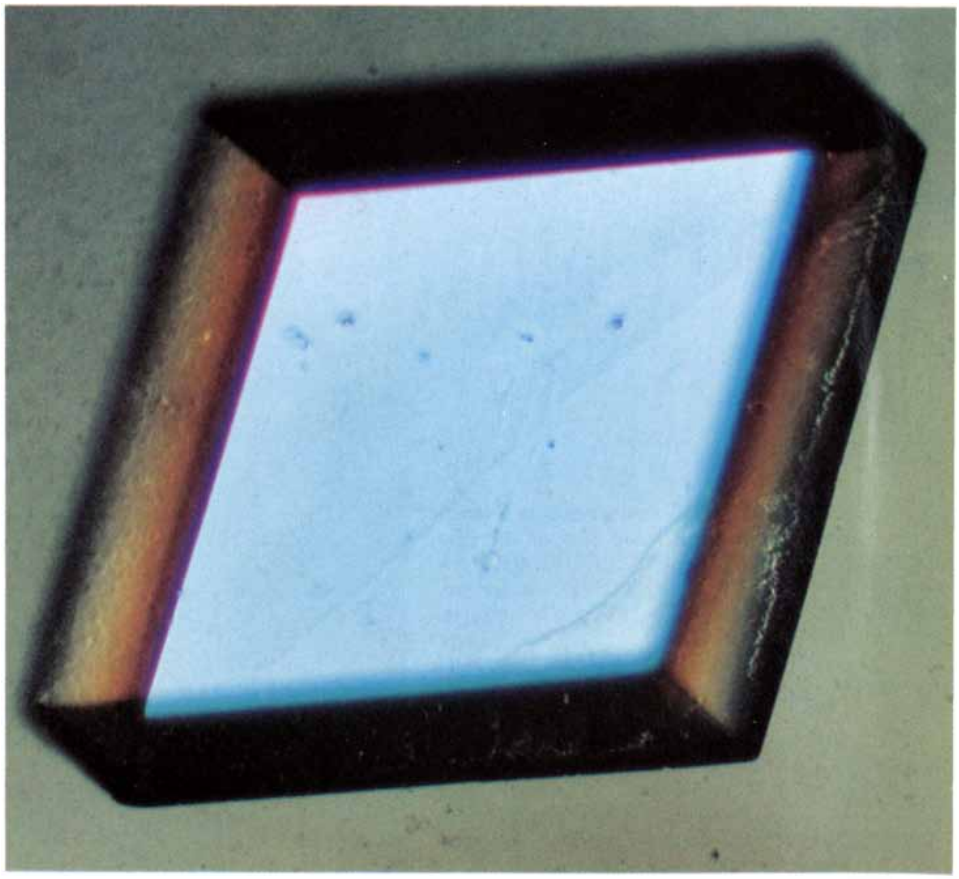
and beans and the resistance of agricultural crops to disease and harsh environmental conditions could be improved. Drugs that interact with proteins and nucleic acids could be designed rationally and systematically if the interactions between a drug and its target molecule could be visualized clearly.

The key to realizing the promise of biotechnology, then, lies in the development of new and reliable techniques for obtaining suitable protein and nucleic acid crystals for X-ray crystallography. If crystallography is ever to be fully exploited, the physical and chemical principles of macromolecular crystallization must be given greater attention.

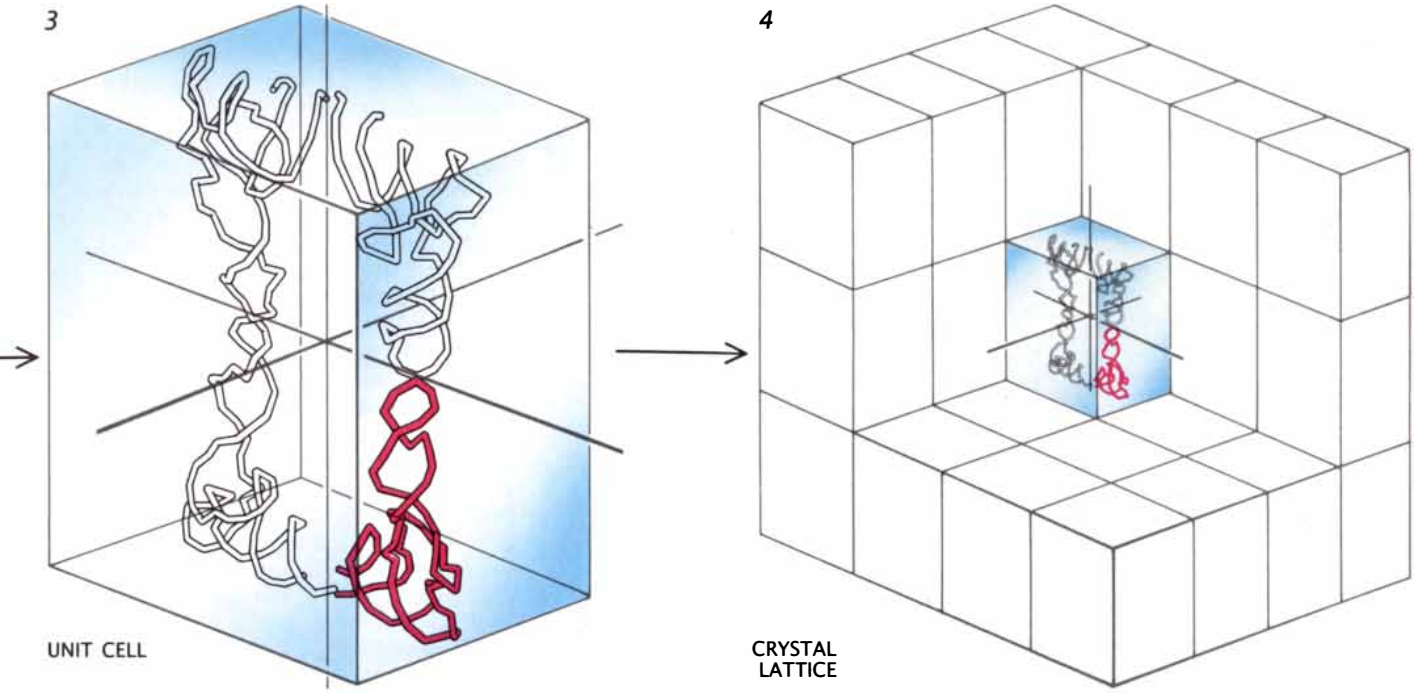
Crystals of any kind, whether they are crystals of table salt or of the most complicated protein, are ordered arrays of molecules characterized by a unique set of parameters. The parameters describe the distribution of atoms in a repeating "unit cell" and the arrangement of the unit cells that make up the crystal.

The fundamental component of the crystal, the so-called asymmetric unit, is usually a single molecule of the substance forming the crystal. Several asymmetric units pack together in a symmetrical configuration the mathematical description of which is called a space group. There are 230 possible space groups allowed in a crystal; only 65 space groups are allowed in protein crystals, however, because the amino acids that make up protein molecules impose certain symmetry constraints. Often one type of molecule can assume several different symmetrical arrangements described by a variety of space groups.

The unit cell is an imaginary box of minimal volume that encloses the set of asymmetric units defined by the space group. Unit cells have six possi-

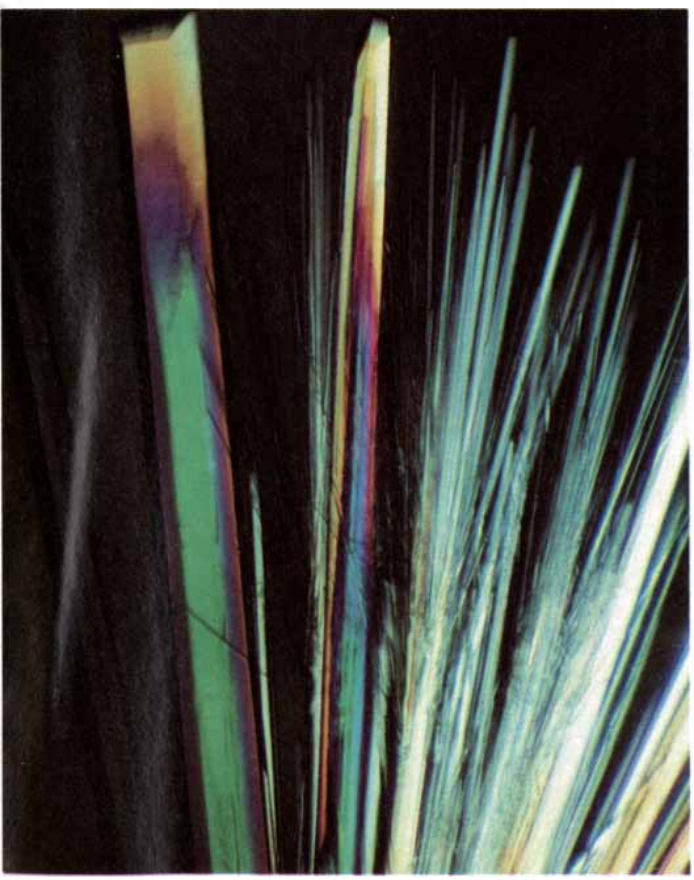


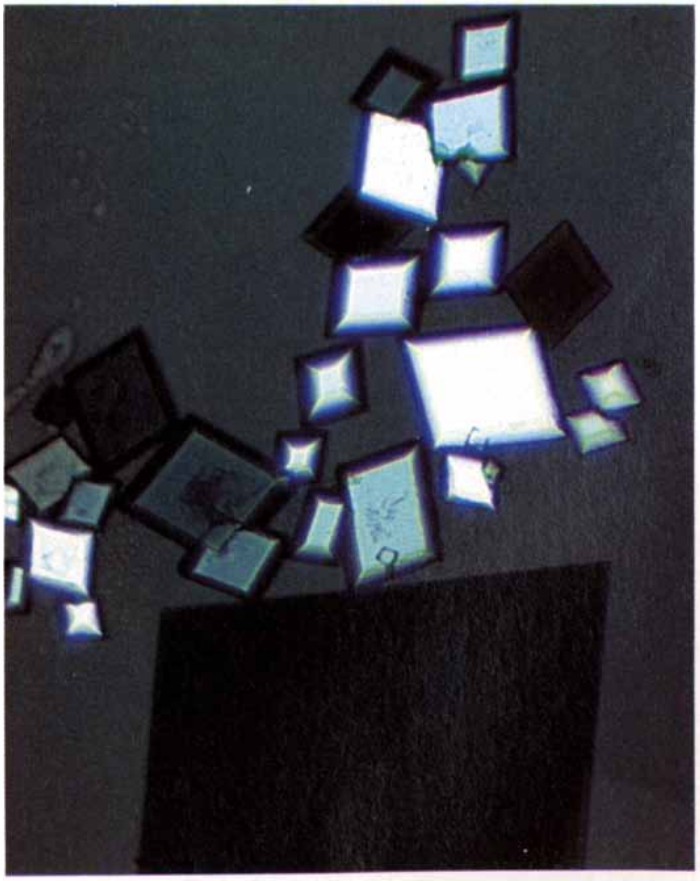
TYPICAL MORPHOLOGIES of crystals, seen through a light microscope under polarized light, include a rhombohedral prism that measures more than one millimeter along each edge* (left)*. The crystal was made from the plant protein canavalin. When canavalin crystals are grown under slightly different conditions of acidity and temperature, a spray of hexagonal prismatic rods results* (middle)*. So-called monoclinic plates, which are flatter than prisms, are also common; these plates were grown from a fruit-mold enzyme* (right)*.

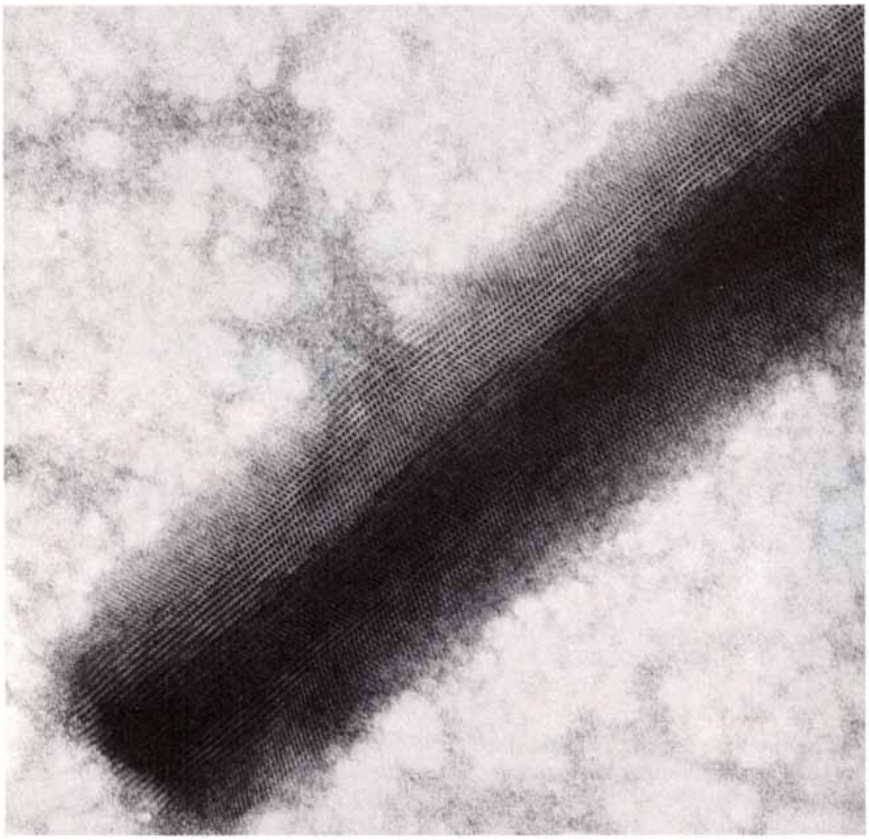


ble shapes; the lengths of the cell's edges vary between 50 and 150 ten-billionths of a meter for most protein crystals. Crystalline structure can be represented as a repetition of the unit cell at periodic or regular intervals.

The spontaneous self-arrangement of molecules in an ordered lattice might seem to violate the second law of thermodynamics, which maintains that all matter tends toward a state of maximum entropy, or disorder. It turns out that the formation of stable chemical bonds that accompanies crystallization drives the process: the release of energy caused by bond formation compensates for the order imposed on the system by crystallization. It can be demonstrated by mathematical and geometrical arguments, and it is intuitively satisfying as well,







***POROUS STRUCTURE** of macromolecular crystals is evident in an electron micrograph of a crystal of a bacterial enzyme* (above). *The dark channels are clearer in the enlargement* (left). *The channels make the crystals extremely fragile, but they also provide paths for the diffusion of small molecules. In the photograph below, color is supplied not by polarized light but by a red drug that has penetrated the solvent channels in the crystal and has bound to the molecules forming the lattice.*

that the maximum energy is released when the bonds are formed in a symmetrical and periodically repetitive manner in the solid state.

To be sure, crystal growth is not energetically favorable in all circumstances. Crystallization can take place only when attractive interactions between the molecules are maximized and repulsive interactions are minimized. In practical terms, these conditions apply in so-called supersaturated, or extremely concentrated, solutions in which there is not enough water to maintain hydration or to shield the molecules from one another completely. To take one of the simplest examples, sea salt crystallizes from tidal pools when the water in the pool evaporates slowly.

Unfortunately the crystallization of macromolecules is not that simple. The number of bonds a macromolecule forms with its neighbors in a crystal is much smaller in proportion to molecular weight than the number formed in crystals of smaller molecules such as salt. Since it is the lattice interactions that maintain the integrity of a crystal, macromolecular crystals are much more delicate than the crystals of smaller molecules.

In addition, individual proteins and nucleic acids readily lose their natural structure under mildly hostile conditions, and so the only conditions that can support the crystal growth of a given macromolecule are those that cause no perturbation of the molecule's properties. Protein and nucleic acid crystals are grown from solutions, known as mother liquors, that maintain a tolerable range of acidity, temperature and ionic strength. If the crystals become dehydrated, they collapse, and so they must always be bathed in mother liquor.

Indeed, crystals of macromolecules can be thought of as ordered gels having extensive solvent channels and cavities. Whereas crystals of smaller molecules such as salt contain little or no solvent, macromolecular crystals are approximately 50 percent solvent (the figure may vary from 30 to 90 percent). Because of the large interstitial spaces and the weak lattice forces, the positions of the molecules in a macromolecular crystal may not be perfectly equivalent, that is, they may vary slightly from one unit cell to another.

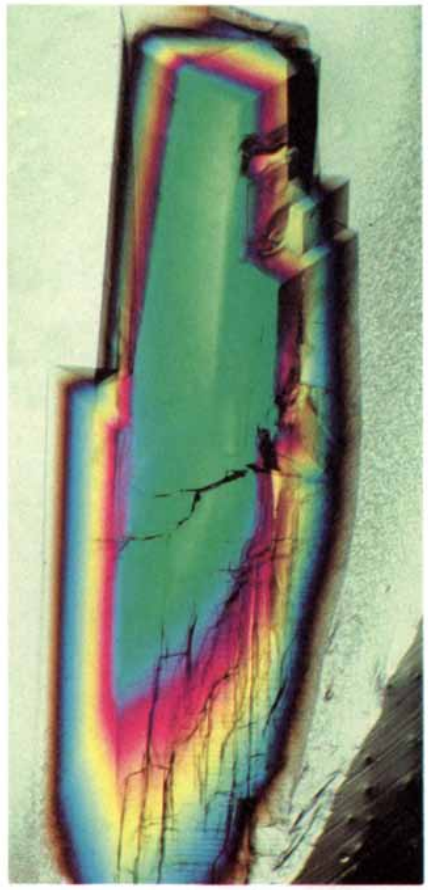
The porosity of protein and nucleic acid crystals has ramifications for X-ray diffraction as well as for the crystallization procedure. The level of detail to which atomic positions can be determined by diffraction analysis

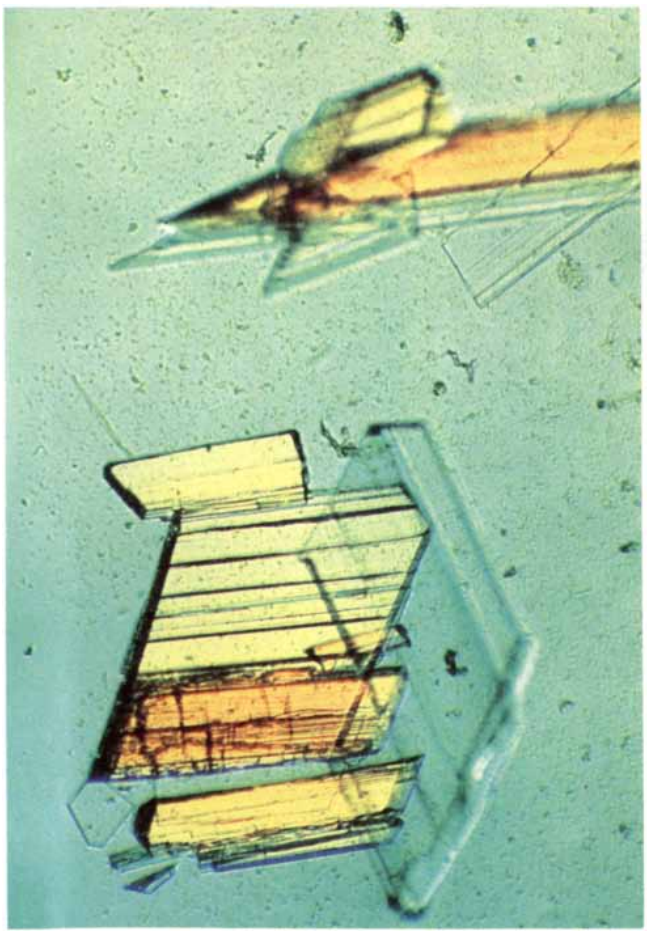
corresponds closely with the degree of crystalline order; any inhomogeneity in a crystal can compromise diffraction patterns. Macromolecular crystals are likely to develop inhomogeneities because of their weak lattice forces. Consequently, whereas the crystals of small molecules often diffract almost to their theoretical limit of resolution, protein and nucleic acid crystals are generally characterized by very limited diffraction patterns.

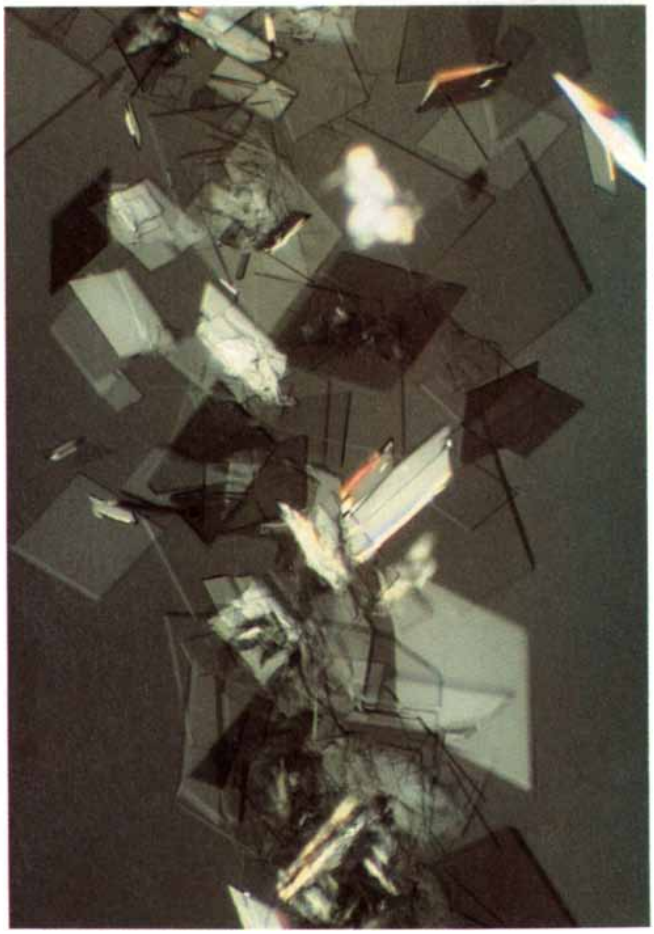
The high solvent content of macromolecular crystals is not, however, an entirely negative attribute. Because of the hydration, molecular structures are virtually unchanged from those that exist in vivo. Hence the biochemist can observe many properties of the protein or nucleic acid as they would appear in their natural context. Moreover, the solvent channels in macromolecules are large enough to admit compounds that interact with proteins and nucleic acids: ions, enzymatic substrates, inhibitors, drugs and so on. Consequently the biochemist can study such interactions by suffusing the mother liquor—and thereby the crystal itself—with the compound in question.

In growing crystals of macromolecules the operative term is "gradual." Any abrupt change in saturation will result in the precipitation of an amorphous solid rather than in crystallization. The standard approach to macromolecular crystallization is to equilibrate a highly concentrated solution with a precipitating agent; if it is introduced slowly enough—by diffusion of a vapor, for instance—the agent will induce crystallization rather than precipitation.

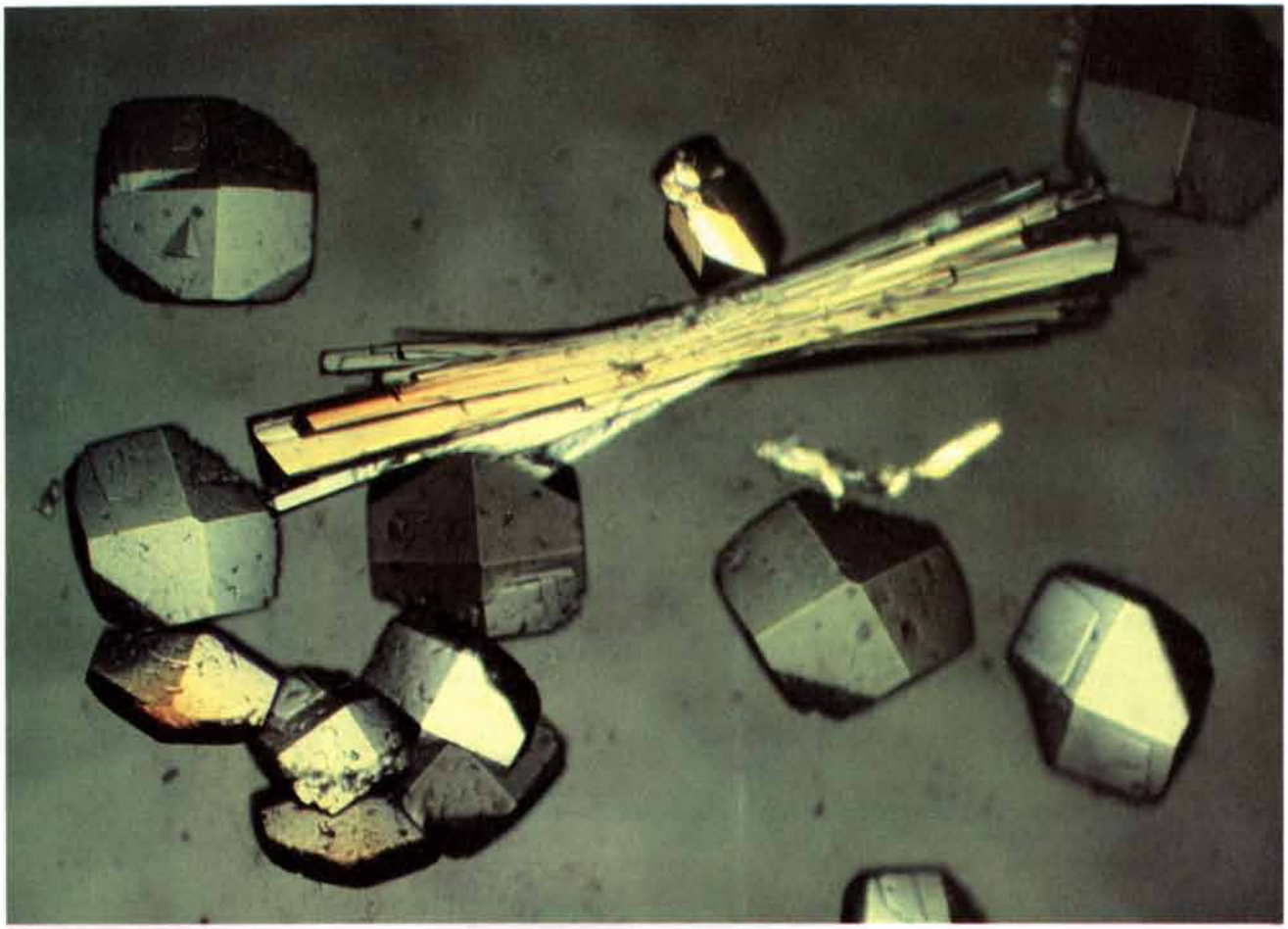
Precipitating agents include salts such as ammonium sulfate, organic solvents such as ethanol and soluble synthetic polymers such as polyethylene glycol. The three agents act in slightly different ways, but all encourage the interaction of macromolecules in solution. Salts compete with the

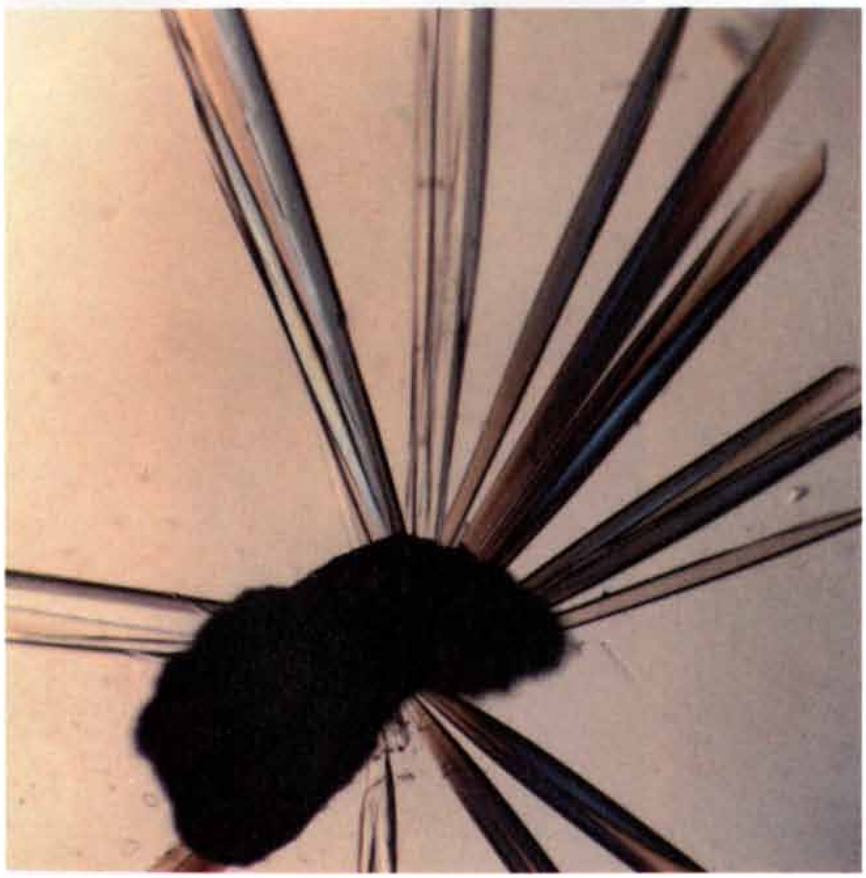






***FLAWED CRYSTALS** display different kinds of imperfections. Cracks and steps resulting from sporadic growth mar the prism at the right. Distortions and surface striations in the crystals at the lower left reflect underlying inhomogeneities in the growth and ordering of the unit cells. The rhombic plates at the lower right are too thin and too small to be used in X-ray crystallographic analysis.*





***MINERAL SUBSTRATE** promotes the formation of protein crystals by providing an ordered base for nucleation. Here a piece of magnetite serves as a platform for crystals of concanavalin B. The author hopes to use mineral substrates to specify forms of crystal growth and to encourage the nucleation of macromolecules that are difficult to crystallize.*

POLYMORPHIC FORMS of the same protein often arise simultaneously during crystal growth in a single "mother liquor." The different forms suggest that a protein molecule can exist in several different conformations under any given set of conditions; each conformation could utilize a different space group and give rise to a different crystal.

dissolved macromolecules for water, thereby encouraging those molecules to interact with one another rather than with water. Organic solvents increase the strength of the electrostatic fields that attract protein molecules to one another. Polymers induce both of those effects and also form a network with water that tends to exclude macromolecules.

Precipitating agents can also serve to bring a highly concentrated solution to a point just below supersaturation; then supersaturation is achieved by gradually modifying some physical property of the solution, such as the acidity or the temperature, in order to reduce the solubility of the protein or nucleic acid. This procedure is analogous to saturating boiling water with sugar and then cooling it to produce rock candy.

Crystallization by any of the current methods is unpredictable. Every macromolecule is a unique three-dimensional structure with distinctive surface characteristics, and every macromolecule has unique physical and chemical properties. For this reason lessons learned by investigating one protein or nucleic acid are seldom applicable to others.

To complicate matters, proteins and nucleic acids seem to be fairly dynamic entities: even the same molecule can exist in several different conformations under identical conditions or in a single solution. Consequently many proteins and nucleic acids grow in a variety of unit cells, resulting in the so-called polymorphic forms seen in single batches of crystals. Often the diffraction properties of one crystalline form are more desirable than those of the alternative forms; the inability to dictate the morphology a particular macromolecular crystal assumes can be quite frustrating.

Recently my colleague Paul J. Shlichta of the Jet Propulsion Laboratory and I have attempted to control the course of crystal growth by seeding protein solutions with fragments of minerals. We were inspired by the example of semiconductor manufacturing, in which the molecular structure of a silicon substrate is used to define the structure of a superposed layer. Our results were encouraging. We found that mineral "platforms" do indeed help the crystals to nucleate; they can also help to confine crystal growth to one morphology. In most cases, however, we have not been able to determine how a particular protein-crystal morphology relates to a mineral substrate's atomic structure.

I have also been collaborating with biochemists across the U.S. to design experiments for crystal growth under microgravity conditions. So far protein-crystallization experiments have flown on five space-shuttle flights, and there are plans for several more this year. We want to find out whether macromolecular crystals grow more uniformly in the absence of the turbulence caused by gravity. The results we have gleaned to date are encouraging but too limited to provide definitive evidence of an advantage.

Other significant factors can contribute to or detract from successful crystal growth. Contaminants can prevent growth or give rise to flaws in crystals, and so preparations of proteins and nucleic acids as well as precipitating agents must be quite pure. Preparations must also be standardized if results are to be accurately reproduced. Furthermore, the nucleation of crystals is still a phenomenon governed by probability rather than predetermination. Hence the crystallization of macromolecules remains a puzzle and a challenge—and an opportunity as well.

FURTHER READING

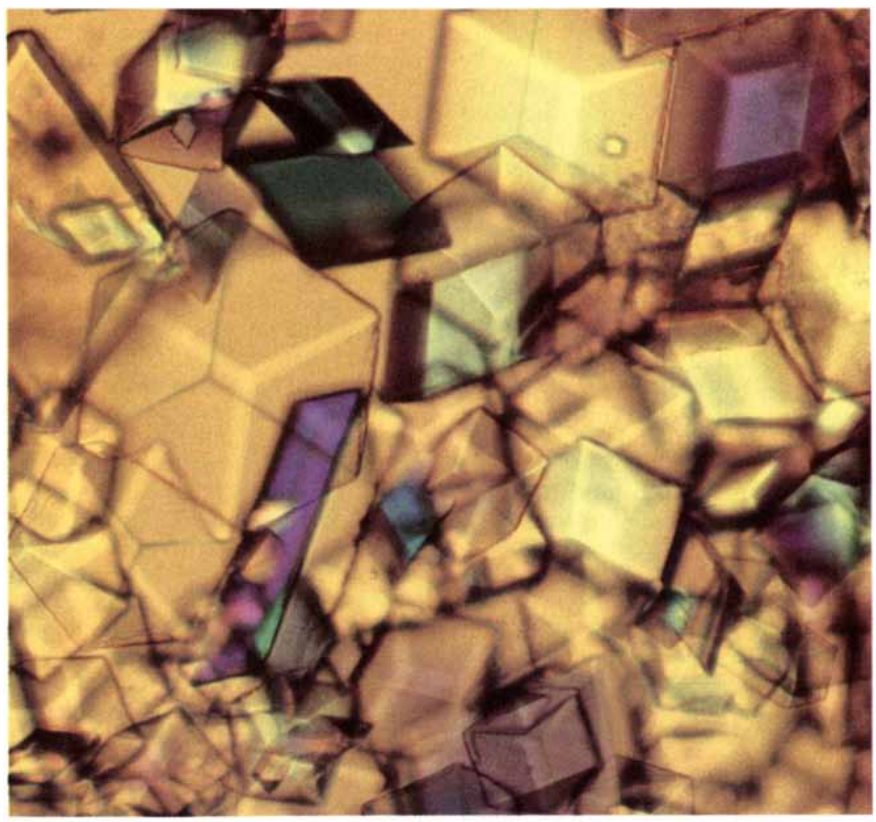
THE USE OF X-RAY DIFFRACTION IN THE STUDY OF PROTEIN AND NUCLEIC ACID STRUCTURE. K. C. Holmes and D. M. Blow. John Wiley & Sons, Inc., 1966.

THE ANATOMY AND TAXONOMY OF PROTEIN STRUCTURES. Jane S. Richardson in *Advances in Protein Chemistry,* Vol. 34, pages 167–339; 1981.

THE PREPARATION AND ANALYSIS OF PROTEIN CRYSTALS. Alexander McPherson. John Wiley & Sons, Inc., 1982.

NEW DIRECTIONS IN PROTEIN CRYSTAL GROWTH. L. J. De Lucas and C. E. Bugg in *Trends in Biotechnology,* Vol. 5, pages 188–192; 1987.

USEFUL PRINCIPLES FOR THE CRYSTALLIZATION OF PROTEINS. Alexander McPherson in *The Crystallization of Membrane Proteins,* edited by Hartmut Michel. CRC Press, in press.



CRYSTALS FROM SPACE were grown on board the space shuttle* Discovery *during its flight last September. The crystals are part of an ongoing series of experiments in microgravity. The absence of gravity-driven convection and sedimentation in the mother liquor may be responsible for the generally perfect growth of such crystals.

Competitive Climate

Industry leaders look to the Government for a new era

Even as fledgling Bush appointees unpack their coffee mugs and tennis rackets, industry is calling on the Administration to address a formidable set of economic problems. A spiraling national deficit, low rates of saving and investment and low rates of productivity growth are just a few of them. At the top of many business leaders' list is international competitiveness. The Government must drastically improve the business climate if U.S. companies are to survive tough global competition, say many of the nation's leading executives.

"The war we're in is an economic one, not a military one," declares Robert N. Noyce, chief executive officer of SEMATECH, a Government-industry research consortium on semiconductor-manufacturing technology. The hurdles companies face—improving their technology, increasing their investments in equipment and facilities, training their work force—have become so overwhelming "that virtually no U.S. company left to its own resources will be successful long term," says Thomas J. Murrin, chairman of the Defense Manufacturing Board.

Business leaders are no longer asking the Government to protect individual industries such as steel or semiconductors. They are calling for it to tackle problems faced by all businesses—the high cost of capital, the expense and risk of research and a poorly educated work force. International competitors, industrialists say, are not shackled by these problems. It is time the Government provided U.S. companies with the same strengths, says John A. Young, chief executive officer of Hewlett-Packard.

Industrialists hasten to add that they are not asking the Government to establish an "industrial policy" that would target areas for special subsidies or artificially raise or lower prices. "We need a partnership approach," Young says.

In such a partnership the Government would employ macroeconomic policies to foster a strong, private-sector technology base. Most people, including Erich Bloch, the director of the National Science Foundation, believe such policies could be developed through existing Government departments. Industry, for its part, would have to take better advantage of the new technology that emerged.

Teaming to compete, changing orbits, thrifty rats

The financial changes business leaders hope for are far-reaching. "We need lower interest rates, R&D tax credits, lower taxes on capital gains, low inflation. The cost of capital is too high and the savings rate too low," says Simon Ramo, cofounder of TRW. The cost of capital in the U.S. ranges between 8 and 10 percent, much higher than the 3 to 4 percent common in Japan, says Frank Press, president of the National Academy of Sciences. The higher the cost of capital, the greater the pressure to produce tangible results. Given the costs, businesses are behaving rationally by trimming their spending on research, concludes Roland W. Schmitt, president of the Rensselaer Polytechnic Institute. Observers think that by encouraging savings, cutting the deficit and consequently lowering the price of capital, the Government can change industry's attitude toward long-term investment.

Several departments are already thinking hard about their roles. Officials at the Department of the Treasury point out that late last year the newly designated secretary, Nicholas F. Brady, formed a departmental study group to examine how tax incentives for research and lower capital-gains taxes might spur investment and long-term growth. Treasury Department officials say their first move will be to encourage pension-fund managers, who typically control large blocks of stock, to hold on to shares longer than is now typical; the move is aimed at lengthening the duration of investment in companies.

The Department of Defense has also indicated that it might push for a larger economic role. In October the Defense Science Board, an independent advisory panel, raised eyebrows when it recommended that the Defense Department "extend its view and influence beyond its own acquisition policy to reach tax, trade, and other economic policies which affect the industrial base." Among other steps, the board suggested that the secretary of defense sit on the President's Economic Policy Council.

So far the report remains a collection of recommendations rather than a map for action. Industry has reacted cautiously. "The DoD ought to be applauded for the courage to endorse those suggestions" and for recognizing that financial issues are at the heart of American companies' ability to make successful products, Murrin says. The Defense Department may not be the most suitable organization to guide the Government on economic issues, he adds, but at least it has stepped forward.

By controlling almost 70 percent of the more than $60 billion U.S. budget for research and development, Defense has traditionally wielded more influence over industrial research than any other department; now, however, the Department of Commerce is hoping to take a more visible role. Last December former commerce secretary C. William Verity established a position for an undersecretary for technology and reorganized several existing Commerce divisions, including the National Bureau of Standards, under the banner of the National Institute of Standards and Technology. (In the fiscal 1990 budget submitted to Congress by outgoing President Reagan, no additional funding was earmarked for the new institute, however.)

Schmitt hopes that Commerce's new stance will catalyze more research cooperation among companies—a trend he thinks will accelerate anyway as antitrust laws are loosened and the shortage of American researchers becomes more acute. William L. Keefauver, vice-president and general counsel at the AT&T Bell Laboratories, agrees with the need for further research collaboration. "The R&D costs [of the most advanced projects] are outstripping the capacity of even AT&T and IBM to do it themselves," he says.

Even as industrial leaders endorse the idea of research consortiums, they concede that the cooperative projects have yet to prove their merit. Although consortium projects are supposed to

Flying low-level missions at night or in adverse weather will be easier thanks to a new night-vision navigation system from Hughes Aircraft Company. The Thermal Imaging Navigation Set (TINS), designed for the U.S. Navy's F/A-18 Hornet, utilizes a thermal imaging sensor that provides pilots with a TV-like image of the terrain ahead. The image is projected onto a heads-up display for viewing at night or in poor visibility. The TINS system is pod mounted in a fixed, forward-staring position, and can be reconfigured into different pods for a variety of aircraft.

A family of display consoles, converters, and switchboards, all designed to process and display information in short response times, successfully completed sea shock testing. Considered the standard display for all U.S. Navy surface ships, Hughes' AN/UYQ-21 display systems are being installed aboard the Navy's Aegis cruisers and destroyers. The system is designed to be used for anti-submarine warfare, command and control, and fire control. Since the UYQ-21 is a family of displays, the combat, sensor, and fire control systems can be tailored for different ship classes. The system also will be used on aircraft carriers, assault ships, and other Navy ships.

A new amplifying device can operate at much higher frequencies, and with lower noise, than traditional field-effect transistors. The High Electron Mobility Transistor (HEMT) device implemented in a new material system, pioneered and developed by Hughes, is fabricated by using indium phosphide as a substrate with gallium indium arsenide and aluminum indium arsenide grown onto it, one layer at a time, using a process known as molecular beam epitaxy. In a HEMT device, the semiconductor material containing the impurities is separated from the region of charge-carrying electrons, allowing the electrons to move much faster, so the device can operate at much higher frequencies, with lower noise, than an ordinary transistor. Potential uses include ultra-high frequency communication systems, high-speed radar signal processing equipment and high-power millimeter-wave circuits.

A spacecraft that was sent to explore the planet Venus for eight months completed its first decade of space service. The Hughes-built Pioneer Venus Orbiter was originally slated to circle the planet for one complete Venus rotation, approximately 243 Earth days, and provide the first topographic maps of the planet. However, its unexpected longevity has given scientists detailed information not only about Venus but also about comets that have streaked within sight of the probe. Despite its 10 years, the orbiter shows little sign of wear and is expected to operate for another four years until its fuel runs out.

Hughes needs engineers, scientists, and programmers to forge new frontiers in aerospace radars, weapon control systems and avionics, airborne displays, aerovehicle data links, airborne countermeasures, and processors. Current openings are for people experienced in electrical/mechanical design, development, test and manufacture for systems engineering, project/program management, design of circuits and mechanisms to bring these to reality through the application of advanced manufacturing techniques. Send your resume to Hughes Aircraft Company, Radar Systems Group, Engineering Employment, Dept. S2, P.O. Box 92426, Los Angeles, CA 90009. Equal opportunity employer. U.S. citizenship required.

For more information write to: P.O. Box 45068, Los Angeles, CA 90045-0068

HUGHES

Subsidiary of GM Hughes Electronics

focus on "precompetitive" research projects, basic and applied research are often difficult to distinguish precisely. Noyce points out that unlike many consortiums, SEMATECH focuses on research that is easily shared: manufacturing expertise. Many chip makers already buy their manufacturing equipment from common suppliers. Companies making chips rely on proprietary designs, not manufacturing processes, to get ahead of their competition, Noyce says.

In addition to financial reforms and research initiatives, industry is counting on the Government to overhaul the education system. Improving primary and secondary schools as well as encouraging more students to pursue advanced degrees in mathematics and science should be a top national priority, Keefauver says. "A few CEO's have told me they could make a product that almost never fails," Press adds, "but the limiting factor is the quality of the work force." Young thinks the U.S. would be wise to boost its support of education, even if it is at the expense of large science projects such as the space station.

Industry leaders are also waiting for President Bush to make good on his campaign promise to upgrade the post of the science adviser to that of special assistant to the president. Ramo says he was disappointed that a science adviser was not named along with the Cabinet appointments.

As Young points out, government can create a favorable climate for renewed international competitiveness but industry must do the rest. "The primary vehicle for competing is the private sector," he says. "And not too many managements feel comfortable about the stewardship of their companies." Developing innovative technology or processes in consortiums such as SEMATECH may prove to be relatively easy for industry, Murrin says, compared with justifying the tremendous investment in new equipment and facilities needed to exploit the technology innovations. —*Elizabeth Corcoran*

Decaying Orbit

The U.S. and France may join forces in remote sensing

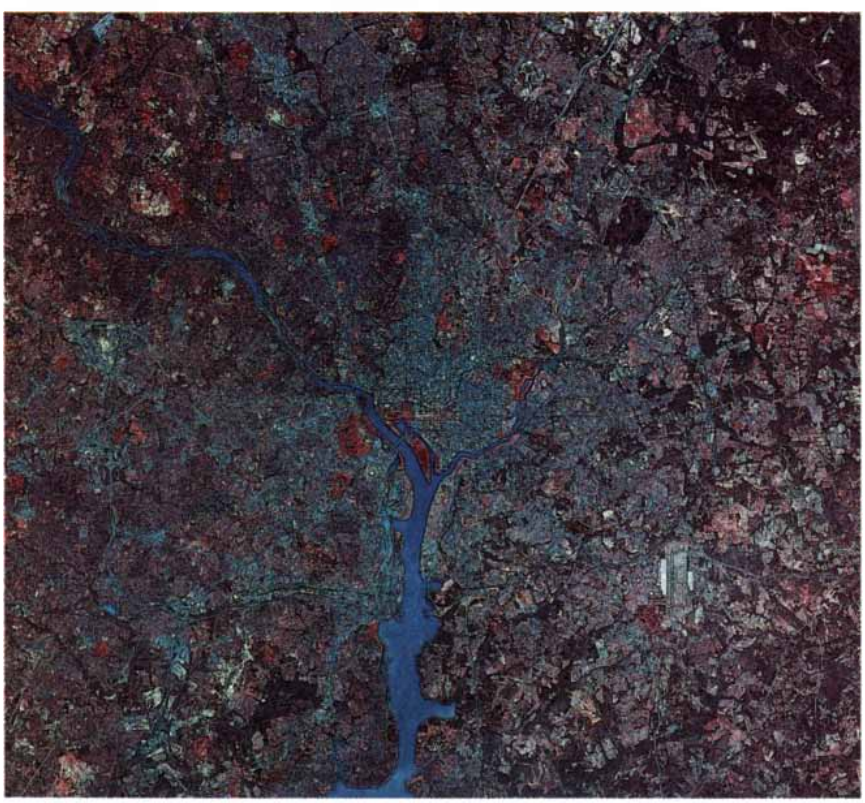
In 1972 the U.S. launched *Landsat 1,* one of the first civilian satellites devoted to sending back data and images of the earth itself. In the early 1980's the Government tried to turn the science it had pioneered into a profitable business of selling data to such users as oil companies and city planners. Now the Government appears to be conceding that the venture is too costly for the U.S. to go it alone.

In January, U.S. Government officials were talking with their counterparts in France about the possibility of merging the two countries' remote-sensing programs in the mid-1990's. A joint working group reviewing the feasibility of such a merger is slated to complete its work by June. Although remote sensing is actively managed in the U.S. by the Earth Observation Satellite Company (EOSAT) and in France by the company SPOT Image, the two governments own the satellites and raw data. Responsibility for the Landsat program falls under the jurisdiction of the National Oceanic and Atmospheric Administration (NOAA), a division of the Department of Commerce; in France the National Center for Space Studies (CNES) oversees SPOT satellites.

David S. Julyan, executive vice-president of SPOT Image's U.S. subsidiary, says that a U.S.-French partnership is likely to entail sharing the design, construction and launch costs for a single remote-sensing program. The first satellite would not orbit before 1996. Although officials are far from working out the specific details of a possible partnership, Julyan says he would expect it to be a "fifty-fifty deal." A joint venture between SPOT Image and a U.S. company would probably market the data; that company may not be EOSAT, however. As of January, EOSAT officials had not been invited to take part in the talks about a partnership.

EOSAT's woes follow several turbulent years. It contracted with the newly formed EOSAT in 1985 to begin commercializing the program. The Government promised to support the operations of *Landsat 4* and *Landsat 5* (launched in 1982 and 1984 respectively) and build two replacement satellites as well as a ground station for receiving data. EOSAT in turn would develop the marketplace for the data and images.

Four years later the scene looks much different. *Landsat 4* and *Landsat 5* are operating beyond their planned lifetimes. *Landsat 6,* originally scheduled to be put into orbit by the space shuttle in December, 1988, had to be redesigned to fit on board alternative launch vehicles after the *Challenger* disaster in January, 1986. Every year the Administration has cut its support for EOSAT, only to have Congress reinstate some part of the funding. "EOSAT



WASHINGTON, D.C., is shown from the vantage of the* SPOT 1 *satellite. The false colors highlight vegetation* (red), *water* (blue) *and buildings* (white to blue, depending on material). *Courtesy of the SPOT Image Corporation.

is not universally admired," says Paul A. Blanchard, president of Space Research & Management, "but they were given a bum deal."

As a result of the ups and downs in funding, EOSAT has cut back the amount of data it collects. But since customers studying regional changes often need data spanning months or even years, spotty archives can be a problem. "Ordinarily we'd like to collect as much data as possible to build a basic data set," says Peter M. P. Norris, executive vice-president of EOSAT. "Now we've changed to a management approach that emphasizes sales in the near term."

EOSAT is also facing fierce competition from SPOT Image, which has been getting high marks for its aggressive marketing. The French were latecomers to the remote-sensing business; *SPOT 1* was launched in 1986. Last year SPOT Image's worldwide sales revenues matched those of EOSAT for the first time. The French company has also enjoyed more consistent support from its government. Whereas NOAA and EOSAT must petition Congress for subsidies to cover the costs of satellite operations each year, SPOT Image pays annually negotiated royalty fees to CNES for the data it sells.

The royalties cover only part of the costs of the satellites and operations, Julyan says. But as long as the company is moving toward full commercialization and meeting its yearly goals, he says, the French government is willing to support the venture. "One of the biggest problems in commercializing space is that too many people are in too big a hurry," says John Egan, president of the Egan Group, a space-business consulting firm. "It will happen, but at its own pace."

That assessment is likely to be echoed in three as yet unreleased reports commissioned by Commerce and delivered last August; they are said to conclude that remote sensing will not become a commercially viable business in the near future.

Blanchard agrees that both U.S. and French remote-sensing companies will continue to need help from their governments in building and launching satellites at least until the year 2000. For instance, building and launching *Landsat 6,* expected to operate for five years, is likely to cost $200 million. EOSAT's 1988 revenues, in contrast, were $18 million. "If you attempt to amortize [space-related costs] by selling information to the user, then the program is doomed to failure," adds Murray Felsher, publisher of the *Washington Remote Sensing Letter.*

A remote-sensing company could be profitable, the consultants say, if it needed to pay only for such ground-based costs as processing the raw data and marketing the information. Although the company could afford to pay part of the satellite and launch costs, Egan says, a government would have to foot the rest of the bill in the near term.

In view of the high costs of business in space and the still small market for remote-sensing data, a U.S.-French partnership might well make business sense. But creating a part-government, part-corporate joint venture, in which the two countries have equal shares, is likely to be a herculean task with heavy political overtones. Finally, will the U.S. be able to support a joint project more consistently than it has supported EOSAT? —*E.C.*

An Economic Animal

Can rat behavior test economic theories?

Most rats interested in pursuing a career in the social sciences have to go into psychology; in a few universities, however, they can also take up economics.

For the past decade John H. Kagel, an economist at the University of Pittsburgh, and Raymond C. Battalio, of Texas A&M University, have employed rats and a few pigeons to test theories about individual choice. Since it is not clear whether human beings make optimal economic choices consciously, the investigators say, animal behavior can serve as a model for human choice and help to sort out conflicts in fundamental economic theory.

Rigorously proving an economic hypothesis is no mean task. Data are incomplete or flawed. Economic models are usually so simplified—positing that consumers have only two goods, for instance—that it is difficult to use real-world observations to support a hypothesis. As early as the 1940's a few economists began trying to devise empirical tests. Many experiments use groups of students (often graduate students in economics) to simulate auctions or markets. But Kagel and Battalio observed that many experiments in psychology relied on animals. If rats and pigeons could test psychological theories, they reasoned, why not economic ones?

Consider, for example, a recent experiment that tested the existence of "Giffen goods," a very rare contradiction to the general axiom that rising prices provoke a fall in demand. The Giffen good every freshman economics student learns about is the Irish potato: during the famine in Ireland in the 1840's, amid rising prices for potatoes, the Irish reportedly bought more spuds. (They were still the cheapest food available.) Conversely, when potato prices finally began to fall, the Irish trimmed their potato demand and splurged on meat. Why? The increase in the price of potatoes meant they could afford no other food.

Yet some economists argue that the potato evidence, as originally recounted by the Victorian economist Sir Robert Giffen, is at best only anecdotal. "Search as they might," Kagel says, "economists have not been able to find unequivocal evidence to confirm this frequently cited exception to the law of demand." Not until the rats entered the picture, that is.

Kagel and Battalio's rats sojourn in cages with vending machines that dispense specific amounts of food and liquid when the animals press the appropriate lever. The six rats that took part in the Giffen-good experiment had specific "budgets" (a fixed number of lever pushes) they could spend on a quinine-and-water mixture and on root beer. The investigators changed the "prices" of the drinks by altering the amount of fluid dispensed per lever push. Like the potatoes, quinine was always the cheaper good.

Given a choice, most rats prefer root beer to quinine, Kagel says. The workers consequently lowered the rats' income and discovered that, true to the definition of a Giffen good, three of the animals soon began drinking significantly more quinine. When the price of quinine fell and those rats found they were getting more quinine for their efforts, they cut back on quinine and began drinking root beer. (The other three rats continued to drink root beer and quinine in about the same proportions, regardless of the price changes.) Such counterintuitive demand curves demonstrate that Giffen goods can exist, the investigators conclude.

In spite of Kagel and Battalio's successes, not many other economists are working with animals. "The experimental designs are interesting and seriously done," observes Thomas Romer, an economist at Carnegie-Mellon University, who reviewed a collection of essays, *Laboratory Experimentation in Economics,* recently published by Cambridge University Press. "But most economists wouldn't find it disturbing if animals didn't behave in the way predicted by theories." —*E.C.*

Modeling the Geochemical Carbon Cycle

Natural geochemical processes that result in the slow buildup of atmospheric carbon dioxide may have caused past geologic intervals of global warming through the greenhouse effect

by Robert A. Berner and Antonio C. Lasaga

Imagine for a moment a doomsday scenario in which all life on the earth is suddenly annihilated. Imagine further that all the carbon in this dead organic matter is burned to form carbon dioxide and that the carbon dioxide is released into the atmosphere. The amount of carbon dioxide this scenario generates is less than human beings will have produced by burning fossil fuels within 200 years of the Industrial Revolution.

This short exercise shows that most carbon is stored not in living tissue or in recently killed plants and animals but in sedimentary rocks. It is not surprising, then, that the familiar biological carbon cycle—in which atmospheric carbon is taken up by plants, transformed through photosynthesis into organic material and then recovered from this material by respiration and bacterial decomposition—is only one component of a much larger cycle: the geochemical carbon cycle.

Just as the biological carbon cycle governs the transfer of carbon between plants and animals and their surroundings, so the geochemical carbon cycle governs the transfer of carbon between sedimentary rocks at or near the earth's surface and the atmosphere, biosphere and oceans. Of paramount importance in both cycles is the role of carbon dioxide; it is as carbon dioxide that carbon is chiefly stored in the atmosphere. And as can be inferred from the hypothetical doomsday scenario, by the year 2050 the burning of coal and oil—which represent only a tiny fraction of the carbon in sedimentary rocks—will have significantly changed the level of atmospheric carbon dioxide. Because the manmade release of carbon dioxide is so much faster than its natural release, humans are "short-circuiting" the carbon cycle. This has led to great concern that fossil-fuel burning could critically influence the climate through the greenhouse effect.

The greenhouse mechanism has been understood at least since 1896, when the Swedish chemist Svante August Arrhenius recognized that carbon dioxide allows short-wavelength solar radiation to penetrate the atmosphere but traps this energy when it is reradiated by the earth at longer wavelengths. The effect of the carbon dioxide is the same as that of glass over a greenhouse, and it leads to a warming of the climate; the more carbon dioxide, the more warming.

Any current greenhouse warming is primarily due to the burning of fossil fuels. Variations in the biological carbon cycle also contribute to changes in atmospheric carbon dioxide over periods of tens of thousands of years. But over geologic time scales, measured in millions of years, much slower natural geochemical processes are by far the major controls on atmospheric and oceanic carbon dioxide. Such processes may have led to previous periods of greenhouse warming. Therefore it is important to understand the geochemical carbon cycle, not only for purely scientific reasons but also to see better how human beings are interfering with it.

Some aspects of the cycle were understood in the past century by the geologist Thomas C. Chamberlain and later by others (most notably Harold C. Urey), but it has been only in the past decade that the geochemical carbon cycle has been placed within the comprehensive framework of plate tectonics, a marriage we shall discuss more fully below.

The easiest way to follow the cycle is to begin with rocks on the continents [*see illustration on pages 76 and 77*]. Carbon is stored primarily in sedimentary rocks that contain two types of compounds: kerogen and carbonates. Kerogen (also known as sedimentary organic matter) represents the remains of the soft tissues of ancient plants and animals, whereas carbonate rocks represent for the most part the skeletal debris of ancient (usually marine) organisms. Kerogen is found mainly in shales; carbonates are found mainly in limestone and dolestone. Coal and oil are also derived from the soft tissues of plants and animals but are not nearly as abundant as kerogen [*see top illustration on page 78*].

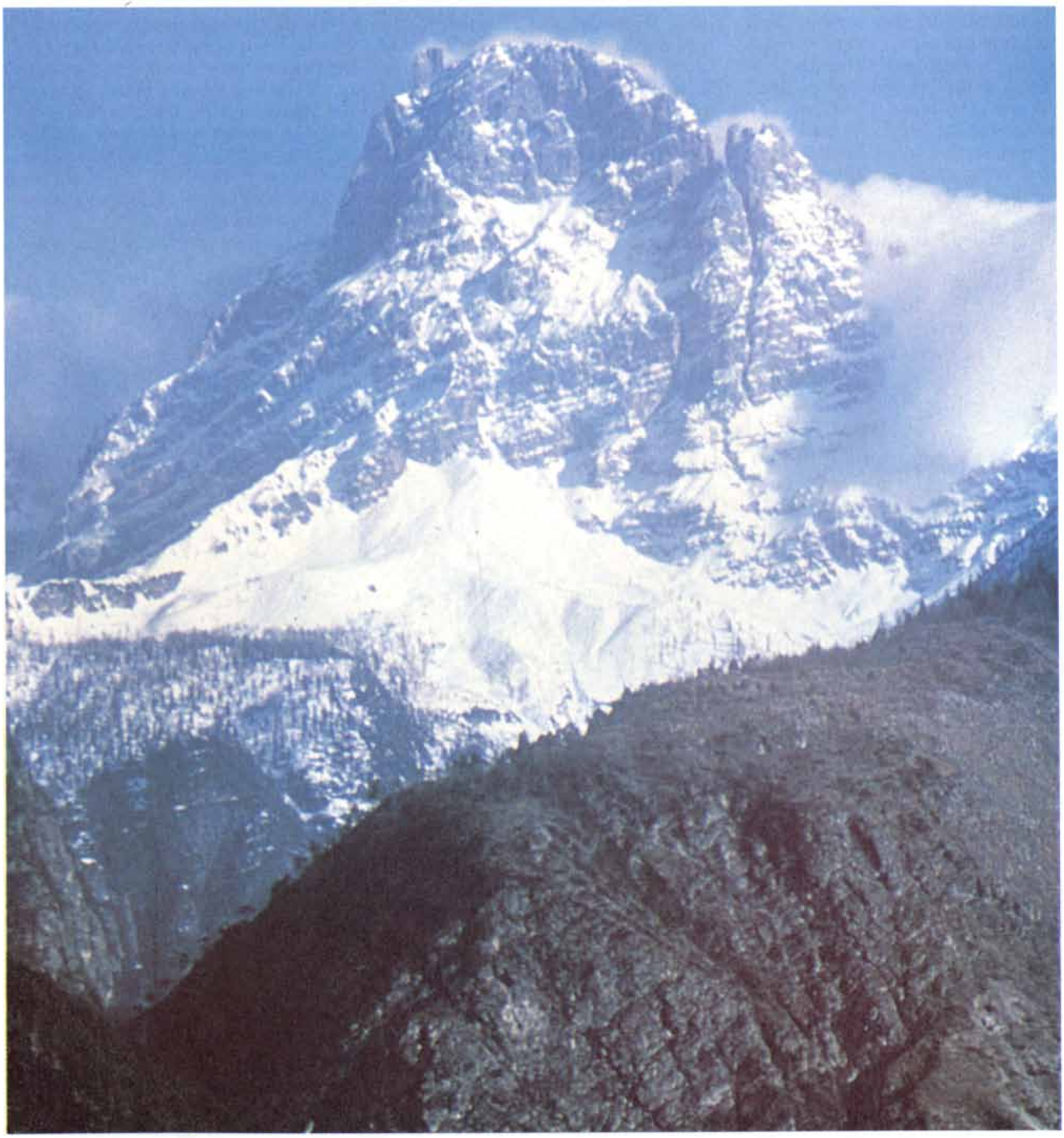
In the course of chemical weathering, whereby rocks in soils are broken down under exposure to soil gases and various acids, the kerogen merely reacts with oxygen to produce carbon dioxide (CO_2), which eventually escapes to the atmosphere. This is the natural, slow analogue to the burning of fossil fuels.

The weathering of carbonates is

ROBERT A. BERNER and ANTONIO C. LASAGA are in the department of geology and geophysics at Yale University and have collaborated for seven years on the investigation of the geochemical carbon cycle. Berner earned a B.S. and an M.S. at the University of Michigan and a Ph.D. from Harvard University in 1962. He is a past president of the Geochemical Society, is a member of the National Academy of Sciences and is the Alan M. Bateman Professor of Geology and Geophysics at Yale. Lasaga got a B.A. at Princeton University in 1971 and an M.S. and a Ph.D. from Harvard, the latter in 1976. He is currently professor of geochemistry at Yale and his research interests focus on the application of quantum mechanics to the emerging field of mineral physics. Together the authors perform Latin music, and they claim to have recorded *La Bamba* before it hit the charts.

more complicated. Magnesium and calcium carbonates ($MgCO_3$ and $CaCO_3$) contained within the minerals dolomite and calcite are attacked by acids present in the groundwater. Principal among them is carbonic acid (H_2CO_3); it is formed in the soil when the carbon dioxide that arises from the decomposition of organic matter in the biological carbon cycle reacts with water. (Ultimately the source of carbonic acid is atmospheric carbon dioxide, which is converted into organic matter through photosynthesis.) The action of the carbonic acid on the carbonate rock yields dissolved calcium, magnesium and two bicarbonate ions (HCO_3^-). In these reactions one bicarbonate ion comes from the magnesium or calcium carbonate and one from the carbonic acid.

Reactions between carbonic acid and magnesium and calcium carbonate are not the only source of bicarbonate ions. Silicate minerals, such as the common feldspars found in granites and basalts (represented here by the generalized formula $CaSiO_3$), also produce bicarbonate ions under carbonic acid weathering. Since silicates lack carbon atoms, all the carbon in the final bicarbonate ions must come from the carbonic acid. Yet, as we shall



MOUNT CAVALLO is in the Italian Dolomites, which have given their name to dolomite, a mineral containing calcium, magnesium, carbon and oxygen ($CaMgCO_3$). Dolomite formations such as Mount Cavallo are among the earth's major carbon reservoirs. Carbon removed from dolomite and other rocks during acid weathering enters the geochemical carbon cycle.

demonstrate below, silicate weathering is more important than carbonate weathering as a long-term control on atmospheric carbon dioxide.

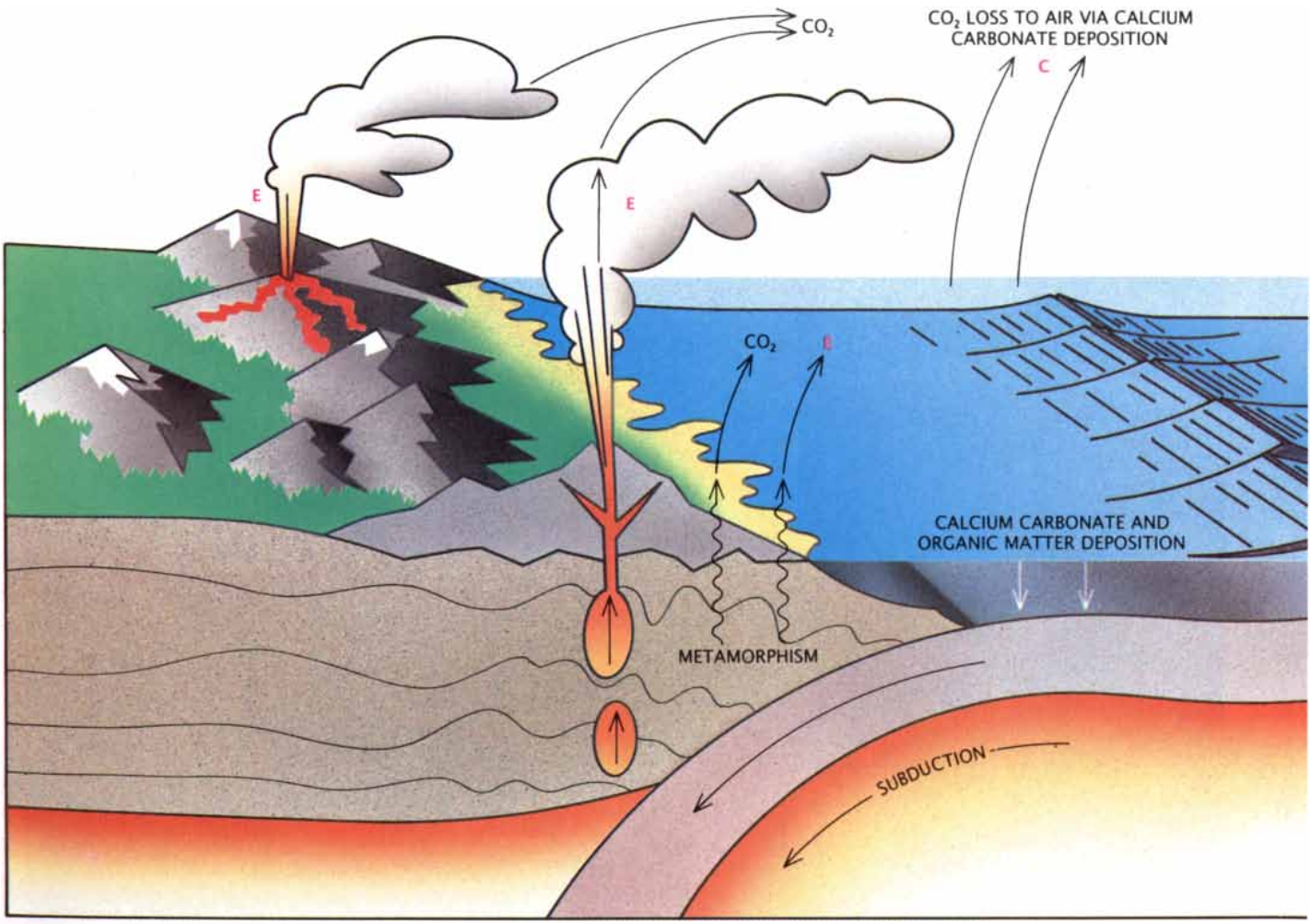
Once they have been produced by the weathering process, the bicarbonate and calcium ions are carried by groundwater to nearby streams, then to rivers and finally to the oceans. In the oceans, marine organisms such as plankton and corals take up the calcium and bicarbonate ions to construct skeletons or shells of calcium carbonate. When the organisms die, the calcium carbonate is deposited on the sea floor and eventually buried. The burial of carbonate is responsible for about 80 percent of the carbon deposited on the ocean floor; the other 20 percent is contributed by the burial of dead organic matter that has arisen from photosynthesis on land and in the sea, with land-derived material carried to the ocean by rivers.

During the reactions involving carbonates, only half of the bicarbonate ions are processed into calcium carbonate to be deposited on the ocean floor. The other half are transformed into carbon dioxide, which eventually escapes to the atmosphere, perhaps after going through yet further recycling by the photosynthesis and respiration of the biological carbon cycle.

To sum up this part of the cycle, each molecule of atmospheric carbon dioxide produces a molecule of carbonic acid in the soil. The carbonic acid molecule dissolves carbonate minerals to produce two bicarbonate ions. One bicarbonate ion is transformed by marine organisms into calcium carbonate and buried on the sea floor, eventually to become sedimentary rock; the other is transformed into carbon dioxide. In this way all the atmospheric carbon dioxide taken up during carbonate weathering is ultimately returned to the atmosphere.

The same is not true for silicate weathering. Recall that during silicate weathering two bicarbonate ions are produced. Calcium ions are also produced, since silicate minerals (for example feldspars) do contain calcium. When the bicarbonate and calcium are washed into the ocean, they are combined by marine organisms to produce calcium carbonate. In other words, the final product of silicate weathering is calcium carbonate. Tracing the reactions shows that only half of the carbon dioxide taken up from the atmosphere during silicate weathering is returned as carbon dioxide to the atmosphere. Therefore silicate weathering is extremely important because it results in a net loss of atmospheric carbon dioxide.

Were it to go unchecked for a peri-



GEOCHEMICAL CARBON CYCLE governs the transfer of carbon among the land, oceans and atmosphere. Carbon dioxide (CO_2) is taken up by plants and fixed by them in the soil, where it combines with water to form carbonic acid (H_2CO_3). Carbonic acid weathers carbonate minerals (here represented by $CaCO_3$) and silicate minerals (here represented by $CaSiO_3$) to produce bicarbonate ions (HCO_3^-), calcium ions (Ca^{++}) and dissolved silica (SiO_2). These products are transported by rivers to the ocean, where marine life incorporates the calcium and bicarbonate ions into calcium carbonate again and liberates CO_2,

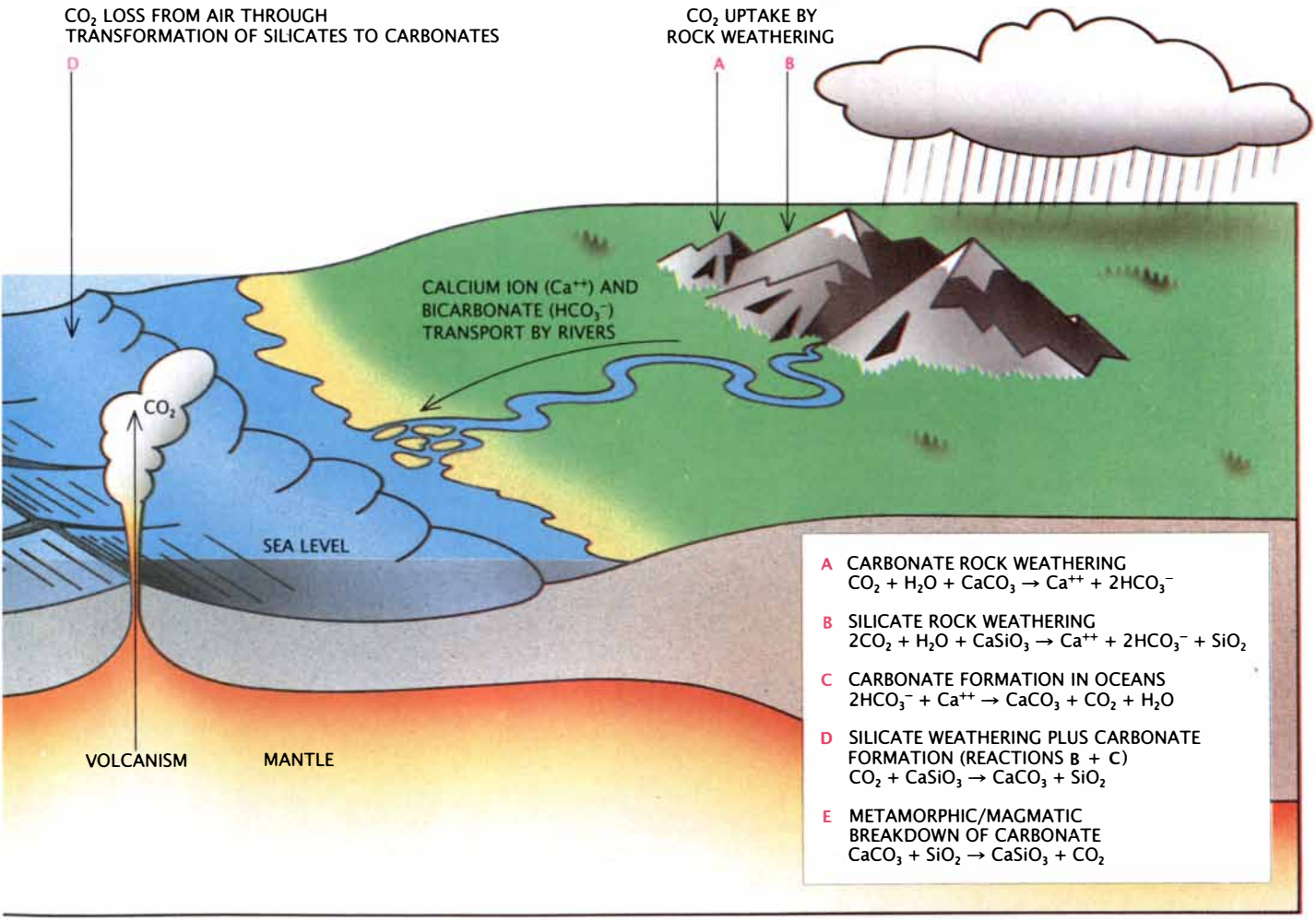
od of about 10,000 years (or about 300,000 years if gas exchange with the ocean were maintained), silicate weathering would result in the total depletion of atmospheric carbon dioxide. Clearly that has not happened or life would have ceased to exist. Some mechanism must restore atmospheric carbon dioxide. The mechanism is the release of carbon dioxide that accompanies volcanic eruptions and related phenomena.

When calcium and magnesium carbonates are buried in the earth to depths of several kilometers, they encounter temperatures high enough to initiate reactions between the calcium and magnesium and the surrounding silicates. The reactions produce new silicates and carbon dioxide. (Reactions that transform the crystalline state of a rock are termed metamorphic; if appreciable melting is involved, they are called igneous.) The carbon dioxide eventually finds its way to the atmosphere, sometimes dramatically in the course of a volcanic eruption, sometimes subtly as a soda spring, where it may end up in a bottle of Perrier. Along with the release of carbon dioxide from the carbonate deposition discussed above, such degassing is the main mechanism by which the carbon is returned to the atmosphere, thereby closing the geochemical carbon cycle.

The dynamics of degassing, and the geochemical carbon cycle in general, can best be understood in the context of plate tectonics. Although degassing can take place in a variety of locations, it is particularly common over subduction zones, regions where two of the great plates into which the earth's surface is divided collide. In the process one plate slides under the other, dragging carbonate sediments down to great depths toward the mantle, where they are subjected to intensive heating.

Ivan Barnes, Willam P. Irwin and Donald E. White of the U.S. Geological Survey have published a map that demonstrates the excellent correlation between the location of soda springs and areas of the earth's surface undergoing plate subduction, for example Japan, thereby establishing the connection between subduction and degassing. One should not conclude from this that degassing does not occur elsewhere; it does, for instance in the middle of oceanic plates or along midocean ridges.

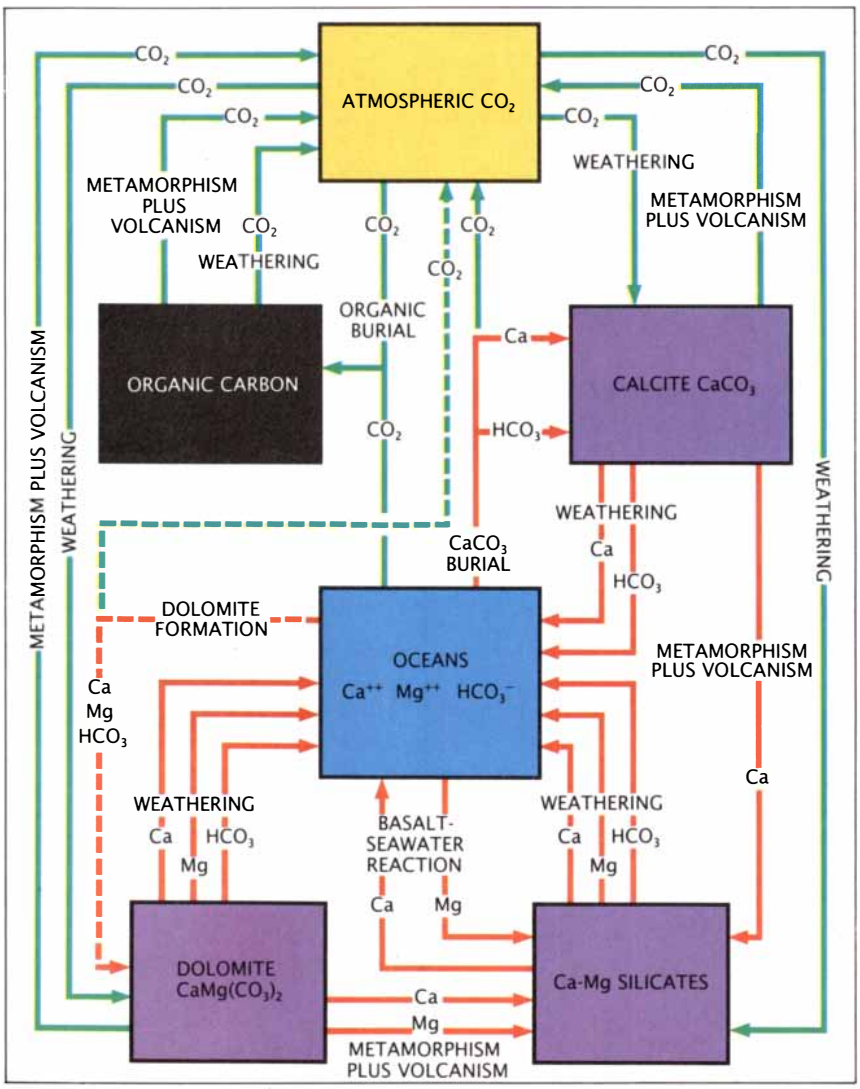
Many aspects of the geochemical carbon cycle, such as weathering, burial and degassing, can be made more quantitative with the help of computer modeling. In collaboration with the late Robert M. Garrels of the University of



which eventually escapes to the atmosphere. Carbonate weathering results in no net loss of atmospheric CO_2. The bicarbonate and calcium ions produced in silicate weathering, however, also combine to form calcium carbonate; in these reactions only half of the CO_2 is returned to the atmosphere, resulting in a net loss of atmospheric CO_2. The balance is made up deep in the earth, where calcium carbonate and silicon dioxide are heated until they combine and produce calcium silicate and CO_2. The CO_2 is degassed by volcanic eruptions and soda springs and returned to the atmosphere, completing the cycle.

FORM	CARBON MASS (10^{18} GRAMS)	RELATIVE TO LIFE
Calcium carbonate (mostly in sedimentary rocks)	35,000	62,500
Ca-Mg carbonate (mostly in sedimentary rocks)	25,000	44,600
Sedimentary organic matter (as kerogen)	15,000	26,800
Oceanic dissolved bicarbonate and carbonate	42	75
Recoverable fossil fuels (coal and oil)	4.0	7.1
Dead surficial carbon (humus, caliche, et cetera)	3.0	5.4
Atmospheric carbon dioxide	0.72	1.3
All life (plants and animals)	0.56	1

AMOUNT OF CARBON found on the earth in various forms is listed both in units of 10^{18} grams and relative to the amount found in all life. Far more carbon is stored in carbonates (the fossil remains of animal skeletons) and kerogen (the remains of soft animal tissue) than in living or recently dead organic matter, indicating that the geochemical carbon cycle is ultimately responsible for regulating atmospheric carbon dioxide over geologic time scales, which are measured in millions of years.



COMPUTER MODEL of the geochemical carbon cycle created by the authors quantifies the processes shown on the preceding two pages. Rocks that participate in chemical weathering (dolomite, calcite and silicates) are shown in purple, organic matter in black. Note that reactions involving magnesium as well as calcium are included in the model. The weathering products are carried to the ocean, where calcium carbonate deposition and burial result in the release of CO_2. All fluxes involving CO_2 are shown in green. The lines related to dolomite formation are broken because, although the processes were important in the past, they are no longer operative.

South Florida, we have constructed a computer model that has enabled us to better understand the geochemical carbon cycle and "run it backward" to calculate planetary carbon dioxide levels in the distant past. Before discussing what the model retrodicts, we should discuss its assumptions and inputs in some detail.

Clearly the rate of chemical weathering of rocks is important in determining the rate of carbon dioxide consumption. Weathering in turn depends on the amount of rock on land exposed to soil acids, and so it is reasonable to assume that the worldwide rate of weathering at a given time is proportional to the total continental land area. By the same token, changes in the weathering rate should be proportional to the change in land area. Estimates of land area over the course of geologic time are available, and so assuming that one knows the rate of weathering today one can extrapolate the weathering rate into the past.

There is a complication to be introduced here, however. As we mentioned above, the carbonic acid in soils is a result of the biological carbon cycle, that is, a product primarily of plant respiration and the decay of organic matter. That being the case, one should properly include in the weathering-rate calculations only the vegetated fraction of continental land area. Unfortunately this quantity is not known with any degree of accuracy. There are some clues. The fossil record does indicate, for example, that during the Triassic period, from about 240 to 210 million years ago, the percentage of land covered by deserts was higher than it is now. Also, before the late Silurian period, about 420 million years ago, there were essentially no vascular plants and hence probably less weathering due to biological processes. Nevertheless, because of the many uncertainties, we have simply included the total continental land area in our model.

Weathering rate is also affected by temperature, which influences both biological activity and the amount of rainfall. A variety of observations show that increasing the mean annual temperature increases the rate at which carbonates and silicates are dissolved by acids to form bicarbonate ions. For instance, chemical analyses of groundwaters show that dissolved bicarbonate content increases with temperature. Similarly, the dissolved silica content in the world's rivers is highest in areas with the hottest climate. In addition, theoretical climate modeling readily shows that the glob-

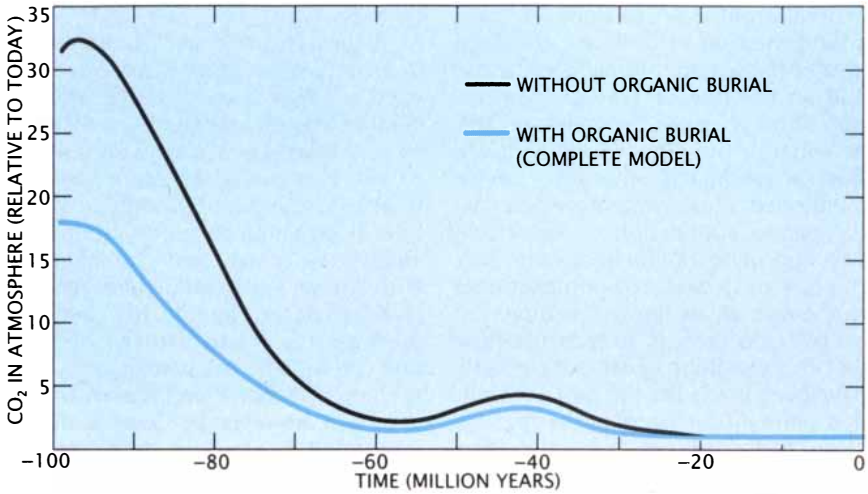
al evaporation of water is accelerated by higher temperatures; worldwide rainfall is accelerated as well, resulting in a higher weathering rate.

In order to model the past geochemical carbon cycle one therefore needs an estimate of past mean annual temperatures as an input. Rather than depend on estimates made by paleoclimatologists, we have assumed that past temperature is primarily controlled by the amount of atmospheric carbon dioxide acting as a greenhouse gas. Working with results from a general-circulation greenhouse model of the atmosphere by Syukuro Manabe and Ronald J. Stouffer of Princeton University, we have derived a simple correlation between the mean surface temperature and the level of atmospheric carbon dioxide.

We employ this relation in our model; the carbon dioxide level gives the temperature, which in turn is fed into the weathering rates. This is important because it provides a negative feedback against an excessive rise of atmospheric carbon dioxide. High carbon dioxide means higher temperature, which in turn implies a faster depletion of atmospheric carbon dioxide through weathering. As part of our procedure the atmospheric carbon dioxide level is recalculated at each time step as a function of all the other processes in the model. The new value is then used to calculate new rates of weathering and so on. Although it is complex, this approach has an advantage: the independent estimates of past temperature by paleoclimatologists serve as a check on the model.

Once again there arises a complication we have chosen to ignore. Many experimental studies have shown that plant growth is increased by exposure to higher levels of atmospheric carbon dioxide, implying higher carbon dioxide levels in the soil and increased weathering rates, independent of temperature. A higher carbon dioxide level might also result in a greater fraction of the land being vegetated. On the other hand, in many places, such as tropical rain forests, plant growth is limited by the availability of nutrients, and so it should not increase with higher carbon dioxide levels. In the absence of solid data on the worldwide response of plants to changes in atmospheric carbon dioxide, we have not included carbon dioxide's direct effect on weathering; the weathering rate is assumed to be an indirect function of atmospheric carbon dioxide through its effect on temperature.

A final complication that would influence weathering rates is topography: uplifted rugged land is subject to greater rates of erosion than low flat land, increasing the rate of carbon dioxide uptake. Once again, however, hard data are lacking and so topographic effects are not included in our current model.



CARBON DIOXIDE content of the atmosphere relative to today's level is plotted for the past 100 million years, with a given set of parameters. The abundance of CO_2 is affected by the deposition and burial of organic matter (the soft carbon-bearing remains of plants and animals) in ancient swamps and sea floors. Such burial effectively removes CO_2 from the atmosphere, a fact reflected in the curves: the lower curve includes organic burial and results in less atmospheric CO_2, whereas the upper one is a model without organic burial, resulting in more atmospheric CO_2. The hump at about 40 million years ago reflects an increase in the rate of sea-floor spreading, leading to an increased degassing and atmospheric abundance of CO_2.

To complete the geochemical carbon cycle one needs to know the deposition and burial rate of organic matter representing the soft parts of organisms—the parts that eventually become kerogen. For this purpose we have relied on two independent methods that serve to check each other.

The first relies on the two common stable isotopes of carbon, ^{12}C and ^{13}C. In photosynthesis plants preferentially absorb the former. On the other hand, the incorporation of carbon into the calcareous skeletons of marine organisms and their deposition on the sea floor treats the isotopes equally. Consequently the $^{13}C/^{12}C$ ratio in a given layer of carbonate sediment in the sea floor records the relative levels of the two isotopes in the ocean at the time the layer was deposited; this in turn reflects the relative rates at which organic and carbonate materials were then being created.

Although there is some argument over details, investigators agree that in the course of Phanerozoic time (since the end of the Precambrian era, approximately 570 million years ago) the $^{13}C/^{12}C$ ratio has varied from low values between 570 and 350 million years ago through high values between 350 and 250 million years ago to intermediate values since then. Our colleague Garrels and Abraham Lerman of Northwestern University have fitted the $^{13}C/^{12}C$ data into a mathematical model, which we have incorporated into our code.

The second method by which we calculate the burial rate of organic matter relies on data collected by Aleksandr B. Ronov, Mikhail I. Budyko and their co-workers at the Leningrad State Hydrological Institute. The group has compiled data on the relative abundances of various types of sedimentary rocks over geologic time. Knowing the original abundances and the organic carbon content of the major rock types—marine sandstones and shales, coal-basin sediments and continental red beds—one can compute the worldwide burial rates of organic matter for the past 570 million years. The results agree well with the rates calculated by the $^{13}C/^{12}C$ method, giving some assurance that both approaches are reasonable.

The most vexing problem we have encountered in modeling the geochemical carbon cycle is to calculate the rate of degassing of carbon dioxide due to igneous and metamorphic activity. Our approach is to assume that the degassing rate is directly proportional to the rate of

generation of new sea floor. The reasoning here is as follows: the total area of the sea floor remains constant, and so the rate of sea-floor generation through plate tectonics should be equal to the rate at which old sea floor is subducted. Since, as we have mentioned, degassing generally accompanies subduction, a reasonable first approximation is to assume that the rate of degassing is proportional to the rate of sea-floor spreading.

The trick, then, is to measure how fast the sea floor spreads. Estimates have been made for the past 100 million years by determining the age of a given region of sea floor and then making certain assumptions about how much of that region has already been lost to subduction. Unfortunately estimates arrived at in this way differ by factors of two, and so we have been forced to include several different formulas in our code that relate the degassing rate to the spreading rate.

When one goes further back in time, the problem is even more difficult: there remains little sea floor more than 150 million years old that has not been subducted, and direct age measurement is impossible. One can circumvent this obstacle by relating worldwide sea level to rates of sea-floor generation, using a model developed by Walter C. Pitman III and his co-workers at the Lamont-Doherty Geological Observatory. When new sea floor is created along midocean ridges, the ridges become wider. The increased volume of the ridges displaces seawater, which rises with respect to the continents. Estimates of sea level over the past 600 million years have been made based on studies of shallow-water sediments, for instance. With Pitman's correlation between sea level and spreading rate, the past rate of degassing relative to the present rate can then be calculated.

There is a major problem with this approach, however. Sea level is affected not only by sea-floor spreading but also by many other processes, such as the breakup and formation of continents and the large-scale removal of ocean water to form glaciers. Before relating sea level to the rate of ocean-floor spreading, account must be taken of these factors. Attempts have been made, but they involve processes extending back only as far as 150 million years ago, only a fourth of the 600 million years for which there are data on sea level.

Clearly more work is needed to quantify rates of worldwide carbon dioxide degassing. To date we have depended on various estimates that relate the area of the sea floor to its age, as discussed above, in order to obtain degassing rates for the past 100 million years. Better methods would enable us to improve the model and include carbon dioxide production for the past 600 million years.

WHITE ISLAND in the Bay of Plenty, New Zealand, lies over the Tonga-Kermadec–New Zealand subduction zone, where the Pacific plate slides under the Indian plate. The subduction process drags carbonate sediments down to great depths, where they undergo intense heating. Reactions between carbonates and silicates produce CO_2, which finds its way to the atmosphere, often through a soda spring or a volcano.

As we mentioned above, we assume the degassing rate is directly proportional to the rate of generation of new sea floor. This approach enables us to determine ancient rates relative to today's rate; to get absolute rates it is necessary to fix the constant of proportionality by determining the current rate of degassing. In principle one could do this by adding up all sources, such as volcanoes, soda springs and the like, but this would be at best tedious and at worst nearly impossible.

It is easier to build on an assumption that the level of atmospheric carbon dioxide has not fluctuated wildly with time. In that case volcanic-metamorphic degassing must essentially balance the carbon dioxide taken from the atmosphere in weathering, calcium carbonate deposition and the other processes we have discussed. The rates of these processes are known reasonably well; their sum must equal the present degassing rate.

The assumption that carbon dioxide has not fluctuated wildly with time is clearly valid, at least to a good approximation: life exists. If the degassing rate were halved (while holding the other rates constant), all the atmospheric and available oceanic carbon dioxide would be exhausted within 600,000 years, resulting in the cessation of photosynthesis. If the degassing rate were doubled, the excess carbon dioxide would lead through the greenhouse effect to the extinction of plant and animal life in just a few million years.

The necessity of balancing the carbon dioxide fluxes does not mean that carbon dioxide cannot and has not fluctuated somewhat over sufficiently long periods. The amount of carbon dioxide in the atmosphere at any given time is small compared with the amount entering and leaving it, so that slight imbalances in the fluxes can produce large changes in the atmospheric carbon dioxide level. As will become clear, this is in fact the principal result of our calculations: slight imbalances in the fluxes between the components of the geochemical carbon cycle have led to a general decrease in atmospheric carbon dioxide over the past 100 million years. Consequently, owing to a lessening of the greenhouse effect, the earth has cooled as well.

How do our calculations compare

with independent observations and how do we interpret the results? There is no doubt that the mean surface temperature of the earth was higher during the Cretaceous period (135 to 65 million years ago) than it is today. Much evidence exists for this assertion: fossils of plants and animals (for instance alligators) that live today in warm climates have been found in what were Cretaceous polar regions. The oxygen composition of deep-sea marine fossils also shows that temperatures were higher during the Cretaceous than they are today.

Eric J. Barron and his co-workers at the National Center for Atmospheric Research have found with the help of a general-circulation climate model that they can match the warm high-latitude Cretaceous temperatures by raising the carbon dioxide level of the atmosphere. Increasing the ocean circulation rate in the model, changing the position of the continents or increasing the amount of solar radiation did raise the temperature, but not to Cretaceous levels. From four to eight times the present amount of carbon dioxide was needed to match the Cretaceous polar temperatures estimated by paleontologists and geochemists.

Such results have convinced us that climatic changes in the geologic past have been caused largely by changes in the atmospheric carbon dioxide content. This "paleogreenhouse" view has been forcefully championed by Alfred G. Fischer, formerly of Princeton University. He has shown that over the past 600 million years there is a good correlation between higher temperatures, high sea levels and a greater abundance of igneous rocks (the last two being indicators of increased carbon dioxide degassing). During glacial periods he finds low sea levels and less igneous rock. Fischer has suggested that the earth periodically passes from "greenhouse periods" to "icehouse periods," depending on the amount of tectonic activity and carbon dioxide degassing.

With our computer model we have indeed shown that the primary factor affecting the carbon dioxide level is actually the rate of degassing. We are therefore in accord with Fischer that atmospheric carbon dioxide and the world climate in general are mainly controlled by tectonism, processes taking place deep within the earth.

Second in importance for controlling the carbon dioxide abundance turns out to be the rate of burial of organic material. Since the burial of carbon in the sea floor effectively removes it from the atmosphere, Cretaceous burial rates that are high in relation to the present rate tend to lower the Cretaceous carbon dioxide level (and temperature).

Because the weathering of organic matter is carried out through oxidation by atmospheric oxygen, it is largely independent of the carbon dioxide level; moreover, to a good approximation it does not depend on temperature and degassing rate. The rate of burial of organic matter is independent of these factors too and can be calculated for all Phanerozoic time. Our results show that organic burial rates have fluctuated above and below the present rate by a factor of about two; this reflects the value of the oceanic $^{13}C/^{12}C$ ratio discussed above.

According to our model, the Carboniferous and Permian periods (from 350 to 250 million years ago) show the highest organic burial rates. We believe this is because of two factors. First, before that time the only source of organic matter was marine life and the organic burial rate would have been low. Just before the Carboniferous period, about 400 million years ago, vascular plants appeared and spread over the continents, providing a new source of organic matter resistant to bacterial degradation. Such material (for example lignin, an essential part of woody fiber) would have entered the geochemical carbon cycle once it was deposited in swamps, lakes or the ocean; the organic burial rate would have increased.

The second reason for the excessive organic deposition rates during the Carboniferous and Permian periods was the presence of vast lowlands exposed above sea level. Most of the world's land was joined together in one supercontinent called Pangaea. The formation of Pangaea caused a regression of the sea from the land and a drop in sea level. Land that had been under water was now exposed and in humid regions was covered by vast freshwater swamps. In contrast to upland soils, swamps protect organic matter from bacterial decay, and so once the dead plant matter was deposited in the swamps it would have been preserved and turned into kerogen and coal. According to coal-abundance estimates, the Carboniferous and Permian periods were the most important coal-forming epochs in the earth's history; this agrees qualitatively with the times of highest organic burial rates in our independent calculations based on the $^{13}C/^{12}C$ ratio.

Now, high organic burial rates have an important additional effect on the composition of the atmosphere. Burial of organic matter represents an excess of photosynthesis (net oxygen production) over destruction by bacterial respiration (net oxygen depletion). Consequently, if all other factors remained constant and the burial rates we calculate for the Carboniferous and Permian periods are correct, the atmosphere at that time would have been richer in oxygen than at any time before or since.

To sum up the various threads of our argument, our model calculations have enabled us to offer a preliminary explanation of climatic conditions in the earth's distant past. Our principal result is that slow natural fluctuations of atmospheric carbon dioxide over time scales of millions of years may rival or even exceed the much faster changes that are predicted to arise from human activities or that may result from the biological carbon cycle. By the same token, establishing that carbon dioxide has been a major influence on global temperature in the past gives some indication of what may be expected in the future if fossil-fuel burning continues at the present rate.

The main purpose of modeling the geochemical carbon cycle, however, is to expose how little is known about the rates of important global processes and how seemingly unrelated processes—such as tectonism and climate—are linked. Modeling the earth's climatic history is a multidisciplinary activity that draws on, among other things, geochemistry, geophysics, biology, soil science, paleontology, meteorology, climatology, oceanography and astronomy. Further progress in this field will require the continued collaboration of scientists from these and other disciplines.

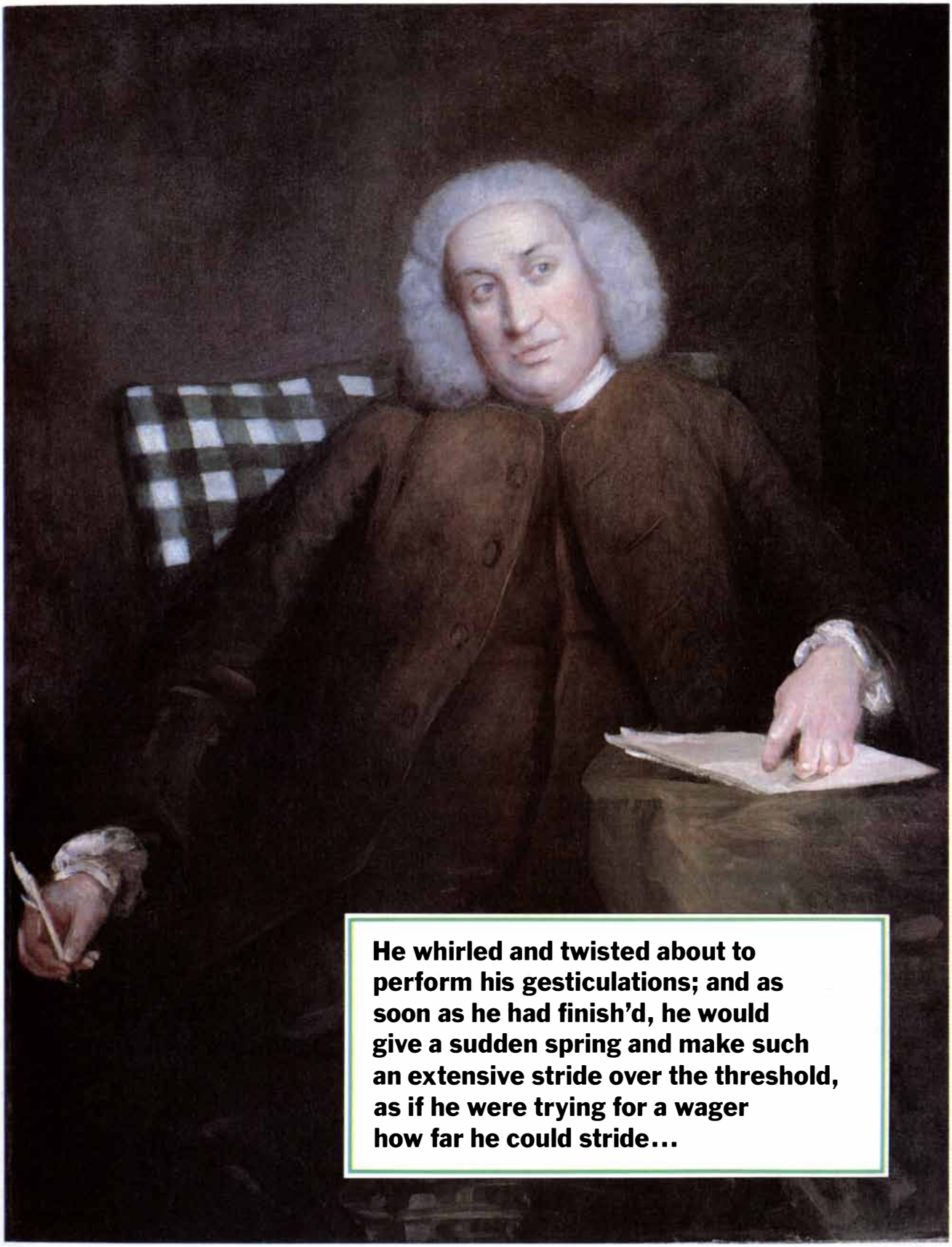
FURTHER READING

THE CARBONATE-SILICATE GEOCHEMICAL CYCLE AND ITS EFFECT ON ATMOSPHERIC CARBON DIOXIDE OVER THE PAST 100 MILLION YEARS. Robert A. Berner, Antonio C. Lasaga and Robert M. Garrels in *American Journal of Science,* Vol. 283, No. 7, pages 641-683; September, 1983.

THE CHEMICAL EVOLUTION OF THE ATMOSPHERE AND OCEANS. H. D. Holland. Princeton University Press, 1984.

THE TWO PHANEROZOIC SUPERCYCLES. A. G. Fischer in *Catastrophes and Earth History,* edited by W. A. Berggren and J. A. Van Couvering. Princeton University Press, 1984.

ATMOSPHERIC CARBON DIOXIDE AND THE GLOBAL CARBON CYCLE. Edited by J. R. Trabalka. U.S. Government Printing Office, 1985.



SAMUEL JOHNSON, the 18th-century English critic, author and lexicographer, displayed various obsessive and compulsive symptoms. The portrait was painted in about 1756 by Sir Joshua Reynolds. Inset into the portrait is a description by a contemporary of Johnson's curious ritualistic behavior when passing through any door. In their severest form obsessions and compulsions constitute obsessive-compulsive disorder (OCD), a disease that can cripple the victim's social existence.

The Biology of Obsessions and Compulsions

Advances in pharmacology and in brain imaging suggest that severe obsessions and compulsions are biologically rooted. Several new drugs, first formulated as antidepressants, relieve them

by Judith L. Rapoport

Sergei is a 17-year-old former high school student. Only a year or so ago Sergei seemed to be a normal adolescent with many talents and interests. Then, almost overnight, he was transformed into a lonely outsider, excluded from social life by his psychological disabilities. Specifically, he was unable to stop washing. Haunted by the notion that he was dirty—in spite of the contrary evidence of his senses—he began to spend more and more of his time cleansing himself of imaginary dirt. At first his ritual ablutions were confined to weekends and evenings and he was able to stay in school while keeping them up, but soon they began to consume all his time, forcing him to drop out of school, a victim of his inability to feel clean enough.

Sergei's condition is called obsessive-compulsive disorder, or OCD. Previously thought to be quite rare, it is now known to affect perhaps as much as 2 percent of the U.S. population. OCD is resistant to family counseling, psychotherapy and most drugs for treating anxiety and depression. Yet trials my colleagues and I have conducted at the National Institute of Mental Health show that OCD yields to certain new antidepressant medications. The capacity of these drugs to reduce obsessive-compulsive behavior is quite distinct from their effect on depression. This is one piece of evidence (among others) that OCD is not a function of mood but a specific, biologically rooted syndrome. Although the details are far from clear, a new, biological model of OCD is rapidly emerging.

A key feature of the model is the idea that certain behavioral "subroutines" related to grooming and territoriality have been programmed into the human brain over the course of evolution. Ordinarily the evidence of the senses (that one is clean or that the stove is off) are sufficient to keep such subroutines suppressed. If higher brain centers malfunction, however, the subroutines may be replayed repeatedly, and sufferers from OCD are left at their mercy, unable to stop washing or checking the stove even though consciously they understand that such behavior is "crazy." The new biology of obsessions and compulsions represents a great step forward in understanding; because of the availability of new drugs with anti-OCD activities, it has been accompanied by equally great strides in therapy.

The terms obsessive and compulsive are by now part of everyday language. People often say "He's compulsive" when what they mean is that the person is an uptight bore; they say "She's obsessed by him" when they mean that a friend is a lovesick ninny. That is not how these terms are applied here. Obsessive-compulsive disorder is a severe, chronic psychiatric problem. It manifests itself through obsessions (recurrent, persistent ideas, thoughts or impulses that are experienced, at least initially, as intrusive and senseless) or compulsions (repetitive, purposeful behaviors—perceived as unnecessary—that are performed in response to an obsession, or according to certain rules or in a stereotyped fashion).

The form these processes take in OCD typically include irresistible urges to wash, to check doors (to make sure they are locked) or appliances (to make sure they are turned off) or to count repetitively, and by the presence of intrusive thoughts (frequently of dangerous or unacceptable behaviors). The difference between OCD and the milder forms of compulsion seen in otherwise healthy people is that these behaviors have become so demanding and time-consuming that they interfere with the patient's life to a considerable degree—as in Sergei's case.

One of the surprising things about sufferers from OCD is that their disease is limited: in other areas they are quite reasonable. What is more, they understand that their obsessive-compulsive behaviors are irrational, and yet they cannot do much to control them on their own. As a result these patients suffer painfully. When the symptoms are severe, they can make the patient appear ridiculous, as Samuel Johnson, a sufferer from the disorder, must have seemed to his companions when he performed odd ritualistic hand movements before leaping through doorways. The victim of OCD may even become an unwashed hermit, as Howard Hughes did. (Although Hughes's fate seems paradoxical in light of the cleanliness compulsions of OCD patients, a plausible explanation is that his cleanliness rituals became so demanding that he could not complete them and as a result became psychologically paralyzed.)

My own medical and psychiatric training had given me little exposure

JUDITH L. RAPOPORT is chief of the Child Psychiatry Branch of the National Institute of Mental Health. She studied psychology as an undergraduate at Swarthmore College and got her M.D. at the Harvard Medical School. After residencies in both child and general psychiatry she did further graduate training in experimental psychology. Her research has generally addressed the biological aspects of child psychiatry. Rapoport is the author of *The Boy Who Couldn't Stop Washing,* the first popular book on obsessive-compulsive disorder.

to patients with OCD. The reason is not that OCD is such a rare disorder; it is that sufferers from OCD rarely seek psychiatric help. Yet as a child psychiatrist I had been intrigued even by the few cases I saw, partly because they seemed identical with the few adult cases that were familiar to me. For a psychiatric disorder to appear in identical form in children and adults is somewhat unusual; it is more common for a particular disease to appear most frequently at a specific point in the life cycle (as schizophrenia, for example, appears most frequently in late adolescence).

Yet in our study I have seen children whose odd, repetitive acts (such as walking in circles or counting or tapping a certain number over and over again—acts they felt forced to carry out against their will) began as early as the age of three. Such children have their own explanations for why they carry out these strange acts. For example, at age seven Stanley saw a television program in which friendly Martians contacted human beings by putting odd thoughts into their heads. On the basis of that program Stanley decided his compulsion to do everything in sequences of four was a sign that the Martians had picked him as their "contact man" on earth.

OBSESSIONS	REPORTED SYMPTOM AT INITIAL INTERVIEW	
	NUMBER	PERCENT
Concern with dirt, germs or environmental toxins	28	(40)
Something terrible happening (fire, death or illness of self or loved one)	17	(24)
Symmetry, order or exactness	12	(17)
Scrupulosity (religious obsessions)	9	(13)
Concern or disgust with bodily wastes or secretions (urine, stool, saliva)	6	(8)
Lucky or unlucky numbers	6	(8)
Forbidden, aggressive or perverse sexual thoughts, images or impulses	3	(4)
Fear might harm others or oneself	3	(4)
Concern with household items	2	(3)
Intrusive nonsense sounds, words or music	1	(1)

COMPULSIONS	REPORTED SYMPTOM AT INITIAL INTERVIEW	
	NUMBER	PERCENT
Excessive or ritualized handwashing, showering, bathing, toothbrushing or grooming	60	(85)
Repeating rituals (going in or out of a door, up or down from a chair)	36	(51)
Checking (doors, locks, stove, appliances, emergency brake on car, paper route, homework)	32	(46)
Rituals to remove contact with contaminants	16	(23)
Touching	14	(20)
Measures to prevent harm to self or others	11	(16)
Ordering or arranging	12	(17)
Counting	13	(18)
Hoarding or collecting rituals	8	(11)
Rituals of cleaning household or inanimate objects	4	(6)
Miscellaneous rituals (such as writing, moving, speaking)	18	(26)

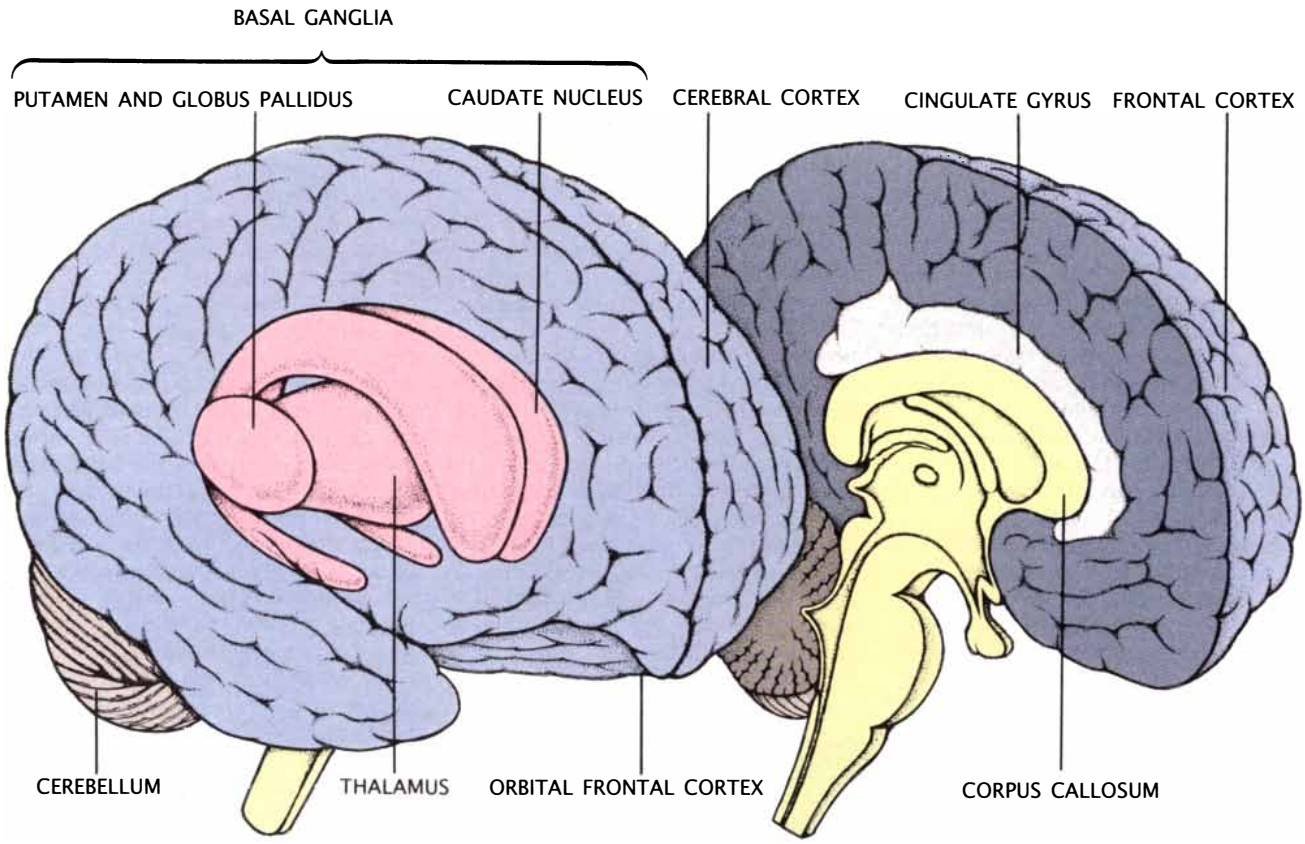
TABLE OF SYMPTOMS indicates the most frequent obsessions and compulsions among 70 children and adolescents who were diagnosed as having OCD by the author and her colleagues at the National Institute of Mental Health (NIMH). The proportions total more than 100 percent because many sufferers have more than one symptom.

After two years of sterile counting rituals, no contact had been made and Stanley gave up this explanation. He did not, however, give up counting. Nor do other OCD patients generally give up their ritualistic behavior without treatment, in spite of their knowledge that the behavior is "crazy." Often, though, there will be a progression of symptoms: from counting in childhood to washing rituals in adolescence to obsessive thoughts in early adulthood. From a third to a half of all the victims of OCD first began to experience these behaviors in childhood or adolescence.

Perhaps in part because they understand that their rituals and thoughts are senseless, OCD patients may go to great lengths to hide them. These patients simulate normality as long as they can by limiting their rituals to private hours and avoiding intimate social situations in which their compulsions would be discovered. Typically the symptoms are concealed for years and help is sought only when the symptoms can no longer be managed because they have started to make the patient avoid school, work or social obligations.

Such concealment is the reason that there were initially thought to be few sufferers from OCD. Indeed, when my colleagues and I began drug trials at the NIMH, we did not know whether we would be able to find enough subjects to complete the work. Those trials began in the mid-1970's, triggered by early reports from Spain, England and Sweden that a new drug called clomipramine (CMI), formulated as an antidepressant, had specific antiobsessional activity. At the beginning we were quite skeptical that any drug could have a specific anti-OCD effect, particularly a drug that was closely related chemically to a standard antidepressant, as clomipramine is.

More than a decade later the situation is quite different. First, we now know that CMI can relieve obsessive-compulsive symptoms. Second, as a result of the publicity our trials received, we now have hundreds of patients requesting treatment. What is more, the new attention focused on the disorder has led to careful epidemiological studies that show OCD is much commoner than was once thought. The prevalence of OCD was recently measured in five U.S. communities among more than 18,000 people interviewed in their homes as part of the NIMH Epidemiological Catchment



NEUROANATOMY OF BASAL GANGLIA is shown on a three-dimensional view of the human brain. The basal ganglia consist of several structures, including the caudate nucleus, the putamen and the globus pallidus, that lie under the cerebral cortex. The basal ganglia are connected to the frontal lobe by a variety of pathways, including one that contains tissues in the cingulate gyrus. The author proposes that aberrations in this brain circuitry may underlie the symptoms of OCD.

Area program. The overall prevalence of OCD ranged from 1.9 to 3.3 percent in the five communities: a rate from 25 to 60 times greater than previous estimates.

The prevalence rates in the catchment-area study were lifetime rates, that is, the rate that might be expected in a group of chronological peers at the end of their life span. That rate is supported by a study of more than 5,000 high school students carried out by my colleagues and me in collaboration with Agnes H. Whitaker of the Columbia University College of Physicians and Surgeons and her collaborators. Among this much younger group the cumulative prevalence of OCD was about 1 percent. If the two studies are correct, from four to six million people in the U.S. suffer from the disorder.

The dimensions of the problem can be appreciated when one considers how much suffering there is among those four to six million people. Even for those who can "pass" in school or on the job, OCD may mean a lifetime of worry and isolation. OCD patients are more likely than others to divorce or not to marry in the first place—perhaps partly to avoid discovery of their private rituals. Follow-up studies of people who have been clinically treated show that the disorder is chronic and recurrent: at least 50 percent of those treated with psychotherapy or older drug therapies suffered from the disease seven to 20 years after the end of their treatment.

The experience my colleagues and I have acquired as the result of the trials beginning in 1975 have led us toward an ethological perspective on OCD. Ethology is the scientific study of animal behavior. Some of the best work in ethology has been done by Konrad Lorenz, who described nest building, grooming, courtship and defensive behavior patterns in young animals, mostly of avian species. These patterns appeared without learning models, and Lorenz hypothesized that they are "hardwired" into the brain circuitry. Many of the behaviors shown by OCD patients seem to resemble the fixed-action patterns described by Lorenz. It is obvious that cultural and physical stimuli account to some degree for a particular patient's symptoms, but the ritualized aspect of the behavior and its startling uniformity—along with the fact that children and adults show identical symptoms—suggest biological preprogramming.

In addition to the uniformity of the behaviors, certain other features of obsessions and compulsions led my colleagues and me to believe in a biological basis for OCD, which had been suggested by others but was buttressed by the new findings. For one thing, the disease is more prevalent among relatives of people with OCD than among the general population, suggesting a possible genetic cause. In addition, an association between OCD and certain neurologically rooted problems implied that the disease was closely linked to the anatomy of the nervous system. About 20 percent of patients also display motor tics: involuntary movements that are usually blinks of the eye or facial grimaces.

The connection between involuntary motor movements and obsessive-compulsive symptoms has been known for some time. In 1896 Sir William Osler, who was then physician-in-chief at the Johns Hopkins Hospital, described a seven-year-old girl who suffered from a syndrome that included the compulsion to count as well as odd, chorealike motions. (Chorea is a general term for rapid, jerky movements that appear to be willed but are actually involuntary.) Since that time it has been shown repeatedly that OCD occurs in association with several types of neurological disorders: Sydenham's chorea, epilepsy, postencephalitic Parkinson's disease and toxic lesions of a part of the brain called the basal ganglia.

Now, this association is intriguing, in part because all the disorders affect the basal ganglia, a group of structures lying under the cerebral cortex that are known to be "way stations" between sensory inputs and the resulting motor or cognitive outputs. Is it possible that in obsessive-compulsive patients disturbances in these way stations have somehow short-circuited the loop that normally connects sensory input with behavioral output, thereby releasing stored, hard-wired behavioral packages?

In our own work we now have several independent lines of evidence implicating the circuitry of the basal ganglia in OCD. As I noted, about 20 percent of our patients show chorea-like twitches resembling those of Osler's patient, and standard tests designed to identify specific neurological function indicate that our patients have functional deficits in the frontal lobes or the basal ganglia or both. We have now gone beyond such preliminary data to complete three additional types of study pointing toward the involvement of these brain regions in obsessive-compulsive disorder.

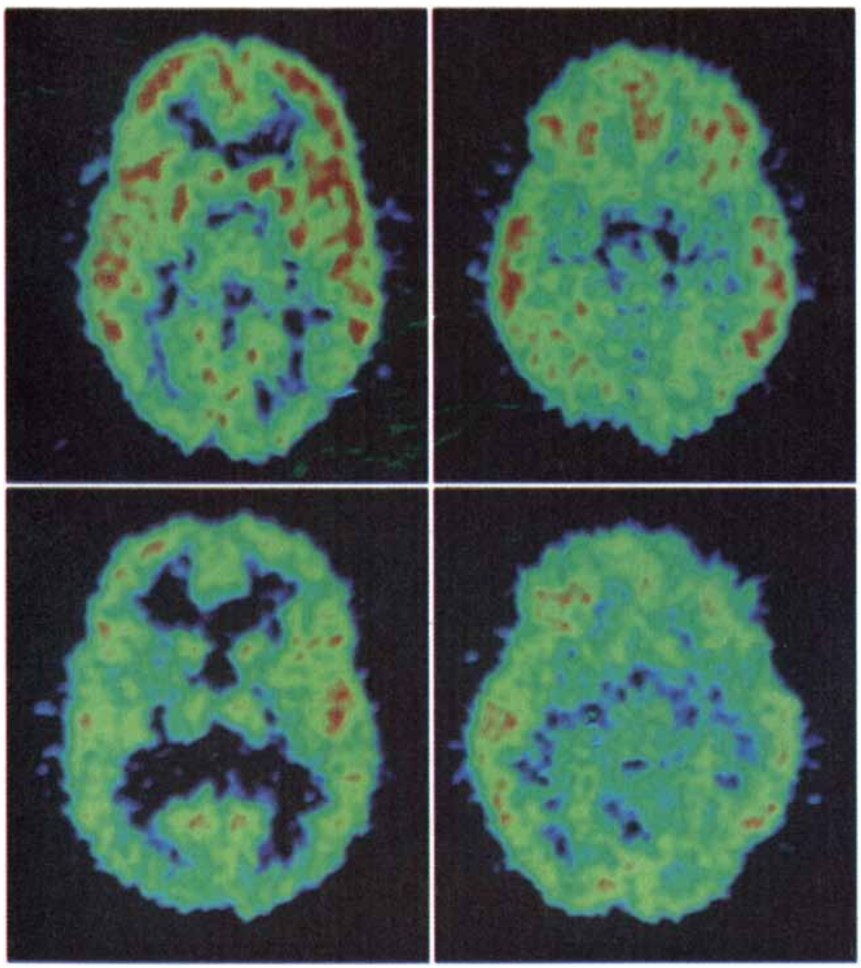
A study in which we collaborated with Jay Luxenberg of the National Institute on Aging and his colleagues compared computerized axial tomography (CAT) scans of the brains of OCD patients with those of a matched control group who did not have the disease. In particular we were interested in the volume of the caudate nucleus, one of the structures in the basal ganglia; our patients had smaller caudate volumes.

A second study took advantage of an "experiment of nature" that unfortunately was affecting many adolescents. There has recently been a resurgence of rheumatic fever in parts of the U.S. Rheumatic fever is a chronic inflammatory disease of the heart and joints that sometimes develops after certain streptococcal infections. About 20 percent of rheumatic-fever patients also develop Sydenham's chorea, probably as the result of an autoimmune response to the basal ganglia, leading to potential damage there.

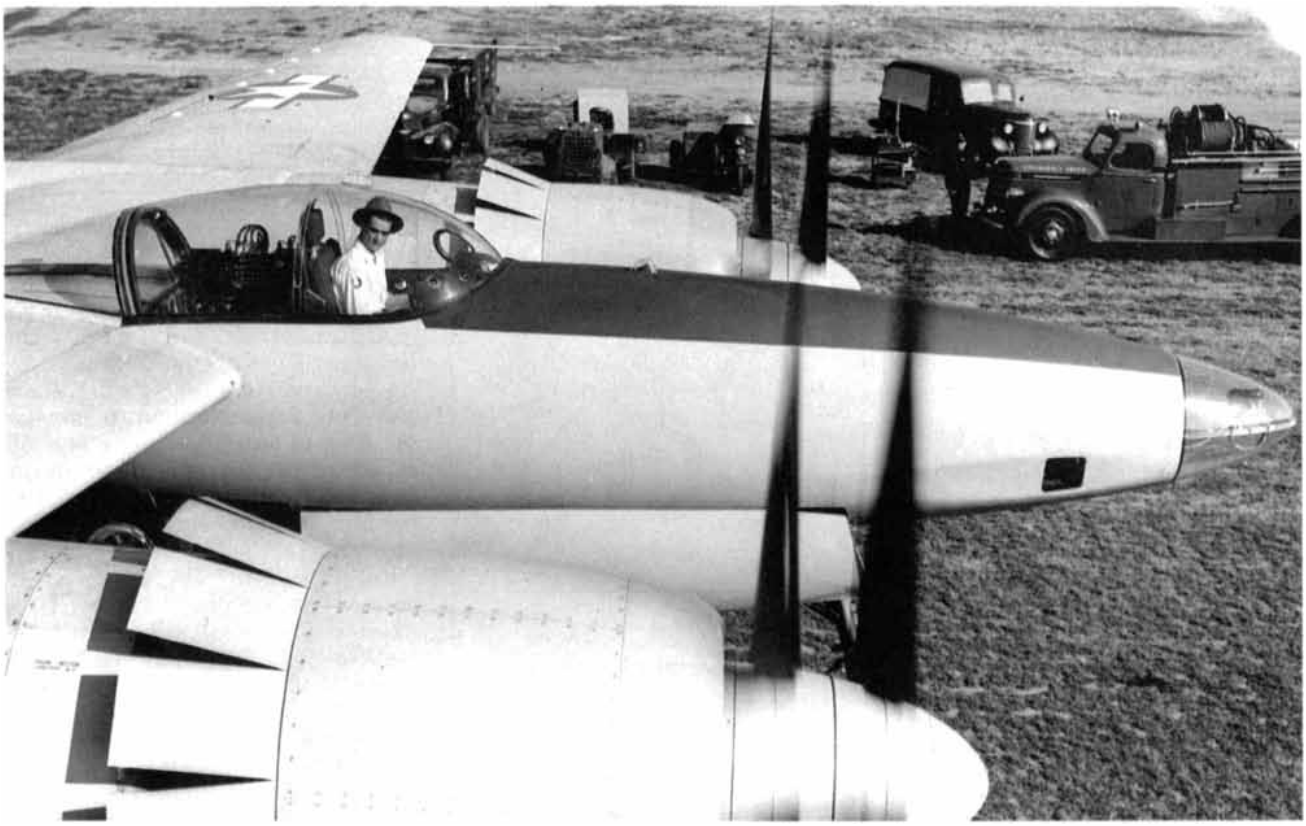
My colleagues and I conducted a survey of 23 Sydenham's chorea patients and 14 patients who had rheumatic fever but did not have chorea (all the subjects the five collaborating pediatric departments could locate). In "blind" evaluations (in which the interviewer did not know the medical diagnosis) scores for obsessional symptoms were significantly higher among those with Sydenham's chorea. What is more, three chorea patients—but no rheumatic-fever patients—met our diagnostic criteria for full-fledged OCD. This result again suggests that OCD, at least in some patients, is due to dysfunction of the basal ganglia.

The third study was carried out by my colleague Susan Swedo in collaboration with Mark Schapiro and Cheryl L. Grady of the National Institute on Aging. The three investigators compared obsessive-compulsive patients and normal controls by means of positron-emission tomography (PET). PET scans yield brain images in which metabolic activity can be measured noninvasively in resting subjects. The study showed that OCD patients had higher levels of glucose metabolism in an area of the frontal lobe and in the cingulate pathway, which connects the frontal lobe and the basal ganglia. These results confirmed earlier ones by Lewis R. Baxter of the University of California at Los Angeles School of Medicine and his colleagues; in both studies the elevated glucose metabolism was correlated with measures of the severity of OCD.

The body of recently accumulated evidence makes it seem quite likely that obsessive-compulsive symptoms,



PET SCANS reveal metabolic differences between brains of OCD victims (*top*) and those of normal controls (*bottom*). In these PET (positron-emission tomography) scans red corresponds to the highest level of metabolic activity and blue to the lowest level. OCD sufferers have elevated levels in several areas, including the basal ganglia.



HOWARD HUGHES was a 20th-century victim of OCD. Throughout his life Hughes was preoccupied with dirt and germs; this preoccupation led him to take various ritualized precautionary measures. Toward the end of his life he became increasingly isolated, unwashed and ungroomed. The author hypothesizes that Hughes's rituals eventually became so time-consuming that they could not be carried out; when he was unable to fulfill his ritual precautions, Hughes became a hermit.

at least in their severest form, have a distinct biological basis. Although the precise mechanism by which the symptoms are formed is yet to be elucidated, there are, as I have described, some clear hints about which brain regions are involved. At the same time as those strides were being made in elucidating the basis of the disease, corresponding advances were being made in treatment.

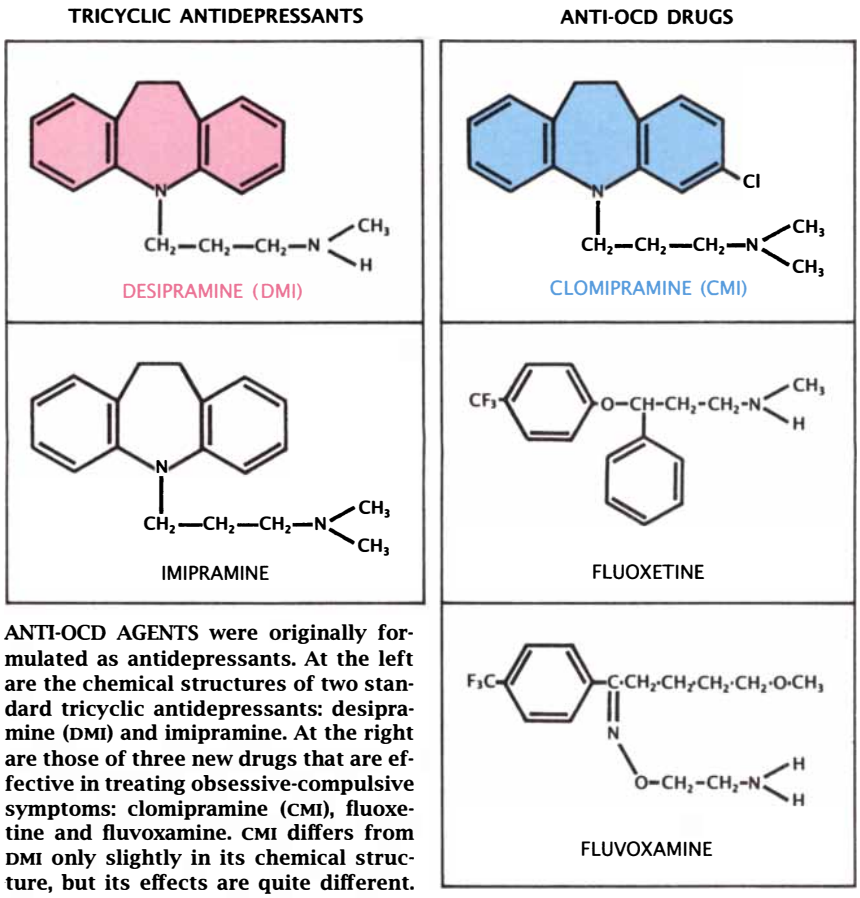
Two very different types of treatment may be effective in dealing with OCD. The first is behavior therapy, which entails repeated exposure of the patient to the stimulus that sets off the ritualistic acts. For example, if a patient has a compulsion that causes him to wash his hands 20 or 30 times a day, his hands may be deliberately dirtied, after which he is prevented from washing them. Although such treatment may sound cruel, it has proved to be effective in severe cases in which traditional forms of psychotherapy had failed. Behavior therapists, including Isaac M. Marks of the Maudsley Hospital Medical School in London and Edna Foa of the Medical College of Pennsylvania, have documented such improvement extensively and shown that exposure to the feared situation is crucial for recovery.

At first glance it may seem contradictory to claim simultaneously that OCD has a strong biological basis and that behavioral conditioning is effective in reversing it. Yet the contradiction is largely superficial. Ethologists have shown that many fixed action patterns in animals, which stem in part from hardwiring, can be extinguished by repeated training. Moreover, since the brain is both a biological organ and the recipient of sensory and psychological inputs, it is only to be expected that strictly psychological causes can have biological effects.

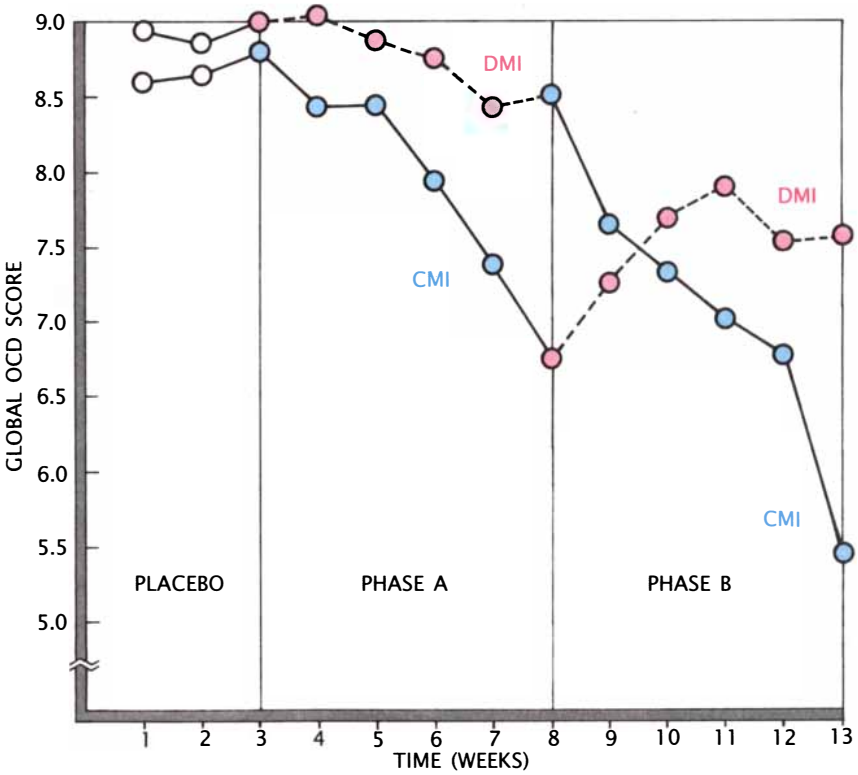
Behavior therapy seems to be more effective in treating compulsions than in treating obsessions, which generate fewer outward signs. The new drug treatments for OCD, on the other hand, seem to be effective in lessening both obsessions and compulsions. Three drugs have been shown to have anti-OCD effects; in addition to CMI, they are fluvoxamine and fluoxetine. All three of these were formulated initially as antidepressants. Indeed, CMI differs functionally by only a single chlorine atom from a standard antidepressant called desipramine (DMI).

It might be thought that the antiobsessional effects of these new agents are simply by-products of their potency in lifting depression, but that does not seem to be the case. Most antidepressants are not particularly helpful in the treatment of obsessive and compulsive symptoms. Furthermore, the benefit seen among our patients in treatment does not depend on whether or not the patient is depressed; it seems to be an independent clinical consequence. A new psychiatric drug treatment for OCD has been established.

Of the three new drugs, CMI is the one that has been most extensively studied. In one recent study the anti-OCD effects of CMI and DMI were compared by a group of my colleagues led by Henrietta Leonard. After two weeks of receiving a placebo, two groups of patients were given either the new drug CMI or the standard antidepressant DMI for five weeks. At the end of that period the groups "crossed over" and the members of each group received the other drug for an equal



ANTI-OCD AGENTS were originally formulated as antidepressants. At the left are the chemical structures of two standard tricyclic antidepressants: desipramine (DMI) and imipramine. At the right are those of three new drugs that are effective in treating obsessive-compulsive symptoms: clomipramine (CMI), fluoxetine and fluvoxamine. CMI differs from DMI only slightly in its chemical structure, but its effects are quite different.



"CROSSOVER" STUDY carried out by the author compared DMI and CMI in treating OCD. Two treatment groups were given one of the two drugs and were then "crossed over" to the one they had not received initially. CMI is effective in treating obsessive and compulsive symptoms; the standard antidepressant seems to have little effect.

period. The results of the crossover study were striking: patients improved on CMI and relapsed on DMI; those who were given DMI first improved only after receiving the newer drug.

To date 14 double-blind studies demonstrating CMI's antiobsessional efficacy have been published. As a result of such studies along with multicenter trials by the CIBA-Geigy Corporation, the drug's manufacturer, and our own multicenter clinical trials, CIBA-Geigy has received permission from the U.S. Food and Drug Administration to provide the drug to patients under a special, expedited license. (Only four drugs have had such handling by the FDA, the first of them being AZT, the anti-AIDS drug.)

CMI works. Why? The answer is not as clear as the drug's efficacy. The effectiveness of CMI (along with fluvoxamine and fluoxetine) is probably related to the physiology of the neurotransmitter serotonin. The role of serotonin in human behavior is not well understood, but there is evidence that the transmitter is important in suicide, appetite and the control of aggression, among other functions. Neurons that respond to serotonin are widely distributed in the brain; they are present in the frontal lobes and they are particularly concentrated in the basal ganglia.

Like other neurotransmitters, serotonin is released into the synapse (the gap between two nerve cells) and later must be removed from the synapse—in the process called reuptake—before the presynaptic cell can be fired again. CMI, fluvoxamine and fluoxetine all block the reuptake of serotonin in the synapse, and work by my colleague Dennis L. Murphy suggests that this may be the reason they are effective against obsessive-compulsive symptoms. Yet although serotonin or its metabolic product can be found in the blood and spinal fluid, no one has shown that OCD patients have abnormal levels of these substances there; evidence for the role of serotonin in the disease remains indirect. What is more, CMI is known to affect the action of other transmitters (particularly dopamine), and the precise mode of action of CMI remains to be elucidated.

In spite of such gaps, what is known is sufficient to formulate a hypothesis about the possible biological basis of OCD. It seems possible that latent behavioral patterns stored in the basal ganglia are somehow triggered by abnormally functioning inferior frontal lobe areas. The initiating impulse

is conveyed to the basal ganglia by pathways mediated by serotonin. Successful drug treatment might alter the role of serotonin in those pathways, thereby damping the spark from the frontal regions.

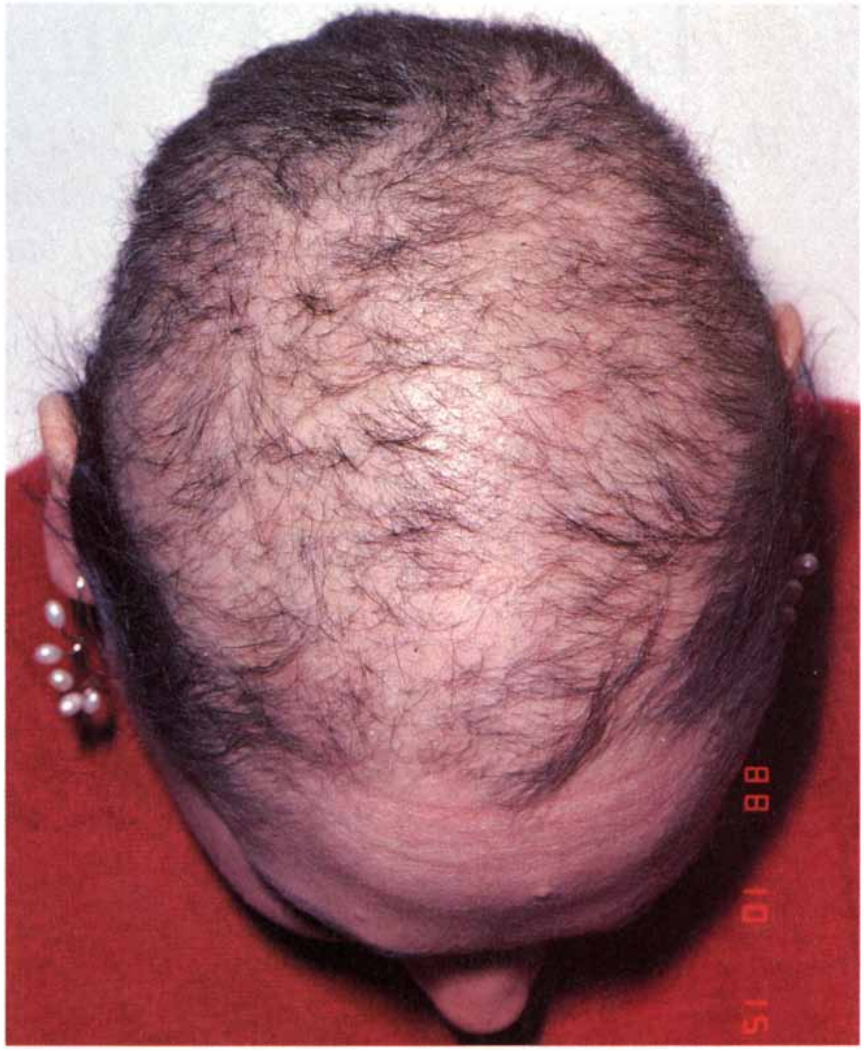
The successful treatment of severe obsessive-compulsive disorder led us to try CMI as a mode of therapy in other conditions that may be biologically related. After our studies were covered on a national television program in March, 1987, we were approached by thousands of patients, among them many women with only a single compulsive symptom: they pull out their hair one strand at a time. This problem, known to physicians as trichotillomania, had not generally been linked to OCD. It may be quite severe; many women who suffer from it have never been seen by their families without a wig.

Following some initial trials, 14 women with this symptom were treated in a double-blind comparison of CMI and DMI resembling the one carried out among our OCD patients. CMI reduced the hair-pulling habit; DMI, the standard antidepressant, was not helpful. PET-scanning studies of the trichotillomaniac patients by Schapiro, Grady and Swedo are now under way, and we suspect they will reveal the same pattern of abnormalities in the frontal lobe–basal ganglia circuit that was seen in patients with the more easily recognized obsessive-compulsive disease.

The selective and successful treatment of trichotillomania raises questions about some other behaviors that are undesirable but subjectively compelling. Might nail biting, for example, respond to the same therapy? How about other "uncontrollable" impulse disorders such as kleptomania? Clinical trials in which such disorders are treated with antiobsessional drugs are waiting to be undertaken. Perhaps the new biology of OCD will one day lead to a reclassification of some other disparate behaviors under a broader category of compulsive syndromes linked by similar patterns of inheritance, by similar brain-imaging results and by the fact that they respond to specific therapeutic agents.

It is my feeling that in the even longer run this work could lead to a fundamental shift in the understanding of evolution, the mind and human rituals. In our model the basal ganglia are the repository of units of behavior that have been organized over the course of evolution. The objects of phobias have been shown to include threats to humanity that are significant from an evolutionary viewpoint (such as snakes, spiders and heights). In much the same way, I believe the formation of obsessive-compulsive rituals may be interpreted from an evolutionary and ethological viewpoint. It is clear that cleanliness, grooming and the checking of order and territorial boundaries had crucial functions during human evolution.

Perhaps under stress these fixed software packages become coupled with certain stimuli that are perceived as dangerous. Once set in motion, this loop cannot be interrupted: the patient becomes an ultimate skeptic who cannot credit his sense data or his attempts to refute the obsession by means of logic. The sufferer cannot accept reassuring information, such as the fact that the door is locked or the light is off. In effect, the patient's inability to believe sets off the ritualistic behavior that appears to be hardwired into the brain at an unexpected level of complexity. Although such thoughts are speculative, they may lead to the profoundest scientific contribution of the work on OCD: the development of a biology of doubt and ultimately a biology of belief.



TRICHOTILLOMANIC PATIENT is compelled to pull out her hair, one strand at a time. Early results from the NIMH show that trichotillomania can be treated with anti-OCD agents. Ultimately a variety of other conditions that respond to such drugs may be reclassified according to their underlying relation to obsessions and compulsions.

FURTHER READING

OBSESSIVE COMPULSIVE DISORDERS. Edited by Michael A. Jenike, Lee Baer and William Minichiello. PSG Publishing Company, Inc., 1986.

OBSESSIVE COMPULSIVE DISORDER IN CHILDREN AND ADOLESCENTS. Edited by Judith L. Rapoport. American Psychiatric Press, Inc., 1988.

THE BOY WHO COULDN'T STOP WASHING: THE EXPERIENCE AND TREATMENT OF OBSESSIVE-COMPULSIVE DISORDER. Judith L. Rapoport. E. P. Dutton, 1989.

Lessons of *Sunraycer*

The solar car's goal was to win the World Solar Challenge race. Its longer-term benefit may be in getting people to think about practical alternatives to fuel-burning vehicles

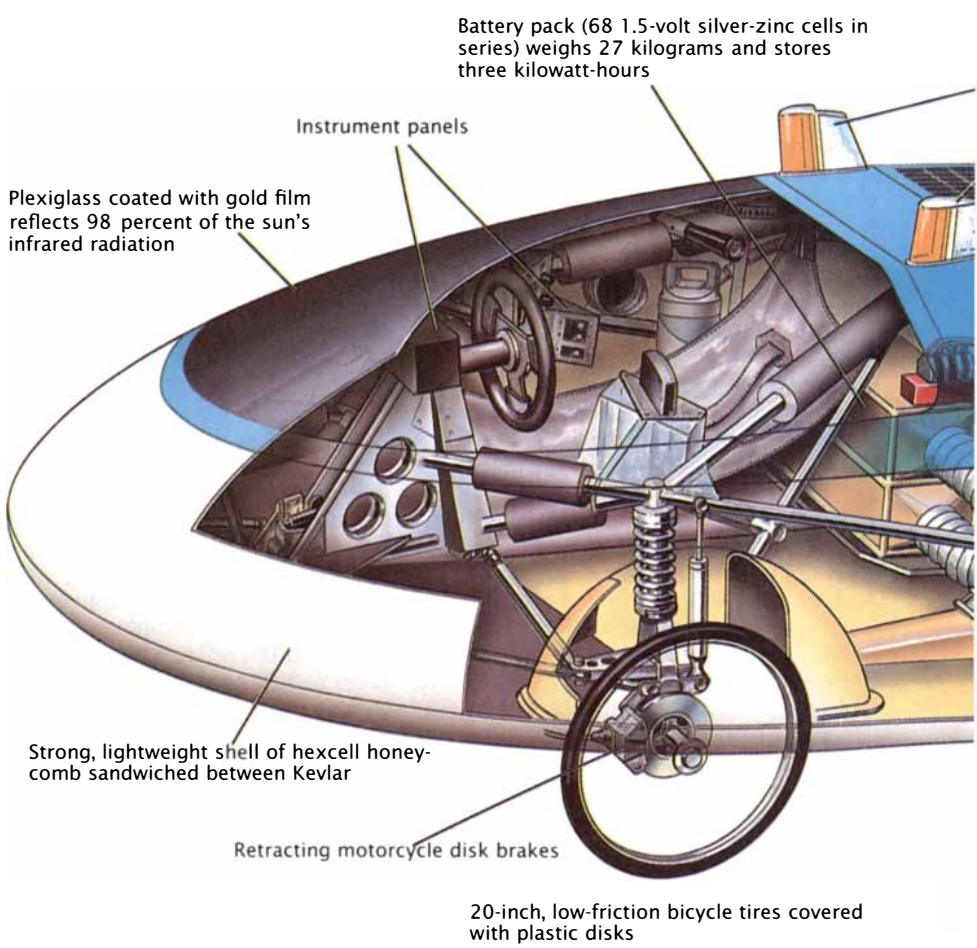
by Howard G. Wilson, Paul B. MacCready and Chester R. Kyle

The recently paved Stuart Highway slices through Australia's barren outback, spanning the continent from sea to sea. For days as we traveled south we had found the highway deserted, but now as we approached the city of Adelaide crowds lined the road to watch *Sunraycer* win the 1,867-mile Pentax World Solar Challenge race. Powered by sunlight shining on the thin layer of solar cells tiling its back, *Sunraycer* completed the journey from Darwin to Adelaide in 44 hours 54 minutes of running time at an average speed of 41.6 miles per hour. We had crossed the continent on an amount of solar energy equivalent to what an automobile engine delivers from about five gallons of gasoline.

Sunraycer had set out six days earlier, on November 1, 1987, from the coastal city of Darwin along with 22 other solar-powered cars. The competing cars, many with saillike solar panels designed to track the sun, resembled a glittering, space-age flotilla. Nine were from Australia, four from Japan, four from the U.S., three from Germany and one each from Denmark, Pakistan and Switzerland. The second-place finisher, entered by the Ford Motor Company of Australia Ltd., arrived on November 9, 23 hours of running time behind *Sunraycer*. In all, 14 cars completed the race, the last one on December 2.

The idea for the race had originated with Hans Tholstrup, an Australian adventurer and visionary, who was an outspoken advocate for fuel conservation, renewable energy and reduced air pollution. In 1983 Tholstrup built and drove a pioneering solar car 2,566 miles from Perth to Sydney in 20 days. Although his car averaged only 12 m.p.h., he was encouraged to create a competition that would stimulate global interest in the inexhaustible and nonpolluting resource of sunshine. The prize offered to the winner was a trophy of gold and silver, donated by the Broken Hill Associated

HOWARD G. WILSON, PAUL B. MACCREADY and CHESTER R. KYLE worked together on *Sunraycer*. Wilson, a vice-president at the Hughes Aircraft Company since 1984, was overall program manager. He has a B.S. and an M.S. in electrical engineering from the University of California, Berkeley, and joined Hughes in 1949, where he has worked on airborne radar, the lunar soft lander and synchronous communications satellites. MacCready is president of AeroVironment Inc., which he founded in 1971. He led the team that built the human-powered *Gossamer Condor* and *Gossamer Albatross,* the solar-powered aircraft *Solar Challenger* and a battery-powered flying replica of the giant pterodactyl. He got a Ph.D. in aeronautics in 1952 from the California Institute of Technology. Kyle, adjunct professor of mechanical engineering at California State University, Long Beach, designed *Sunraycer*'s wheels, tires and brakes. He earned a Ph.D. in engineering from the University of California, Los Angeles, in 1969.



DESIGNED TO WIN, *Sunraycer* integrates a host of efficiency-boosting features that enable it to go surprisingly fast on a small amount of power. The car is six meters

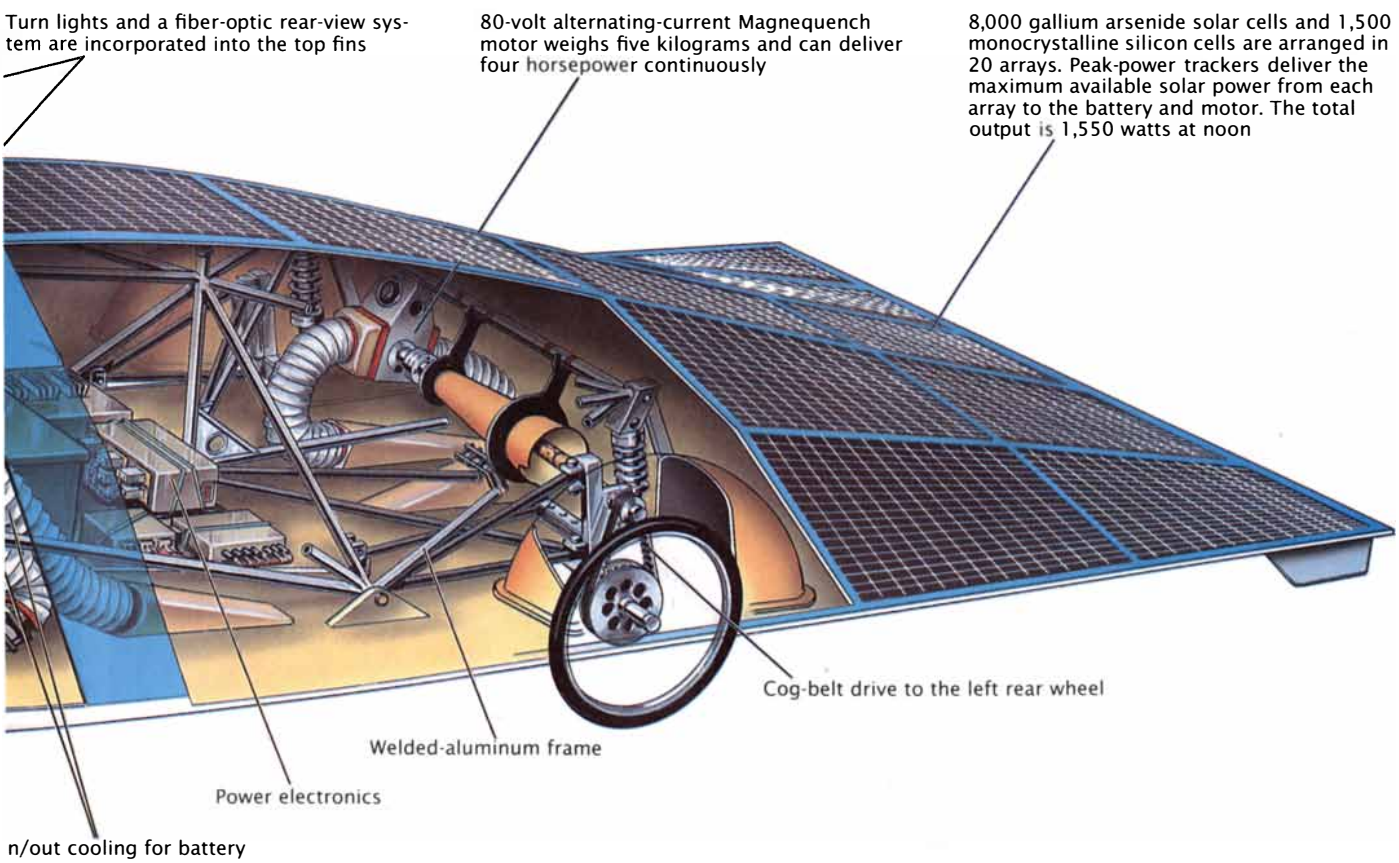
Smelters Pty. Ltd. The basic rules were simple: the cars were to be powered solely by energy from the sun. Tholstrup did not specify how solar energy was to propel the car, but all the successful entrants depended on electric motors energized by solar cells and assisted by batteries charged by the cells.

When we designed *Sunraycer,* our efforts were focused on one specific result: winning the race. But the publicity generated by the Solar Challenge has also encouraged serious discussion about alternative, nonpolluting vehicles, and in particular about the potential of practical battery-powered cars suitable for urban transportation needs. By requiring the cars to rely entirely on a weak power source, the race forced us to push the limits of design. Many of *Sunraycer*'s winning features made use of innovations and state-of-the-art technology that may also be applicable to battery-powered cars. Perhaps the greatest value of *Sunraycer* will be to change people's thinking: to make them seriously consider ways to design practical battery-powered cars that can compete in the marketplace with conventional ones.

We became involved in the Solar Challenge race after Roger B. Smith, chairman of the General Motors Corporation, passed Tholstrup's invitation to participate in the race on to GM's subsidiary, the Hughes Aircraft Company, where it reached one of us (Wilson), whose new job as Vice President–General Motors Programs was to mesh the technologies of the two companies. AeroVironment Inc., a small, creative firm that had developed a pioneering series of low-powered vehicles, was asked to join. Members of the company had built the human-powered airplanes *Gossamer Condor* and *Gossamer Albatross,* as well as a solar-powered aircraft that had flown from Paris to England and the *Flying Fish* human-powered hydrofoil [see "Human-powered Watercraft," by Alec N. Brooks, Allan V. Abbott and David Gordon Wilson; SCIENTIFIC AMERICAN, December, 1986]. The resources and expertise of the two giant corporations, together with the agility and focusing ability of tiny AeroVironment, turned out to be a winning combination.

The race rules stipulated that the vehicle had to fit within a volume six meters long, two meters wide and two meters high, with a minimum height of one meter. The solar collectors had to fit within four meters of the car's length. The driver, with ballast if needed, had to weigh at least 187 pounds. Each car had to have brake lights, turn lights and a rear-view capability, and it had to meet the Australian Department of Transportation's legal standards for highway vehicles. These simple rules left the door wide open for creative and imaginative designs. Indeed, the vehicles entered in the race displayed a wild variety of shapes.

Most of the competitors sought to maximize power output with solar panels that tilted to track the sun or with large, fixed side and top panels that were mounted on the car bodies. These panels increased aerodynamic drag because of the increased height of the vehicle, the large exposed surfaces and the flow interference be-



long, two meters wide and one meter high and weighs 397 pounds without the driver. It recently set the world speed record of 48.7 miles per hour for a solar-powered land vehicle. Assisted by its battery, *Sunraycer* has exceeded 80 m.p.h.

tween the panels and the car body. They also tended to be vulnerable to crosswinds, and their greater mechanical complexity made them heavier and less reliable.

In developing *Sunraycer* we tried to avoid preconceived notions about what might work best. Instead we designed for simplicity, efficiency, light weight and reliability. In the initial three-week planning stage, begun in March of 1987, we quickly considered the tradeoffs involved in a dozen different configurations of solar-panel shape and orientation, body contour and the location, number and structural support of the wheels. (We explored almost every vehicle configuration that finally showed up at Darwin.)

We determined that in order to build a fast solar car all systems would have to be superefficient. The most important factors were to achieve high power output, low aerodynamic drag, low weight, low rolling resistance and high reliability, stability and safety. Our low-drag aerodynamic shape required multiple curved surfaces for the solar panels, which eventually took the form of five gently curved facets. This streamlined, unified shape meant that we had to sacrifice some 10 to 15 percent of the solar energy available each day, but we figured that the gain in aerodynamic efficiency would more than make up for the loss in solar energy. As a bonus our concept happened to result in a sporty look, like something James Bond might drive.

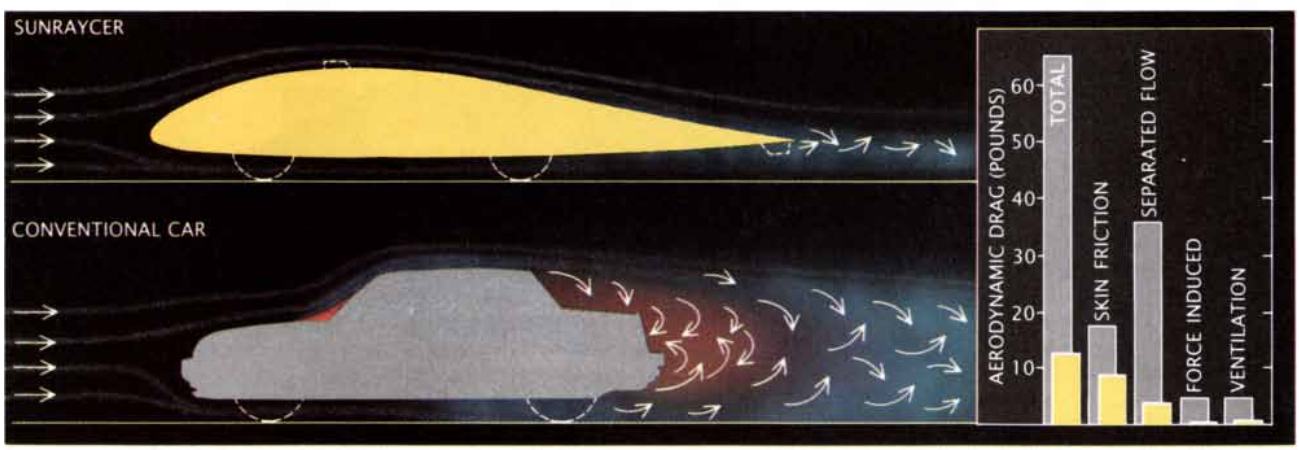
Hughes Aircraft made a major contribution with the design and construction of the solar-cell array. At first we chose the same silicon solar cells that Hughes employs for its communications satellites. As the actual race vehicle was being built, though, we decided to compensate for the loss of array efficiency imposed by the car's shape by switching to better solar cells. Gallium arsenide cells yield from 20 to 30 percent more power for the same amount of sunlight and so we incorporated them into *Sunraycer* to achieve the highest power possible with the streamlined body shape.

We gave high priority to *Sunraycer*'s ability to withstand crosswinds. We were particularly worried about what would happen if the vehicle, traveling at high speed, were to encounter strong side gusts from natural winds or passing traffic. The enormous three-trailer trucks, called road trains, that ply Australia's outback are 170 feet long and weigh 150 tons, and the gusts they generate, particularly on the downwind side, could easily cause a small vehicle to lose control. In qualification tests before the race, all 23 entries had to pass a road train going 70 m.p.h. in the opposite direction. Such tests, however, did not probe extreme conditions.

We were concerned that even small crosswind effects might prove troublesome because of *Sunraycer*'s light weight. Consequently much of the early testing of a quarter-scale model in the California Institute of Technology's wind tunnel focused on crosswind effects. Wherever air speeds up to go around a body, air pressure is reduced and the body experiences an outward force. *Sunraycer*'s basic shape balances the lift over the top with the down force over the bottom for a net vertical force of about zero.

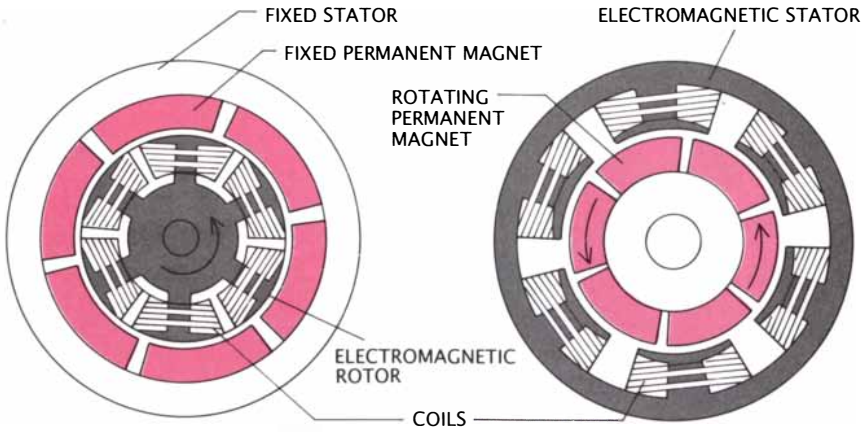
In crosswinds the air tries to flow laterally across the convex bottom, but the wheels redirect the flow lengthwise. The flow bends less sharply in this direction and so has less speedup and therefore less down force. Two small fins on the roof provide a somewhat compensatory effect on the top, minimizing the net vertical force on the rear of the vehicle, which might otherwise lift the rear wheels enough to allow skidding. The fins offered other benefits: we placed the turn signals in them, along with a fiber-optic remote rear viewer. Thus we eliminated external irregularities that would have increased the drag. To improve directional stability in crosswinds, we also placed a series of small vanes underneath at the rear.

Only four and a half months after beginning the project we had a complete working prototype of the car, which we tested in the GM wind tunnel. The prototype had the lowest aerodynamic drag coefficient (.125) for a land vehicle ever measured at that facility. Most automobiles today have



AERODYNAMIC CONTOUR lets *Sunraycer* slip through the air leaving little disturbed flow in its wake. (Flow lines are shown for the center plane only.) The graph shows the aerodynamic drag from four types of flow for a vehicle speed of 55 m.p.h. Aerodynamic drag is proportional to speed squared; the power consumed, drag times the speed, is proportional to speed cubed. *Sunraycer* expends 1.84 horsepower to overcome the drag. A conventional car expends 9.52 h.p. The front third of the *Sunraycer* has a smooth boundary layer (*blue*) and contributes less than 10 percent of the total skin friction. In a conventional car the boundary layer becomes turbulent at the sharp corners at the front. The rough underside makes the turbulent layer grow rapidly. Separated flow (*red*) breaks away from the object and forms eddies, which mix with the outer flow. For *Sunraycer* separated flow behind the wheels and wheel wells causes about a third of the total drag. The much larger drag for a conventional car comes mainly from separated flow in the rear and behind the wheels. Induced drag results when local downward flow lifts the vehicle. *Sunraycer* has no aerodynamic lift and so has no induced drag. At 55 m.p.h. a conventional car would experience 50 pounds of lift and five pounds of induced drag. Internal ventilating flows are also a source of drag.

drag coefficients that are more than twice as high. The aerodynamic engineering of cars is well understood, but the development of *Sunraycer* served to clarify some key points; for example, the large drag from the underside of a car can be greatly reduced by streamlining the bottom. Wheel wells also add substantially to drag and could be improved. Such projecting surface features as joints, rearview mirrors and hood ornaments also increase drag. On the other hand, practical cars must compromise efficiency for certain considerations of comfort and style. For example, in order to provide a comfortable interior it is necessary to increase the car's height, which increases aerodynamic drag. To make a car safe on highways it must have a heavier body structure.



BRUSHLESS ALTERNATING-CURRENT MOTOR provides a lightweight, highly efficient motor for *Sunraycer*. In a conventional direct-current, permanent-magnet electric motor (*left*) a coil-bearing rotor spins between stationary magnets; mechanical brushes switch the d.c. voltage from the power supply in order to maintain a unidirectional torque on the coils. In the *Sunraycer* motor powerful magnets are mounted on the rotor instead and the coils are placed on the stator. The switching is done electronically in a separate motor driver. The motor weighs less, heat losses from brush electrical resistance and friction are eliminated and the coils are easier to cool.

One of the major innovations in *Sunraycer* was its lightweight Magnequench brushless motor, developed by Nady Boules and his colleagues at the GM Research Laboratories. The heart of the motor is a group of Magnequench magnets. This powerful new type of permanent magnet is made from a compound of neodymium, iron and boron, which is melted and then quenched extremely fast in a process developed at GM. GM's Delco Remy division in Anderson, Ind., now manufactures the materials for a variety of applications.

The Magnequench motor weighs only 11 pounds and is 92 percent efficient at the rated load. Although it delivered an average of only about one horsepower over the course of the race, it could provide 4 h.p. continuously and as much as 10 h.p. for short bursts. This made it possible to climb hills and accelerate with a single-ratio transmission. In addition to the "shiftless" transmission, the drive train was unique for a four-wheel vehicle: the power was delivered to only one rear wheel, eliminating the losses associated with a differential.

In conventional direct-current, permanent-magnet motors, mechanical brushes switch the d.c. voltage from the battery in order to maintain a constant, unidirectional torque on the spinning coils. In the *Sunraycer* the motor has an electronic inverter, or motor driver, which exploits advances in new semiconductor switches. The inverter switches (opens and closes) the circuit for variable intervals to modulate the amount of direct current coming in from the battery.

In order to synthesize a sinusoidal-voltage alternating-current waveform from the constant-voltage d.c. battery, the device switches hundreds of times during a single a.c. cycle, controlling the magnitude and direction of current flow. There is very little power loss when the switch is fully on or fully off, and because the new switches are extremely fast, the losses during switching are so small that overall efficiency is very high. Earlier electric cars were limited by poor inverters. The new, fast-switching electronics have decreased this power loss and can now attain efficiencies of from 96 to 98 percent over a wide range of torques and shaft rotation rates. If there is a future for electric cars, it will be largely because of such improvements in electronics.

An element of *Sunraycer*'s efficiency was its use of "regeneration": most of the braking was accomplished by operating the motor as a generator and recycling about half of the energy through the battery instead of dissipating it as heat in brake disks. During this process the inverter works in reverse, converting the a.c. energy from the motor into the correct d.c. voltage for returning energy to the battery. Thus a future electric car, instead of requiring a large and expensive battery charger at a servicing location, can recharge its battery with its onboard inverter.

The basic simplicity of electric cars is most appealing. The motor has only one moving part. The transmission has a single ratio. Most of the braking is done by the motor and much of the braking energy can be recycled. There is no need for a clutch, because the motor stops when the car stops. This eliminates the wasted fuel and pollution generated by idling engines, a particular benefit in cities that suffer from constant traffic gridlock. A motor for a practical electric car is estimated to weigh about 100 pounds and deliver 100 h.p., which would provide a sprightly performance.

When one considers applying *Sunraycer*'s technology to urban transportation needs, it is immediately evident that solar power is an impractical source of energy. A practical solar car would need a battery large enough to complete a trip even if the weather turned cloudy or night came on. It would be better to eliminate the solar cells and charge the battery from an electric outlet.

A key part of *Sunraycer*'s electrical system was the silver-zinc batteries. The cells, built by Eagle-Picher Industries, Inc., of Joplin, Mo., and designed and assembled into battery packs by Hughes Aircraft, had three kilowatt-hours of storage capacity and weighed only 60 pounds. Lead-acid batteries capable of storing an equal amount of energy weigh about four times as much. The top competitors all turned to silver-zinc batteries because of their high energy density.

All the teams had problems managing their batteries. Each team had to develop a race strategy that would make the best use of the limited energy available from the battery. The difficulty of the task was compounded by the fact that silver-zinc batteries can

STUDENTS at the Terraset Elementary School in Reston, Va., are intrigued by *Sunraycer* during a visit last October. A member of the *Sunraycer* team answers their questions about the car. The car has been the catalyst for a nationwide education program to interest students in energy conservation, science and engineering. The program is sponsored by GM.

be damaged if they are allowed to discharge completely. There is no easy way to determine the charge level with simple instruments; the best method is to measure ampere-hours coming in and going out. We paid considerable attention to managing our batteries, with successful results, but several other teams had severe problems because of measurement errors.

The silver-zinc batteries served us well in the race, but they have disadvantages that make them unsuitable for a practical vehicle. In addition to the drawbacks mentioned already, silver-zinc batteries are expensive, and they can be recharged only about 15 times before they lose much of their capacity. Practical electric cars would require a durable, inexpensive and easy-to-maintain battery with a high energy density. To win acceptance as an urban delivery or commuting vehicle, an electric car would have to be able to travel in the range of at least 75 to 150 miles without need for recharging the battery. Because rechargeable batteries wear out eventually and must be replaced, they cannot be too expensive either.

There is now a worldwide effort to develop better batteries. One of the most appealing candidates is the sulfur-sodium battery. Its energy density is four times that of lead-acid batteries; what is more, its ingredients are inexpensive and it intrinsically does not degrade with time or use. At present the ceramic electrolyte components of sulfur-sodium batteries are not reliable, but if that problem could be overcome, sulfur-sodium batteries could probably power a car over a range of between 150 and 200 miles and last for 120,000 miles or more. The most significant drawback of sulfur-sodium batteries is that they must be kept at a constant temperature of 350 degrees Celsius. Thermos-like containment systems have been developed that lose only about 100 watts and can contain the molten contents completely and reliably—an essential safety feature because of the high temperature and extreme chemical reactivity of the material.

There exist today Canadian and German test vehicles on the road that have sulfur-sodium batteries encased in such containments. The Ford Motor Company's electric-van project, sponsored by the U.S. Department of Energy, also incorporates a sulfur-sodium battery. Hughes Aircraft is developing sulfur-sodium batteries for spacecraft and also has an eye to future applications in electric vehicles.

In order to provide adequate range at an affordable cost, the first practical electric commuter cars will probably contain a lead-acid battery weighing nearly 1,000 pounds—about 50 percent of the total weight of a small car. Hence there is a great incentive to reduce the weight of the rest of the vehicle as much as possible. Low overall weight is advantageous in several ways. It requires less energy for acceleration, makes possible higher climbing speed on hills and lowers rolling resistance in the tires.

By relying on lightweight materials and components in all systems we were able to hold *Sunraycer*'s weight to only 397 pounds (without the driver). The aluminum-tube "space frame," a cagelike body frame employed in racing cars, weighed only 15 pounds and yet supported the weight of the rest of the car and the driver—584 pounds in all. The body shell, constructed from a sandwich of two plas-

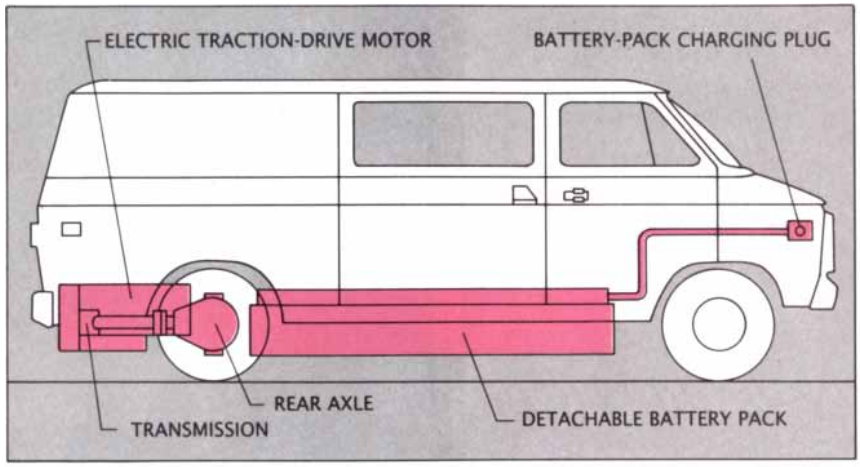
tic materials, Kevlar (used in bullet-proof vests) and Nomex honeycomb, provided a rigid, lightweight body with a smooth aerodynamic surface.

The Kevlar-Nomex sandwich is very strong for its weight and is widely used in racing cars, spacecraft and for other specialized applications. It is much too expensive and difficult to fabricate, however, for use in a commercial automobile. Of the various choices available for commercial electric cars, aluminum appears to be particularly attractive. Aluminum bodies would have to be thicker than steel bodies to provide enough strength, but they would still be lighter than steel. The front and sides in particular would have to be designed to provide protection against collisions. Factories that currently produce steel frames and bodies could be converted to aluminum with relative ease.

Tire rolling resistance is decreased with lower wheel loads, with thinner, more pliable tread and casing and with higher tire pressure. *Sunraycer* rolled along on slick bicycle tires pressurized to 100 pounds per square inch. In the race we actually employed worn tires whenever possible because the thinner tread further diminished the rolling resistance. For a given vehicle weight, bicycle tires have about a third the rolling resistance of standard automobile tires and about half that of motorcycle tires. The top four cars all rode on bicycle tires, but the four Japanese vehicles and several others had heavy motorcycle tires, thereby paying a double penalty in weight and high rolling resistance.

Since the most recent energy crisis, the fuel mileage of passenger cars has improved significantly. Smaller cars and efficient engine management are given much of the credit, but actually a large fraction of the improvement came from the development of tires with lower rolling resistance. Further improvement is possible, but the higher tire pressures that would be required would decrease the smoothness of the ride. This tradeoff may be acceptable for electric vehicles.

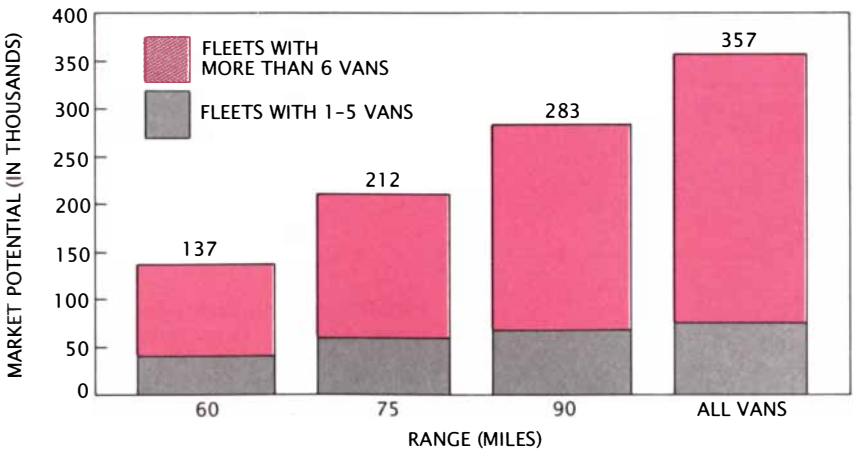
The Solar Challenge race finished near Adelaide in the beautiful Borossa Valley, at the 100-year-old Seppeltsfield Winery. In the library of the ancestral home were several bound volumes of *Scientific American* from the 1890's. Browsing through them, we came across an article describing a marvelous new Columbia electric carriage that would travel at 15 m.p.h. After a brief heyday early in this century, when electric cars outnumbered gasoline vehicles, the electric car dropped out of favor as the gasoline engine was improved, as self-starters were developed and as range and performance became more important. Now, 80 years later, as environmental concerns grow more pressing, electric cars may be ready for a comeback. The perception of electric cars as golf carts puttering about on the grounds of retirement communities will be swept away as new technologies make it possible to develop electric vehicles that can compete effectively with gasoline-powered cars for many transportation needs.



FIRST ELECTRIC VANS would take conventional van models and retrofit them with a detachable lead-acid battery pack, an electronic motor driver and an electric motor.

That vision of the future flies in the face of accepted reason. After all, the energy-storage capacity of batteries is minuscule compared with that of fossil fuel. Enough energy can be extracted from gasoline by a standard motor to lift the initial weight of the gasoline almost 1,000 miles, whereas a lead-acid battery delivers only enough energy to lift its own weight less than 10 miles, or 1 percent as far. Even the most advanced experimental rechargeable batteries can improve the performance by only a factor of four. The low energy density of batteries would appear to doom the electric car. In reality, however, an electric vehicle designed for high efficiency could get much higher mileage per unit of energy, and if it were powered by an inexpensive and durable battery, it would not cost any more to operate than a conventional gasoline-engine vehicle. Moreover, most electric vehicles will be recharged at night, when demand for electricity is low. Electric utilities are now trying to develop electric vehicles as a market for electricity that would not require new plant capacity.

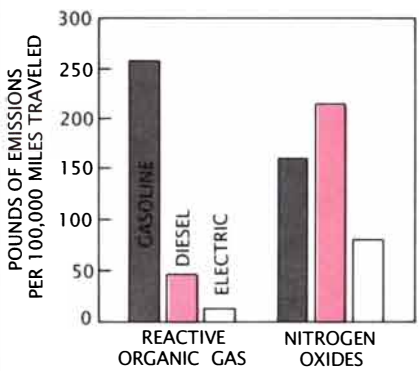
The Electric Vehicle Development



MARKET POTENTIAL for electric delivery vans is shown in this graph, based on a national study by the Electric Vehicle Development Corporation. The market grows sharply as the distance that can be driven without a battery recharge increases. Most of the growth will be in larger van fleets, which account for most vans in the U.S.

SWISS PROTOTYPE shows how a superefficient electric car designed from scratch might look. The car is meant for short trips such as delivering commuters to trains.

Corporation (EVDC) has an active program to introduce electric vans to urban fleets. Such fleets are good candidates for electric propulsion because most individual vans are not driven great distances on a given day, make frequent stops (often with the engine left idling, which wastes fuel and generates pollutants) and are serviced at a central maintenance station. An EVDC survey of the 30 largest markets in the U.S. suggested that 140,000 vehicles could be converted if one assumes that electric vans would have a 60-mile range, which is attainable with present-day lead-acid batteries. The 100-mile range offered by advanced batteries would double the size of the potential market.



EMISSIONS from burning fossil fuel at a power plant to charge batteries for an electric vehicle would be significantly less than tailpipe emissions from conventional gasoline-powered vehicles. Reactive organic gases would be virtually eliminated and nitrogen oxide would be cut by two-thirds. The data apply to converted vehicles. Electric vehicles designed from scratch would be much more efficient and would pollute even less.

In a planned cooperative effort by the EVDC, the Electric Power Research Institute, Magna International, Inc., of Canada, Chloride EV Systems Ltd. of the U.K. and GM, 500 GMC Vandura vehicles will be manufactured with electric power trains and introduced into various regional markets across the country in 1990. These first few G-Vans, as these vans will be called, are expected to sell for less than $30,000 apiece. (Increased production rates would reduce the selling price significantly.) They will cover 60 miles on a single battery charge and attain a top speed of 52 m.p.h. They are equipped with power steering and power brakes, can accelerate from a full stop to 30 m.p.h. in 12 seconds and can carry 2,000 pounds. They will cost 40 cents per mile to operate—comparable to the cost of operating a conventional gasoline-powered van under the same conditions.

A small commuter car, engineered for high efficiency, would consume only about 1.5 cents per mile in electricity. If its battery cost $1,000 and lasted for 30,000 miles, the car would cost a total of 4.5 cents per mile to operate. This would be equivalent to getting 22 miles to the gallon (at $1 per gallon). Government and utilities can help to ease the introduction of such cars into the market by providing economic incentives, such as waived registration fees, and new infrastructure, such as variable-rate electric me-

ters and battery-recharging stations. Utilities might lease batteries and offer discounted rates at night.

Fossil fuel is a convenient, high-energy-density, high-power source of energy, very attractive for propulsion. But it is not replenishable on a time scale relevant to human activity, and it pollutes the air at the site of combustion. Batteries store the energy generated at a central electrical plant and so avoid emitting local pollution, which is an increasingly critical concern in large cities; the pollution emitted at the central plant is less per unit of power output than it is for vehicles with combustion engines. What is more, battery-powered vehicles could as easily get their energy from a nonpolluting, renewable source, such as hydroelectric, solar or wind power, as from conventional fossil fuels. Finally, if vehicles can be made efficient enough to run satisfactorily on batteries, the total amount of energy expended for transportation will be drastically reduced, regardless of the source. Battery-powered cars are not a panacea for the energy and pollution problem, but they can provide substantial improvement and may buy society more time to resolve the fundamental problem of growing global demand for dwindling resources.

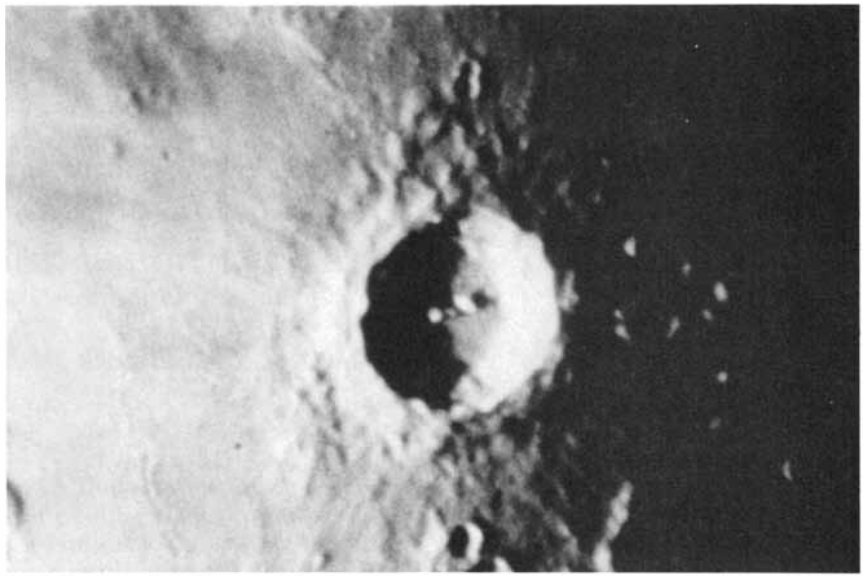
Until recently any effort to design battery-powered vehicles has been inhibited by the assumption that a gasoline-burning alternative can always do the job better, even if fuel prices rise substantially. *Sunraycer* at least demonstrated that a nonpolluting electric vehicle can attain a respectable speed on surprisingly low power. Perhaps the systems-design approach employed so successfully with *Sunraycer,* incorporating advances in aerodynamics, materials, electronics, motors, batteries and tires, will help electric cars to become competitive with cars powered by fossil fuel. No matter what energy sources are exploited in the future, the visionary challenge of the race across Australia has raised expectations for transportation alternatives that make fewer demands on the global environment.

FURTHER READING

ELECTRIC VEHICLES. Sheldon R. Shacket. Domnus Books, 1981.

ELECTRIC VEHICLE TECHNOLOGY. L. E. Unnewehr and S. A. Nasar. John Wiley & Sons, Inc., 1982.

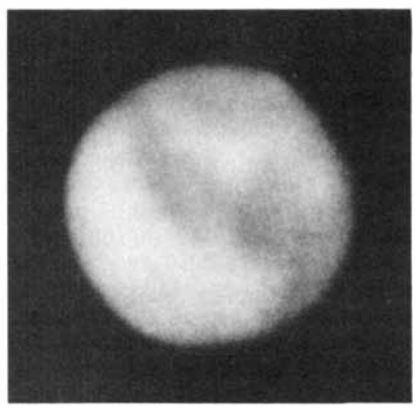
AN ANALYSIS OF THE COMPONENTS OF POWER LOSS IN A SOLAR-POWERED ELECTRIC VEHICLE. D. J. Patterson. Darwin Institute of Technology, Australia, January 30, 1988.



COPERNICUS And MARS: New Questar Photographs Of Old Favorites

Hugh Entrop, whose photographs we have published many times in our literature, photographed Mars with his Questar 7 on October 6, in central Washington at an elevation of 4400 feet. Entrop describes the night as quiet, no air stirring at that elevation but the upper air "pretty bumpy."

He experimented with three films, Ektachrome 200, Kodachrome 64 and Fujichrome 100. His best results were with Fujichrome and Kodachrome and we reproduce here a black and white photograph made from the Fuji 100, which means that what you see here is several processes removed from the original 35 mm. slide. On the original the image of Mars was 3/32", yet the following features are clearly visible on the film and, hopefully, in the magazine reproduction: South Polar Cap positioned at one o'clock on the disk, Mare Chronium to the left, Mare Cimmerium, Hesperia and Mare Tyrrhenum in the central portion, North Polar Hood at 7 o'clock with Sithonius to the left, Syrtis Major near 5 o'clock, and Hellas near the edge midway back to the South Pole Cap.



The moon photo was taken with the Questar 3½ from Entrop's backyard in Seattle and shows a very different view of Copernicus from others in our collection. The crater is right on the terminator with the dark shadow providing a dramatic backdrop and sunlight sharply delineating the peaks rising from the crater floor.

Entrop always comments on the stability and accurate guiding of the Questars, even in a brisk wind which he encounters many times in his photographic locations. He says he sometimes gets his best shots of galaxies and nebulae in areas that are exposed to sudden mountain squalls and gusting winds, in spite of which the Questar keeps guiding steadily along.

A telescope which is so easy to set up and use, and can resolve detail worthy of the performance of a great observatory telescope, could enhance your own observing pleasure. Why don't you move up to Questar?

The Roman Aqueduct of Nîmes

New analyses of the once magnificent water-supply system suggest that the structural and hydraulic engineers who designed it 2,000 years ago were more sophisticated than is generally thought

by George F. W. Hauck

Any warm, sunny day in the south of France will find thousands of sightseers wandering across a 240-year-old bridge spanning the Gardon, or Gard, River near the city of Nîmes. Most of those visitors pay little heed to the sparkling river, the lush Gardon Valley or even the aged structure under their feet. Instead they crane their necks to view a still more magnificent sight: the famous Pont du Gard, a parallel, 16-story-high (48.77-meter) bridge whose three tiers of masonry arches were set in place some 2,000 years ago by workers of the Roman Empire.

In the spring of 1985 I scrambled over the ancient monument like the most intrepid young tourists, but my reasons for visiting the bridge, a major link in the aqueduct that once supplied water to ancient Nîmes—the Latin colony of Nemausus—were more professional than recreational. An engineer, I was evaluating the bridge and other remnants of the aqueduct in order to determine how well various components of this water-supply system were designed.

Conventional wisdom has it that the Roman engineers were empiricists who cautiously modified tried-and-true designs—that they had little understanding of, say, the geometry that would optimize the velocity of water in a conduit or of just how much mass a bridge would need to resist the overturning forces of the wind. They also have the reputation of being excessively conservative, compensating for lack of insight by designing structures to be sturdier and hence costlier than is now known to be necessary.

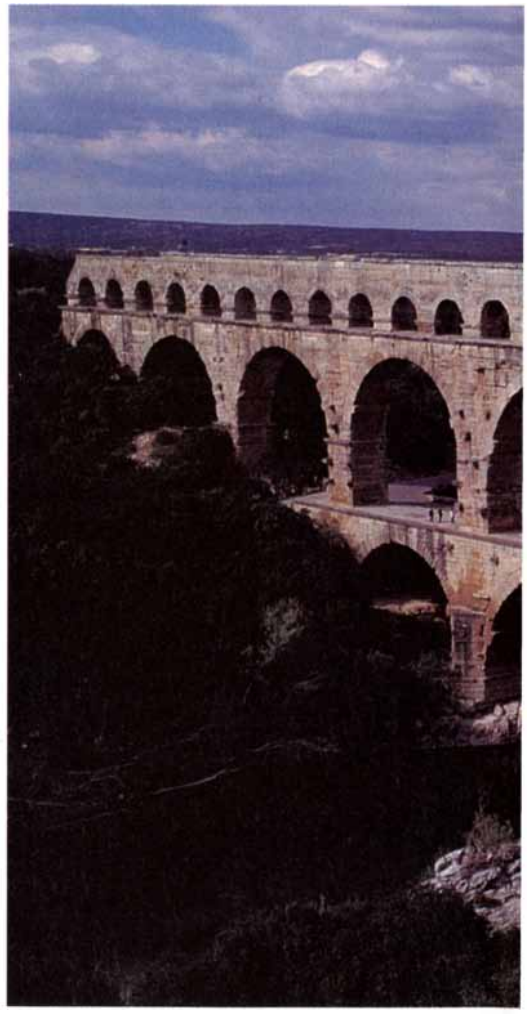
Yet the Pont du Gard, with its extraordinary height and soaring arches, was a dramatic departure from anything the Romans had done before. Such boldness suggested to me that the Roman engineers may have been more knowledgeable and daring than is generally thought. Apparently that is the case. Calculations I did on my own and with Richard A. Novak, a former colleague at the University of Missouri, point to surprising sophistication on the part of the hydraulic and structural designers of the Nemausus aqueduct.

My studies while I was in France were greatly facilitated by the cooperation of Jean-Luc Fiches of the French National Center for Scientific Research. He and other investigators working under his direction helped me to examine the Pont du Gard and other remnants of the aqueduct while they were in the middle of preparing a major photometric survey of those structures. They also allowed me the privilege of assisting them.

The aqueduct we were so carefully measuring in 1985 once carried water by the force of gravity from springs near the hamlet of Ucetia (now Uzès) to a hillside in Nemausus, where a circular "castellum," or collection basin, stood. From there the water flowed to lower elevations through 10 large distribution pipes.

Circumstantial evidence indicates that the aqueduct was built in 19 B.C. by Marcus Agrippa, the lifelong friend and colleague of Caesar Augustus. In addition to being a military and administrative genius, Agrippa was a brilliant engineer and the chief designer of many public works in the Roman Empire, water-supply systems in particular. He is known to have been in or near Nemausus at that time.

By 19 B.C. Nemausus' population of about 50,000 would certainly have had need for an improved water supply. From Syria to Spain, from the North Sea to the Sahara, the people of the empire were enjoying a golden age of civil engineering. The Romans were constructing roads, sewers, bridges, buildings, theaters, stadiums and, of



PONT DU GARD was a major link in the Nîmes aqueduct. It carried water across a deep gorge, the Gardon Valley, and across the Gardon River on the floor of the valley. The bridge is remarkable not

GEORGE F. W. HAUCK, who received a doctorate in structural engineering from Northwestern University, is associate professor of civil engineering at the University of Missouri at Kansas City, where he specializes in the history of his profession. Following the philosopher George Santayana, Hauck says that "engineers who ignore the past can do little for the future."

course, magnificent public baths. The last, a key element of Roman social life, required that clean water be supplied at a rate comparable to the rate in modern cities. The water provided by the local springs and wells would probably have fallen short of the increased demand; only an aqueduct would suffice.

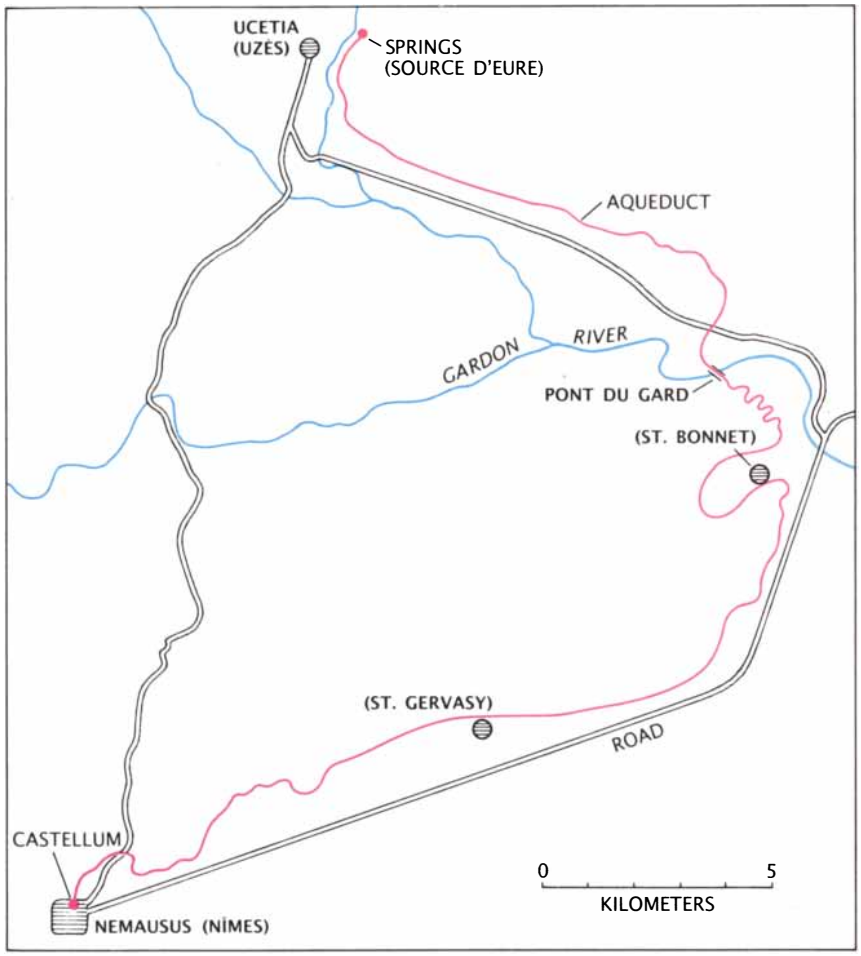
Building an aqueduct for Nemausus was not going to be easy, however; one can readily imagine that the project, with its formidable challenges, was indeed the work of someone of Agrippa's stature. The springs near Ucetia yielded water at a rate that could supply about 600 liters per day for each resident of Nemausus. It was enough. Yet the shortest route from Ucetia to Nemausus, some 20 kilometers as the crow flies, was an obstacle course of hills and gorges that would have required the construction of at least one impossibly long eight-kilometer tunnel; many centuries would elapse before tunnels of that length could be built. A detour to the west was out of the question, again because of hills. The only reasonable solution was a horseshoe-shaped detour to the east.

This route would avoid many hills, but it meant the aqueduct would have to be 50 kilometers (31 miles) long, and it would still have to traverse the deep Gardon Valley, pass through or around rugged outcroppings of stone and cross marshes. One of the toughest problems was presented by the relatively low elevation of the springs: a mere 17 meters (56 feet) above the castellum. While designing tunnels and bridges to contend with the forbidding obstacles, Agrippa and his deputies—whose tools included only primitive levels, abaci, wax tablets and the human brain—would also have to find a way to maintain an unusually slight average decline of no more than 0.34 meter per kilometer in the channel (the 17-meter difference in elevation between Ucetia and Nemausus divided by 50 kilometers of channel). Such a gradient is imperceptible to the eye; a small error could have resulted in flat areas where the water would stagnate.

The Roman engineers managed to do what was required, but I wondered whether they were able to ensure that the water flowed through the channel efficiently, with the least amount of impediment. Such a feat would be truly remarkable, considering that the mathematical formulas engineers now rely on for designing gravity-controlled water pipes were not developed until the 19th century.

With Novak's help I evaluated the

only for its beauty and longevity but also for its size: close to 49 meters (160 feet) tall, the 275-meter-long structure is 16 stories high. The Romans, who never built a taller bridge, achieved the desired height by stacking three tiers. The lowest of them has six arches and is 21.87 meters high and 6.36 meters thick; the middle tier, with 11 arches, is 19.50 meters high and 4.56 meters thick, and the top tier—which holds the massive concrete water channel—is 7.40 meters high and 3.06 meters thick. The largest arches, which are over the Gardon River, span 24.5 meters, the width of a six-lane highway.



ROUTE OF AQUEDUCT (*red*) began at springs situated near the hamlet of Ucetia (now Uzès). From there it followed a 50-kilometer horseshoe-shaped path, traveling first southeast to the Pont du Gard, then southwest, past what are now St. Bonnet and St. Gervasy, to Nemausus (ancient Nîmes). The water poured into a circular castellum, or collection basin, on a hillside in Nemausus and flowed to lower elevations through 10 lead pipes. The shortest route—which would have taken the aqueduct almost directly south to Nemausus—was rejected because large hills stood in the way.

efficiency of the water flow by estimating the typical depth of the water in various parts of the channel. We knew that the ideal depth is equal to half the width of the channel, because at that level the amount of water touching the surfaces of the channel in relation to the volume of water in the conduit (and therefore friction between the running water and the surfaces) would be minimized.

The depth can be determined if one knows the smoothness, the geometry and the slope of the channel as well as the rate at which water flows into it from the source. Fortunately we had information about all these parameters. We knew, for instance, that the channel was quite smooth. As was the Roman custom, it was laid underground wherever possible. Workers excavated a ditch, laid a concrete base, erected walls of stone masonry and applied a layer of pinkish *maltha* concrete: a smooth and impervious liner made of lime, pork fat and the latexlike milk of unripe figs. They then vaulted over the structure with a semicircular stone arch and backfilled the trench.

We also knew that the cross section of the actual water conduit—the floor and the walls (excluding the vault)—was generally an open square measuring 1.2 meters on a side; the total height of the channel, including the archway, was 1.8 meters, just high enough to allow maintenance workers to stand upright after they entered through hatches placed at regular intervals along the channel's route. In places where the conduit was elevated, as on the Pont du Gard, the channel was rectangular; workers made the walls higher and covered them with huge stone plates that were more resistant to weathering than the stonework vaults.

For economic reasons I shall allude to below, the gradient averaged 0.67 meter per kilometer in the leg of the aqueduct above the Pont du Gard but only 0.187 meter per kilometer in the leg below, where it varied from 0.07 to 0.30 meter per kilometer. On the basis of modern hydrological estimates, we assumed that the springs near Ucetia yielded water at a rate ranging approximately from 210 to 450 liters per second and averaging about 350 liters per second.

Neither the designers nor we had to worry about the water depth in the upper leg of the channel. They obviously decided to make the slope there steeper than average to reduce the height of the Pont du Gard and the approach structures as much as possible, thereby minimizing both the expense and the risk of failure. (Even at its reduced height the bridge was the tallest span the Romans had built or would ever build.) They paid a price for the saving in height, however: making the first leg of the channel steeper than average meant the gradient in the leg below the Pont du Gard would have to be flatter than average (which explains why the slope there is particularly slight).

How well did water pass through the critical lower leg of the conduit? It moved quite efficiently, according to our calculations of water depth. We learned that the depth of flow was about optimal (0.6 meter) under several conditions. For example, when the least water was flowing through the channel, the water depth in the flattest segment, just below the Pont du Gard, was ideal; it was also ideal at maximum flow in the steepest (final) segment, just before the channel reached the castellum. The empirically derived formula we applied, known as the Manning formula, further yielded a calculated water level of 0.6 meter for the average gradient (0.187 meter per kilometer) of the entire postbridge leg during times of average flow—a figure that would have been of particular concern to the designers.

We also considered whether the channel was ever in danger of overflowing—whether the water would rise into the area above the 1.2-meter square, causing leaks or crowding out the breathing space of the maintenance workers. We found there was no such danger. Where the channel was flattest—that is, where the velocity of the flowing water would be expected to slow down and the level would be highest—the square part was just

Lufthansa's check-in area is so convenient that you can walk there from your hotel room.

When you fly Lufthansa, you expect quality service on the ground as well as in the air. For that reason, we've installed full-service Lufthansa check-in counters in some of the finest hotels of Central Europe. If you're a Lufthansa passenger staying at one of those hotels, you can buy your ticket, check your bags and pick up your boarding pass right at your hotel.

What other airline shows that much concern for its passengers before they even get to the airport?

People expect the world of us.

Lufthansa
German Airlines

Lufthansa is a participant in the mileage programs of United, Delta, USAir and Continental/Eastern. See your Travel Agent for details.

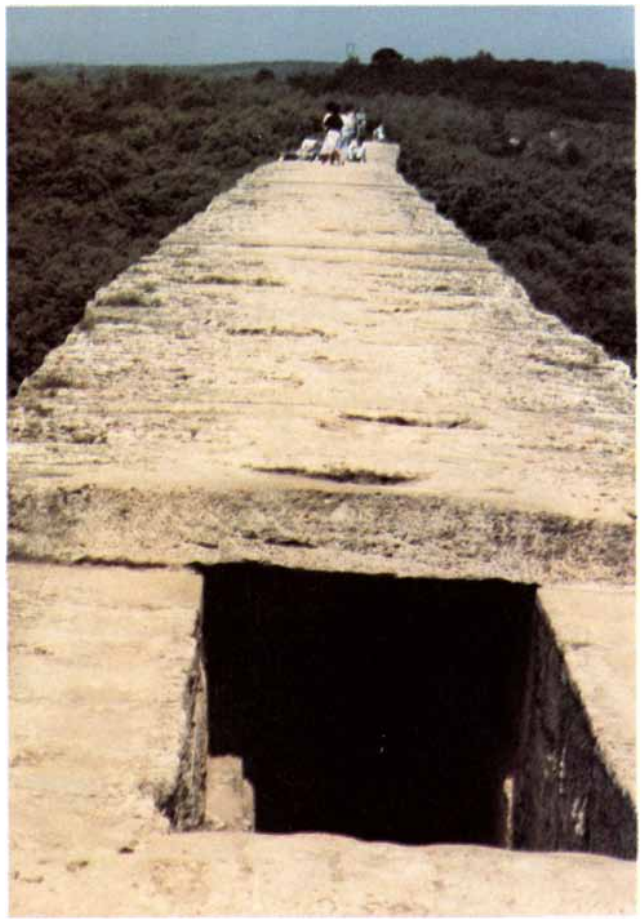
large enough to accommodate the water spewed by the springs at their highest flow. It is difficult to imagine how a channel that had to be reasonably easy to build and maintain could have been designed any better.

If the channel was remarkable, the Pont du Gard is even more so, not only for its lacy appearance, which belies the mass of the six-ton stones of which it was fashioned, but also because it has managed to remain standing all these centuries. True, the span is well balanced: the lower tier is larger and heavier than the middle tier, which in turn is larger and heavier than the top tier, where the channel rests. Yet the valley that the bridge crosses is known for its ferocious winds, and the river below is subject to severe springtime floods. Presumably the engineers had no way to calculate the overturning stresses produced by winds and floodwaters. Such forces are not well understood even today. Did they solve the problem, one wonders, merely by being excessively cautious?

To answer the question I calculated the wind speeds and the associated forces that would be necessary to separate even slightly the masonry joints somewhere in the bridge. The joints are the weakest parts of the structure, and any separation, or cracking, would signal the impending toppling of the bridge. Then I compared those winds with the ones typical of the region.

The process involved estimating the direction and magnitude of the wind forces that might be exerted on a single bent during storms of varying strengths. (A bent is a representative section of the bridge: one arch of the bottom tier and the masonry directly above it on the other tiers.) I then calculated the tension (a lifting stress in this case) the wind forces could be expected to produce on the windward side of the bent, as well as the compressive (downward) stresses the wind would produce on the leeward side [*see illustration on page 102B*]. I also calculated the compressive stresses resulting from the weight of the masonry above any given region, because the compressive stresses from the masonry counteract the tensile stresses produced by the wind, thereby limiting the lifting effect of the wind. This principle is easily demonstrated: a stack of heavy bricks will resist a considerable push from the side, but a stack of Styrofoam bricks can be toppled with a flick of a finger.

I assumed that whenever the tension was greater than the compression, separation of the masonry joints would occur. Such separation can be visualized as the lifting of the windward part of one layer of masonry off the layer below, much as if a wedge of air were inserted between them. The strength of mortar was not an issue because none was used. The calculations indicated that the bases of the piers (the columns supporting the arches) in the middle tier of the bridge are most vulnerable to tension, but only slightly more so than the piers of



SURVIVING SEGMENT of the water channel (*left*), now encrusted with calcium deposits, is typical; most of the conduit ran underground and had a similar shape. The area meant to hold water was usually 1.2 meters square; the total height of the channel, including the vault, was 1.8 meters—just enough to give maintenance workers headroom. Where the conduit was elevated, as on the Pont du Gard (*right*), the walls were higher than 1.2 meters and the vault was replaced by stone plates.

THE LIFESTYLE RESOURCE™

At The Lifestyle Resource we give you all the facts and details necessary to make an informed purchase. Your satisfaction is our primary concern. If your purchase doesn't meet your expectations, return it in original condition within 30 days for prompt refund.

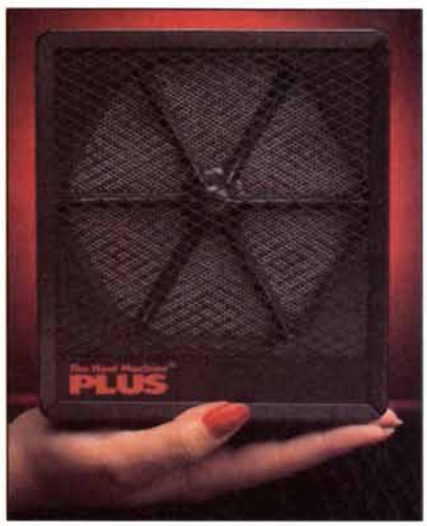
ZONE OF CALM

Noise produces stress in the human organism. Today the volume of civilization nearly everywhere seriously interferes with our abilities to relax, read, sleep, concentrate or otherwise function at optimum efficiency. But you needn't be victimized by noise pollution. The *new* Marsona Sound Conditioner electrically synthesizes a variety of pleasing natural sounds that help mask out and reduce the annoyance of unwanted noise. A 5 inch speaker brings you the sounds of ocean surf, summer rain, mountain waterfalls — as close to nature as you will find. You control not only the volume but also wave pattern, wave or rain rhythm, the seeming nearness or distance of the source. Solid state electronics designed for continuous use. UL listed. Manufacturer's 1 year limited warranty. **$139.95** #2201.

SHARP AS A RAZOR

The Chef'sChoice® Knife Sharpener safely and easily hones the dullest of knives to razor sharpness in minutes. According to *Food and Wine*, the Chef'sChoice Knife Sharpener is "unquestionably the best." *Popular Science* calls it "idiot proof." Its ingenious patented design is the brainchild of former DuPont engineer, Daniel Friel. His unique sharpening method uses diamond abrasives to sharpen a knife in three stages to form a progressively sharper edge. Since magnets hold the knife at the precise angle required for each operation, the guesswork of traditional methods is eliminated. The resulting triple bevel edge stays sharper longer than conventional edges. Only the minimum amount of steel is removed during each stage, making it safe for all high quality steel or alloy knives. World class chef, Pierre Franey, in his *N.Y. Times* column, hailed the Chef'sChoice as "unparalleled, in my experience, for putting an exceedingly sharp and long-lasting edge on knives of all kinds." Two year limited warranty, UL listed, measures 9" x 3½" x 4". **$79.95** #2450.

THE SIX INCH, FIVE POUND FURNACE



Imagine heating an average size room for 24 hours for about $1.00! This ceramic heater not only uses minimal amounts of electricity, it is unique in its size, its exceptional safety, and tremendous heating ability. Two North American companies, GTE and Micromar, have created what we consider the best and safest 1500 watt ceramic heater on the market today — The Heat Machine Plus — and we looked at 12 competitive heaters. GTE spent 18 months of intensive research and development creating, to our knowledge, the *only* ceramic element specifically designed for a portable ceramic heater. This U.S. patented element is manufactured exclusively for Micromar, who was the first company to market portable ceramic heaters in North America. They've built on their experience to design this third generation ceramic heater using the finest components available. Consider some of the features of this remarkable product. SAFETY: This heater operates at a temperature below the ignition point of tissue paper. No fuels, odors, flames or red hot coils — safety grille gives added protection. A "tip-over" switch automatically cycles the heater down if it is turned over. It is designed to run around the clock, is exceptionally safe and great for a child's room. AIRFLOW: The high quality fan is manufactured in West Germany by Papst, who also supplies fans to other top notch companies. The ceramic elements are manufactured using a special pressing process which allows larger holes than any other solid ceramic heater — creating an airflow almost twice as much as some others. That means quicker, more even heat distribution. EFFICIENCY: At full capacity of 1500 watts, The Heat Machine Plus puts out 5120 BTU's per hour. As it reaches the desired temperature you set, electrical consumption drops, thereby reducing your costs. Temperature is controlled by a special thermistor circuit which varies fan speed and eliminates the on-off, hot-cold cycling common to many heaters. VERSATILITY: The Heat Machine Plus goes anywhere, gives instant heat and allows you to lower the heat in other areas of the house and selectively heat those areas you occupy. DURABILITY: The Micromar is built like a tank, with a heavy-gauge metal case, industrial grade components and a meticulous attention to detail. Laboratory tested for over 100,000 on/off cycles. Manufacturer's 5-year limited warranty. Built-in handle and washable air filter. Standard 110 volts AC. UL listed. **$179.95** #2280.

THE LIFESTYLE RESOURCE™

FOR FASTEST SERVICE — CREDIT CARD ORDERS CALL TOLL-FREE 24 HRS A DAY **800-872-5200**

DEPT. SFAC09; 921 EASTWIND DR. SUITE 114: WESTERVILLE, OH 43081

ITEM NO.	QTY.	DESCRIPTION	ITEM PRICE	TOTAL PRICE

SEND TO (PLEASE PRINT)

ADDRESS

CITY STATE ZIP

☐ CHECK OR MONEY ORDER ☐ MASTERCARD ☐ VISA ☐ AMEX SIGNATURE

ACCT. # EXP. DATE

Shipping Charge covers UPS, handling and insurance for guaranteed delivery.
UPS Second Day available for an additional $7.50 per order.

Up to $20	$ 3.95	$50.01 to $60	$ 7.95
$20.01 to $30	$ 4.95	$60.01 to $70	$ 8.95
$30.01 to $40	$ 5.95	$70.01 to $100	$10.95
$40.01 to $50	$ 6.95	Over $100	$12.95

Canadian residents please call (614) 794-2662 for ordering information.

SUB TOTAL	
SHIPPING (See table at left)	
TOTAL	

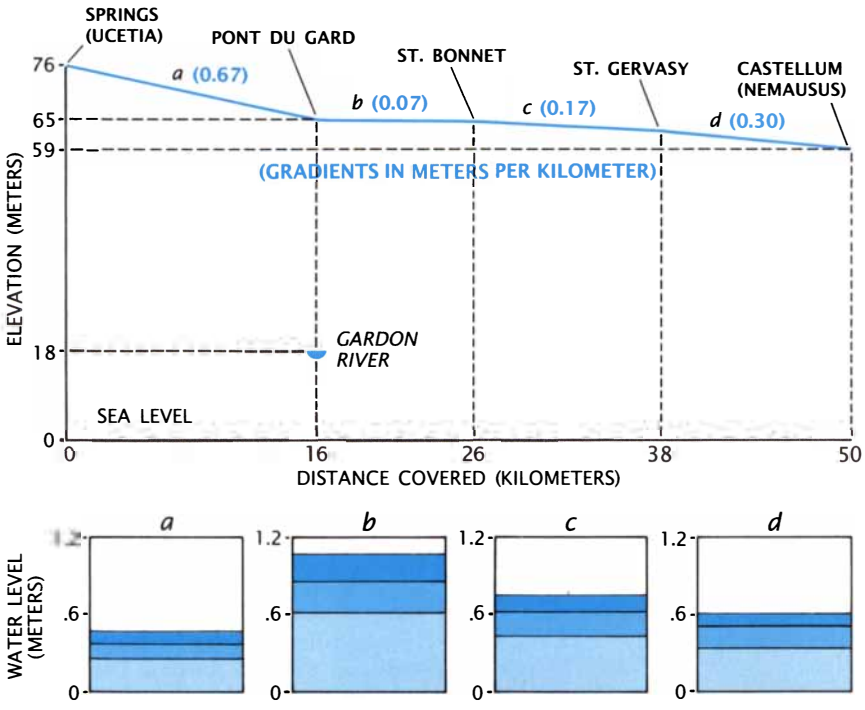
ORDER WITH CONFIDENCE

- *Most orders ship within 48 hours of receipt.*
- *Credit Card orders billed only upon shipment.*
- *No risk 30-day return privilege.*

UPS 2ND DAY AIR

THE MEDIA DEVELOPMENT GROUP

©1989 C12 ADVERTISEMENT



GRAPH plotting the elevation of the channel against the distance covered (*top*) shows the unusually slight slope of the channel. The average gradient was 0.34 meter per kilometer (the 17-meter difference in the elevation between the springs and Nemausus divided by 50 kilometers of channel), but the actual gradient varied (*numbers in parentheses*). Although the channel was relatively flat, its water depth—an indicator of the efficiency of water flow—was often optimal (0.6 meter), or approximately so, somewhere in the channel. The depths of the channel's major segments are indicated (*bottom*) for times when the flow rate from the springs was low (*light blue*), average (*medium blue*) or high (*dark blue*). Note in particular that the depth was optimal in the flattest segment (just below the Pont du Gard) during minimal flow and was optimal again in the final segment (between St. Gervasy and Nemausus) during maximal flow.

the bottom level. When storms cause flooding, so that water pounds against the bottom piers, the bases of these piers become about as vulnerable as those of the middle-level piers.

For both sets of piers I found that hurricane winds with a velocity of about 215 kilometers (134 miles) per hour at ground level are required for net tension to be produced at the bases of the middle and bottom piers. Yet most storms in the region have wind speeds no higher than about 150 kilometers (93 miles) per hour and exert wind pressures that are rarely more than about half those of 215-kilometer-per-hour winds. The Romans, then, built in a safety factor of about two against cracking, a factor that is both reasonable and equivalent to the typical margins of today.

The Romans knew how to calculate the weight of the masonry, but considering that they must have been uncertain about exactly how to calculate wind effects, it is quite impressive that they arrived at such sensible solutions for both tiers. Clearly the structural engineers, like their hydraulic counterparts, not only coped with formidable uncertainties in completing their assigned task but also did so without resorting to wasteful and excessive construction.

As is true of the Pont du Gard, much of the castellum has survived the centuries and holds clues to the logic of its designers. The building that covered the circular basin has vanished, but the basin itself is one of the best preserved installations of its kind. It is almost six meters in diameter and 1.4 meters deep. Grooves in the bottom and sides of the 1.2-meter-square port where the water entered the *maltha*-lined basin from the channel, together with small holes in the stone plates above the port, suggest the presence of adjustable gates of some kind at the entrance to the basin.

During normal operations the water flowed from the basin into the 10 distribution pipes, whose positions and 0.3-meter diameters are still indicated by holes penetrating the basin wall. The pipes were made of lead, which was known to be potentially toxic but could be tolerated because the rather hard water would soon leave a protective coating of calcium. The water could be drained by means of three holes in the floor, which were fitted with valves that were closed during normal operations; these drain holes too were approximately 0.3 meter in diameter. Presumably the water drained into a sewer.

Novak and I began our assessment of the castellum's design by exploring the hydraulic efficiency of the distribution pipes. We found that when the flow of water into the basin was at its maximum, the lead distribution pipes flowed half full (the optimum for a circular conduit), so that once again efficiency was maximized.

Next we attempted to discover answers to a few long-standing questions posed by archaeologists: What purpose did the gates at the entry port serve and how did they operate, and why did the engineers specify three large drains instead of a single small drain, as was typical?

As for the gates, we immediately discounted an old speculation that they served as a valve to control the flow of water into the basin. That theory does not hold water because the vaulting did not. Assuming the flow from the springs continued, a gate system that efficiently blocked the flow of water into the basin (if such a system were possible) would have forced water to back up and accumulate in the channel until it eventually leaked through the permeable vaulting in places where such leakage might be quite harmful. More likely, as others have suggested, the function of the gates was to measure the flow of water at a particular time. Excellent managers of water supply, the Romans would not have overlooked the need for metering.

Novak and I considered a number of possible metering arrangements and found one we believe to be the likeliest. It was a sluice (an arrangement in which water flows under a gate) that was both simple to operate and sensitive to variations in flow. Our findings suggest that the Roman engineers grasped a principle generally thought to have been discovered by 19th-century hydraulic engineers: If one knows the size of an underwater opening through which water passes from a channel into a basin, and if one also knows the head (the difference between the water level in the basin and the level in the channel) it is possible to determine the rate of flow. We think the Romans designed what can

BREAKTHROUGH:
A FAIL-SAFE SYSTEM FOR GETTING YOUR MESSAGE ACROSS.

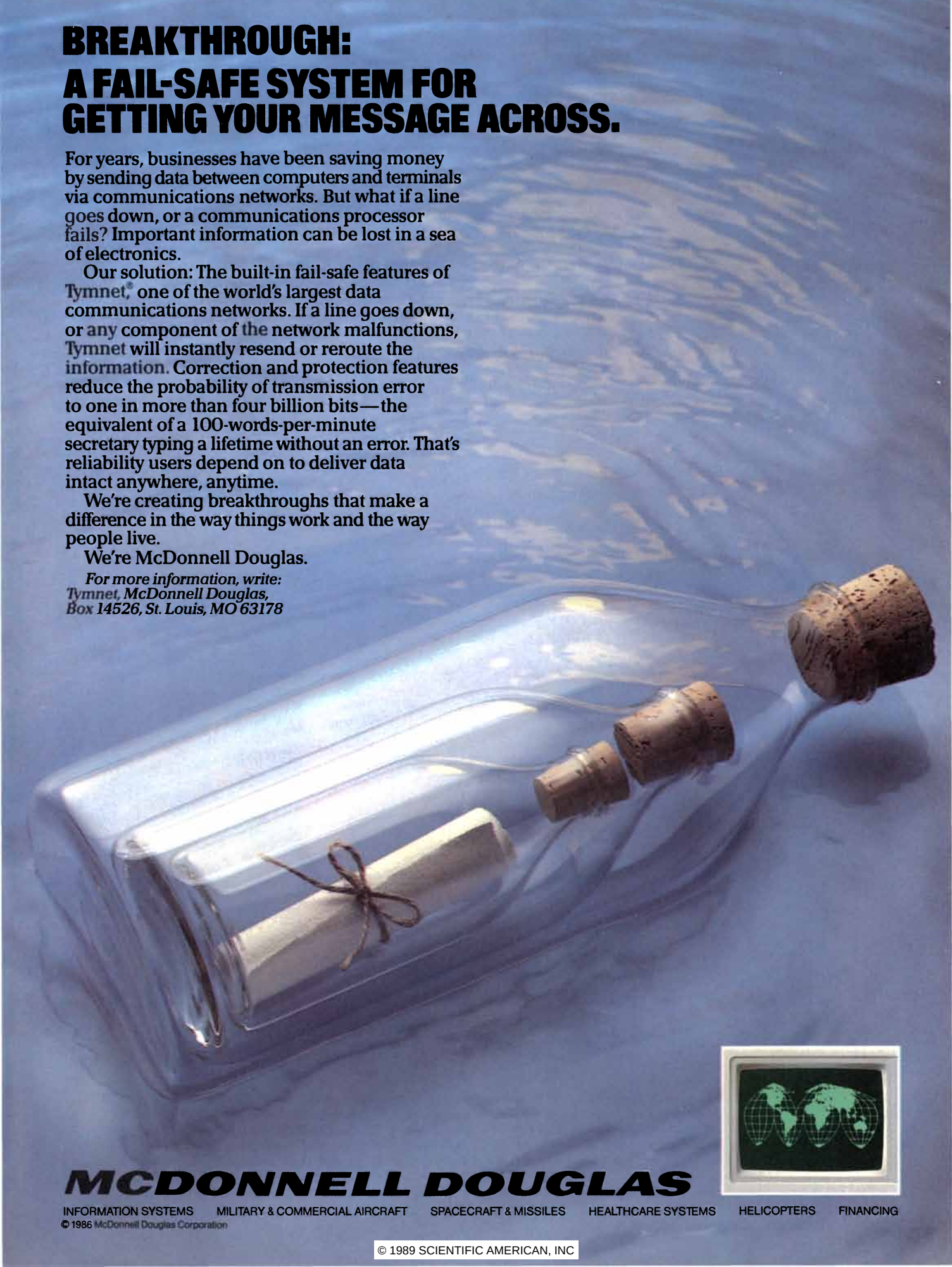
For years, businesses have been saving money by sending data between computers and terminals via communications networks. But what if a line goes down, or a communications processor fails? Important information can be lost in a sea of electronics.

Our solution: The built-in fail-safe features of Tymnet,® one of the world's largest data communications networks. If a line goes down, or any component of the network malfunctions, Tymnet will instantly resend or reroute the information. Correction and protection features reduce the probability of transmission error to one in more than four billion bits—the equivalent of a 100-words-per-minute secretary typing a lifetime without an error. That's reliability users depend on to deliver data intact anywhere, anytime.

We're creating breakthroughs that make a difference in the way things work and the way people live.

We're McDonnell Douglas.

For more information, write:
Tymnet, McDonnell Douglas,
Box 14526, St. Louis, MO 63178



MCDONNELL DOUGLAS

INFORMATION SYSTEMS MILITARY & COMMERCIAL AIRCRAFT SPACECRAFT & MISSILES HEALTHCARE SYSTEMS HELICOPTERS FINANCING

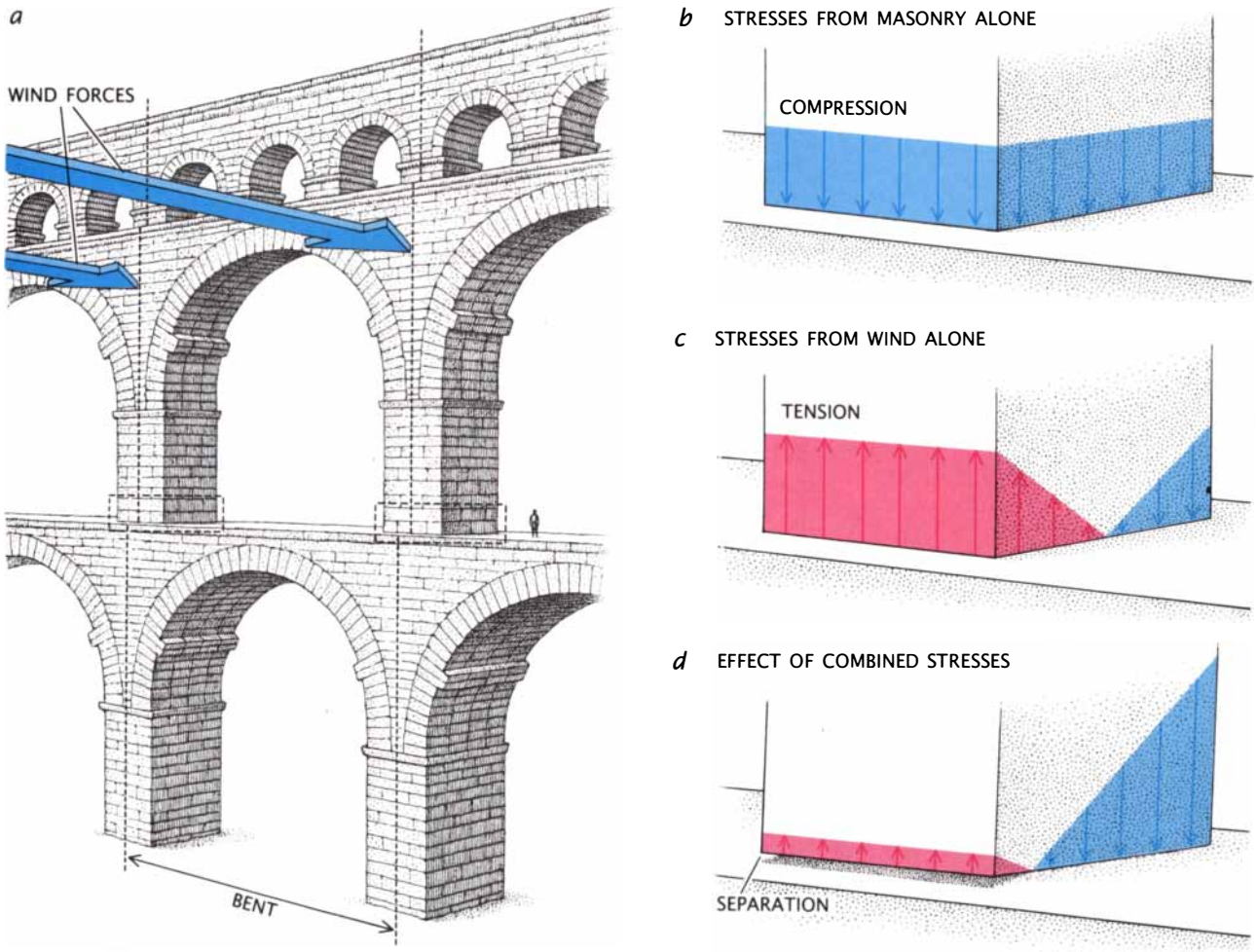
be described as a constant-head gate system [*see illustration on page 104*]. That is, they forced the head to be a fixed size, measured the size of the (variable) orifice needed to produce the chosen head, and from the orifice size derived the flow rate.

The sluice consisted of two gates that were as wide as the channel but less than half as high, positioned one directly behind the other. We imagine that at night no measuring was done and so both gates were raised; the water level in the basin was the same as that in the channel. In the morning the custodian, known as a *castellarius,* adjusted ropes passing through the openings in the stones at the top of the channel to lower the front gate (call it the head gate) until a mark on the gate was aligned with the surface of the water in the basin. That put the gate's top edge at a predetermined height (the height of the desired head) above the basin's water level. The back gate (call it the sluice gate) would then be lowered, thereby reducing the size of the underwater orifice and causing water to rise in the channel behind the gates.

The worker would continue to lower the sluice gate slowly until just a trickle of water passed over the head gate into the basin, indicating that the water in the channel was as high as the top of the head gate and hence that the head was at the desired height. At that point the *castellarius* could read the orifice height from a measuring stick connected to the sluice gate and multiply by a fixed factor to determine the flow rate; it is even possible that the gauge was calibrated to indicate the flow rate directly.

We believe the constant-head approach is the likeliest of several possible solutions because it fits the enigmatic archaeological evidence, gives highly sensitive readings, would have been easy to operate and is the only reasonable option in which the readings are linearly related to flow rates. The Roman designers would have liked the convenience of such a simple and reliable relation. Moreover, the gates we envision would have served a nonmetering purpose as well, one involving the operation of the once mysterious drains.

As for those drains, the explanation for their number and size is probably less complicated than has been



EFFECT OF WIND FORCES on a representative bent, or section, of the Pont du Gard (*a*) was calculated in order to determine which parts of the bridge are most vulnerable to separation of the masonry joints. The results showed that the bases of the piers (pillars) of the second tier (*boxed*) are most vulnerable, and that wind speeds greater than about 215 kilometers (134 miles) per hour are necessary to separate the bases from the ledges below. The critical wind speed was identified (*right*) by estimating the compressive, or downward (*blue*), stresses produced by the masonry (*b*), estimating the tensile, or upward (*red*), and compressive stresses produced by winds of various speeds (*c*) and then calculating the net stresses in the vulnerable region. Separation occurs whenever the tensile stresses exceed the compressive forces of the masonry (*d*).

Climb Aboard The Most Awesome Combat Aircraft in the World. Be there...

IN THE COCKPIT

The Video Library of Today's Airpower

F/A-18 HORNET

BONUS! Includes Actual Footage of Jan. 4, 1989 Shootdown of Libyan MIG-23's.

Spectacular Air Combat Action!

Attack Aircraft in Action!
Carrier Launched Strikes!
Soviet Aircraft Up Close!
High Flying Spy Planes!
...And Much More!

Stand an F-15 Eagle on its tail, and feel its twin 23,000 lb. thrust engines and afterburners rocket you to 60,000 ft.! Catapult off a flat top in an F-14 Tomcat, roaring from 0 to 170 mph in just 3 seconds!

Now you can experience the thrill of flying the most formidable combat aircraft in the world – planes of electrifying capabilities, astonishing supersonic speeds, and dazzling arrays of weaponry. It's all in a new, all-action video library, *IN THE COCKPIT*. Featuring spectacular combat footage and narration written by leading experts on military aviation.

Six Action-Packed Videos

You'll receive six exciting videos, filled with authentic, breathtaking action. Go on a mission aboard the finest reconnaissance aircraft built, the incredible SR-71 Blackbird in **Spies in the Skies.** Fly with dedicated spy planes like the ones used over Cuba and Vietnam. In **Red Star** see close-ups of the new Soviet Mig-29. See the Famous Hind and Halo attack helicopters in remarkable footage from Nicaragua and Afghanistan. Fly carrier launched air strikes against North Vietnam in **Flat Tops,** and search for Soviet submarines in the Hawkeye.

In **Eagles in the Sky,** fire Phoenix missiles from the Navy's F-14 Tomcat, and fly the Air Force's F-15 Eagle and F-16 Falcon on air superiority missions. In **Fly Low, Hit Hard** see ground attack action in Vietnam with the nimble A-4 Skyhawk and the most feared ground attack aircraft in the world like the Thunderbolt II and the battle-proven Israeli Kfir. And in **Target Tank,** witness combat helicopter action on the modern battlefield.

Attractively Priced, Convenient Monthly Installments

IN THE COCKPIT is available exclusively from Easton Press Video. Each set of six videos-60 minutes each- comes in a handsome library case and is priced at just $149.75† payable in convenient monthly installments. To order, simply return the attached coupon or call the Toll-Free number. If you are not totally satisfied, return your set within 15 days for a full refund. Order today to avoid disappointment!

† plus 4.75 shipping/handling per set.

CALL TOLL FREE 1-800-367-4534
In Conn. call: 1-855-8717

Easton Press Video
47 Richards Avenue
Norwalk, Conn. 06857

Satisfaction Guaranteed

For Fastest Service Call Toll-Free: 1-800-367-4534
In Conn. call: 1-855-8717

Yes! Please send me the *IN THE COCKPIT* video library. I understand only VHS format is available.

Name _____
PLEASE PRINT CLEARLY

Address _____

City/State/Zip _____

Signature _____
(All orders subject to acceptance)

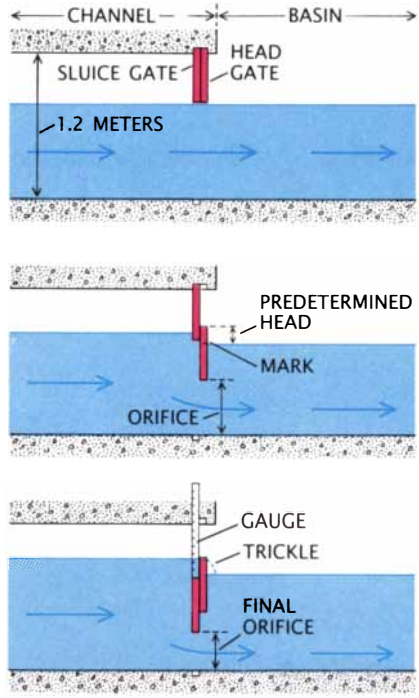
Charge each of 6 equal monthly installments of $25.75* to my credit card:
☐VISA ☐MasterCard ☐American Express ☐Diners Club

Credit Card No. _____ Exp. Date _____

☐ I prefer not to use a credit card and will pay by check. Enclosed is my deposit of $50.00*. I will pay the balance of $104.50* in three monthly installments of $34.83* each.

Allow 3 to 6 weeks for shipment.
*CT residents add 7½% sales tax.

117



CASTELLUM (*left*) probably had two adjustable gates where the channel opened into the basin (*square opening under blocked archway*). The gates, which apparently hung one behind the other (*right*), may have measured the flow rate in the channel, a value derivable from the head—the difference between the water level in the basin and in the channel—and the size of the underwater orifice through which the water entered the basin. A worker presumably moved the gates from their elevated full-flow condition (*top*) and adjusted them so that the head was at a preset height: he lowered the front, or head, gate until a mark some fixed distance from the top of the gate (equal to the desired head) was level with the water in the basin (*middle*); then he gradually lowered the back, or sluice, gate until just a trickle of water flowed from the channel over the head gate into the basin, indicating that the head was now at the desired height (*bottom*). The flow rate could then be determined by reading it from a gauge attached to the sluice gate, or reading a measure of the final orifice and multiplying by some factor.

thought. In order for someone to clean the basin or make repairs, the water in the basin would have had to be emptied. In the past archaeologists assumed that the inflow into the basin could be stopped indefinitely, giving workers unlimited time to wait for the basin to drain. In that case even a tiny drain hole would do. Yet if there was no practical way to block the flow of water into the basin for very long—which was almost certainly the case—a single, small drain would have been inadequate. Under such circumstances the presence of several large drains would have cleverly ensured that the basin could either be drained almost to the bottom while the inflow continued or be drained empty while the inflow was stopped temporarily.

When the flow rate was naturally low, simply opening all the drains would reduce the water level to ankle depth. If workers wanted to drain the basin completely, all they had to do was raise the gates at the same time as they opened the large drains, wait for the water level in the channel to fall and then block off the gateway completely. This operation would evacuate the water in the basin in half a minute after the gates were closed; at low flow the water could be retained behind the gates long enough to ensure that the basin floor remained dry for about 20 minutes—enough time for sweeping the floor or making minor repairs.

Individually each of the findings relating to the design of the aqueduct is surprising. Together they are startling, and they suggest that the Roman engineers were either very lucky or, more likely, had greater wisdom than they have received credit for. How long did their aqueduct fulfill its intended function? It worked properly for about four centuries. Once the Roman civilization faltered and maintenance of the aqueduct declined, the channel became severely encrusted by calcium deposits; by the eighth century the conduit was so clogged as to be virtually useless. As time went by wars and earthquakes destroyed many parts of the aqueduct, and people quarried the stone and the lead for other purposes.

The Pont du Gard too suffered greatly, particularly during the Middle Ages, when someone had one of the more outrageous ideas in technological history: the second-tier pillars were dangerously narrowed to make a passage for travelers along the top of the bottom tier. The ill-begotten roadway remained in use until the adjoining road bridge that now serves two million visitors annually was constructed in the 1740's by Henri Pitot. The original bridge was not adequately restored, however, until 1855, when the emperor Napoleon III finally ordered thorough repairs. Today the Pont du Gard remains an eloquent witness to Roman creativity and daring.

FURTHER READING

STRUCTURAL DESIGN OF THE PONT DU GARD. George F. W. Hauck in *Journal of Structural Engineering,* Vol. 112, No. 1, pages 105–120; January, 1986.

INTERACTION OF FLOW AND INCRUSTATION IN THE ROMAN AQUEDUCT OF NÎMES. George F. W. Hauck and Richard A. Novak in *Journal of Hydraulic Engineering,* Vol. 113, No. 2, pages 141–157; February, 1987.

THE AQUEDUCT OF NEMAUSUS. George Hauck. McFarland & Company, Inc., 1988.

WATER FLOW IN THE CASTELLUM AT NÎMES. George F. W. Hauck and Richard A. Novak in *American Journal of Archaeology,* Vol. 92, pages 393–407; July, 1988.

The History Of The World Comes Alive When You See The Big Picture.

Never has there been a clearer, more concise overview of our history, archaeology, theology and mythology than with The World History Chart from International Timeline, Inc.

When you look at 6,000 years of history in chronological order—on a chart almost three feet high and eight feet long— important dates and events just seem to fall into place!

Test your talent for historical trivia. What do the Great Wall of China and the Chin Dynasty have in common with the Roman Empire? When the Normans migrated to Northern France, where did they come from?

The world takes on new dimensions as you witness the development of 15 major civilizations from 4,000 B.C. to the present. To fully understand the scope of information on the Chart—picture U.S. history as just two inches over the space of eight feet of time!

This invaluable teaching and reference tool has taken years to compile. Its easy-to-read format appeals to everyone from the casual admirer to the scholar.

Now, you can have your own for just $29.95 (U.S. currency) plus $5.00 for shipping and handling ($34.95 each). Add $15.00 for each laminated chart ($49.95). **Simply call toll-free 1-800-621-5559 (1-800-972-5855, in Illinois),** and charge it to your MasterCard, VISA or Choice. Or, complete the coupon below and mail with check or money order payable to: **International Timeline, Inc., 2565 Chain Bridge Road, Vienna, VA 22180.** (Virginia residents add 4½% sales tax. Outside U.S. add $10.00 instead of $5.00 for shipping and handling. Allow 4-6 weeks for delivery.)

Please send me the World History Chart for $29.95 (U.S. currency) plus $5.00 for shipping and handling for each chart ordered ($34.95). Add $15.00 for each laminated chart ($49.95). Outside U.S. include an additional $5.00 for shipping and handling.

6,000 Years Of History At A Glance For Only $29.95!

Send to:_____ Address:_____
(please print)

City:_____ State:_____ Zip:_____

Of Charts Desired:_____ Laminated_____ Unlaminated

Enclosed is____ check ____ money order for $_____.

Virginia residents add 4½% sales tax. Please allow 4-6 weeks for delivery.

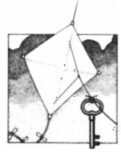
Checks payable to: **International Timeline, Inc.**

SA-3

2565 Chain Bridge Road, Vienna, VA 22180

THE AMATEUR SCIENTIST

How to get the playground swing going: a first lesson in the mechanics of rotation



by Jearl Walker

At the right age a child can climb onto a playground swing and learn through mimicry and experiment how to start it moving and how to "pump" it. Yet explaining the scientific principles behind swinging is hardly child's play. How does one initiate the pendulum motion other than by simply shoving off from the ground or having someone push? And then how does one build the amplitude (the angular extent of the swinging) when either standing or sitting in the swing?

Although the pastime of swinging is ancient, the relevant mechanics got little attention until 1968, when Peter L. Tea, Jr., and Harold Falk of the City College of the City University of New York published a study. They limited their examination to the case of someone standing in a swing that is already in motion. As you may have learned as a child, you can increase the amplitude by standing and squatting in synchrony with the swing's motion, standing when you pass through the lowest point and squatting at the two highest points. If the swing has rigid supports instead of ropes or chains, and if you are persistent, you may be able to swing higher than the swing's support bar. In fact, if the attachments of the supports to the bar allow it, you could even end up moving in full circles; some circus performers do it.

Why does the procedure of standing and squatting build the swing's amplitude? Tea and Falk pointed out that several factors are involved, but what is paramount is the matter of energy and just how you work to increase it. Picture yourself on a rope-supported swing; disregard any loss of energy to friction and air drag [*see top illustration on opposite page*]. Suppose you are squatting when you reach the highest point of a swing to the rear. Your energy just then is entirely potential energy, which depends on the height of your center of mass from some reference level. Let the lowest point of the arc through which your center of mass swings be the reference level; assume that your center of mass is two meters above it.

As you descend, the energy is gradually transformed into kinetic energy and you gain speed. When you reach the lowest point, your energy is entirely kinetic energy and you are moving at peak speed. As you begin to ascend on the arc, the transformation is reversed: you slow down and then stop momentarily at the top of the arc.

How high you go depends on what you have done during the swing. If you have continued to squat, the upward motion is a mirror image of the downward motion, and your center of mass ends up just as high as when you began the forward swing, that is, at a height of two meters. If instead you stand when you are at the lowest point, you swing higher.

There are two reasons for the greater height. One of them is that after you stand your center of mass is higher as you begin the upward swing. Suppose standing raises your center of mass by .5 meter. Assume that standing does not change your kinetic energy (an assumption I shall further examine below). The amount of kinetic energy you have at the lowest point determines how high your center of mass goes during the rest of the forward swing. Since you have descended by two meters, you must have enough kinetic energy for your center of mass to swing back up by two meters. With the extra .5 meter added by the standing, your center of mass reaches a height of 2.5 meters at the top of the path. As you approach the top you rotate the swing platform around your center of mass, so that it too goes higher than before.

The amplitude is larger, then, but if you continue to stand, you will just swing back and forth to heights of 2.5 meters. If you want to pump the swing to greater heights, you must squat each time you reach the end of an arc so that you can stand again when you pass through the lowest point. Of course, by squatting you lower your center of mass somewhat and sacrifice some of the height at the end of the arc. Yet you do not lose as much height as you gained by standing. The reason is that when you squat, the swing is at an angle to the vertical; you shift your center of mass along a slanted line rather than straight down. As you swing higher and thereby increase the angle, the height you lose with each squat decreases. If you can manage to swing up to an amplitude of 90 degrees, squatting does not lower your center of mass at all.

The other reason for the increase in amplitude when you stand at the lowest point is subtler. By standing you move your center of mass toward the axis about which you rotate, and that actually increases your kinetic energy rather than leaving it unchanged, as I assumed above. The increase may seem strange: normally when you lift something—including yourself when you stand up—you do not increase the kinetic energy of what you lift. Instead your work in the lift merely adds to the potential energy, because you have increased the height of what you lift.

The lift during swinging is different, because you are moving in circular motion. To see the point, consider a different and more conventional demonstration. When an ice skater spins on point with outstretched arms and then pulls her arms inward, she spins faster. This common textbook example displays a powerful but often nonintuitive principle of physics: the conservation of angular momentum. That form of momentum is the product of the rate of rotation and a factor called the moment of inertia, which depends on the quantity of mass and its distribution with respect to the axis of rotation. For the ice skater the moment of inertia is initially large because of the outstretched arms; it is smaller when the arms are pulled closer to the axis of rotation.

The angular momentum of an object can change only when a force acts on it in a certain way. If the force is not radial (if a mental extension of the force does not pass through the axis of rotation), the force creates a torque that changes the angular momentum. If the force is radial, though, it does not create a torque and the angular

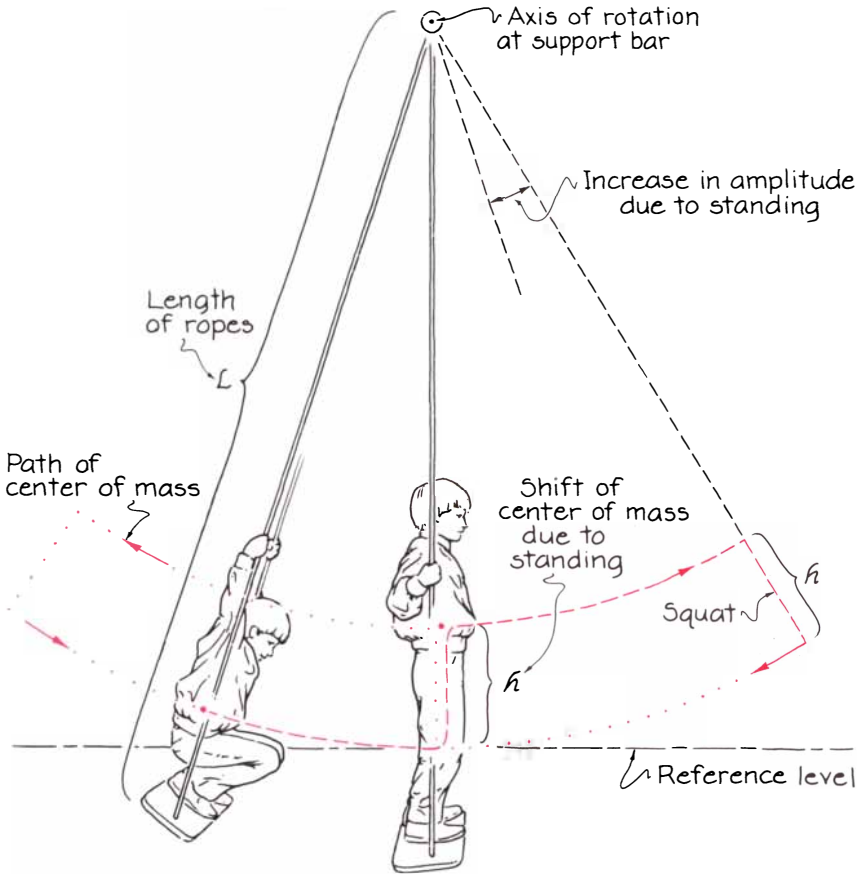
momentum does not change. When the ice skater pulls in her arms, the forces are radial and her angular momentum is unchanged by the action. The decrease in her moment of inertia is therefore matched by an increase in her rate of rotation.

She also increases her kinetic energy, which depends on her moment of inertia and the square of the rate of rotation. Although the moment of inertia is decreased, the change in the kinetic energy is dominated by the increased rate of rotation. The increase in energy comes from the work she does in pulling in her arms: she works, in effect, against centrifugal force. (Such a force is usually only a fictional device, because in most examples of rotation the notion that a force tugs outward on a rotating object is one of convenience rather than reality. Still, one can state mathematically that the ice skater does work against a centrifugal force, and the work increases her kinetic energy.)

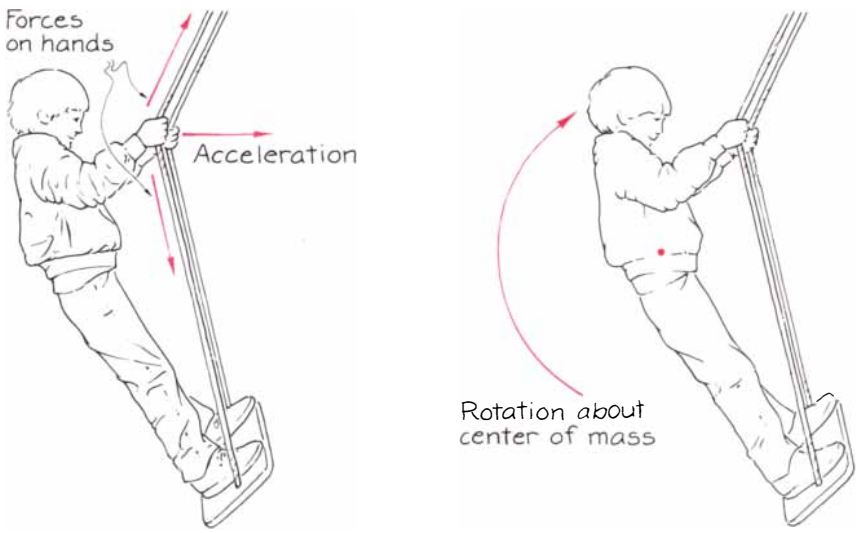
The same principles apply when you stand up in a swing as it passes through its lowest point. By standing you decrease your moment of inertia, because you shift your center of mass toward the axis about which you rotate—the support bar of the swing. The forces you exercise are upward and radial to the rotation, and therefore they do not produce torques. Since your angular momentum cannot change while you stand, the decrease in your moment of inertia must be matched by an increase in your speed around the bar. As a result your kinetic energy increases too. The added energy comes from the work you do against centrifugal force. No such considerations apply to a squat at the top of the arc, because there the swing has slowed to a stop and no centrifugal force acts on you.

To see how your work against centrifugal force changes the height to which you swing, assume that the length of the ropes is five meters. You squat at a height of two meters, swing down, stand up and swing to the top of the arc. Your center of mass is then at a height of 2.97 meters; the additional .47 meter comes from your work against centrifugal force. After you squat your center of mass is at a height of 2.52 meters, which is somewhat higher than the previous result.

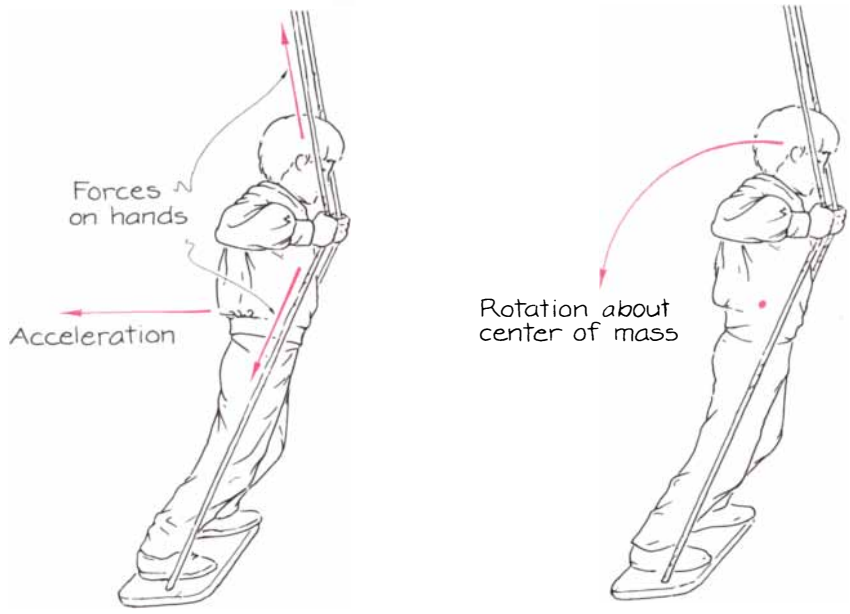
I do not mean to imply that your angular momentum remains constant throughout your swing, only that the act of standing does not change it. Except when you are at the lowest point in the arc, a mental extension of the force of gravity on you (your weight) does not pass through the axis of rotation, and the torque it creates continuously alters your angular momentum. When you swing upward, the torque counters your rotation, decreasing your angular momentum until you have none when you reach the top of the arc. Then, as you descend, the torque reestablishes your angular momentum until it is largest when you reach the lowest point. One way to describe the advantage of standing at the lowest point is in terms of the angular momentum. Since it is largest there, the decrease in your moment of inertia when you stand up results in a



Pumping a playground swing



The results of pulling hard on the ropes



The results of pushing hard on the ropes

larger increase in your speed than if you stood at some other point.

If you have been on a swing lately, you may have noticed that as you keep swinging the standing up becomes more difficult and the increase in amplitude with each pass through the arc becomes larger. At an early stage, standing is easy because the centrifugal force on you is small owing to your slow speed through the lowest point. When you raise your center of mass, you therefore do little work against the force and increase your kinetic energy only modestly. Also (because the angle of swing is small) when you squat, the decrease in the height of your center of mass is almost as much as the increase in its height when you stand. Since the work is small and the loss of height is large, there is only a slight increase in amplitude with each pass through the arc.

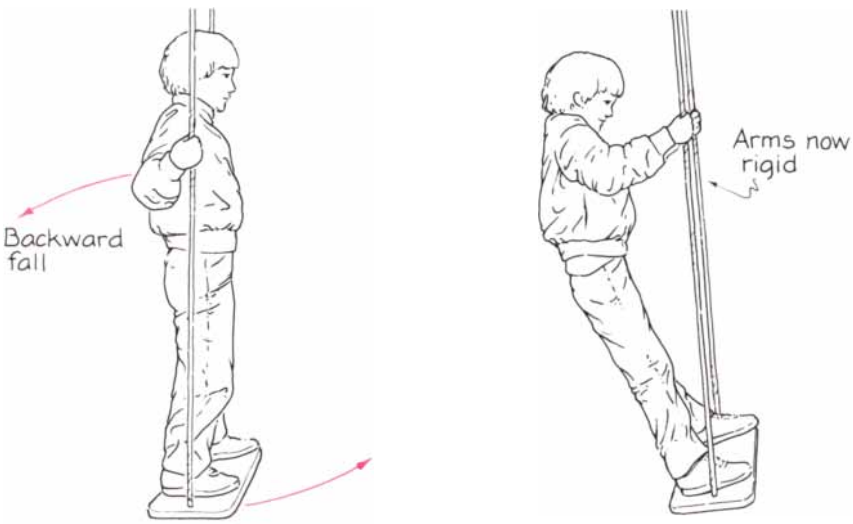
Later, when the swing is going better, the centrifugal force is larger because of your faster speed through the lowest point. You might then have to struggle to stand, but in doing so you work more and increase your kinetic energy more than before. And when you squat, you lower your center of mass less. As a result the increase in amplitude with each pass through the arc is larger.

Pumping by standing up and squatting makes a swing go higher, but how do you get the swinging started without pushing off from the ground or having someone help you? One way was explored in 1970 by Bryan F. Gore of Central Washington State College. Imagine that you are standing on a stationary swing suspended by lightweight ropes. If you suddenly lean back and pull hard on the ropes, you can move them off the vertical toward you [*see bottom illustration on preceding page*]. The forces exerted on your hands by the ropes are then partially in the forward direction, and so you begin to accelerate forward. Since the ropes are tied to a fixed bar and to the platform that bears your weight, your pull on them is effectively directed against the bar and your weight.

The forces also yield a torque that rotates you about your center of mass, with the top half of your body tending to rotate forward and the bottom half (along with the platform) tending to rotate backward. The rotation is superimposed on your general forward motion and that of the swing; unchecked, it would move the top half of your body in front of the platform. As you reach the highest point in this first, short arc, you might stop the rotation about your center of mass by resisting it with your arms until you become aligned with the ropes. Instead you can actually shove the ropes forward, so that their distortion creates forces on you that are partially toward the rear [*see top illustration on this page*]. The torque from the ropes then eliminates your initial rotation and may send you rotating about your center of mass in the opposite direction.

Gore showed that after you first get the swing moving you can pump it by timing your pull and push on the ropes. The procedure works whether you are standing or sitting. When you swing forward, pull the ropes toward you so that their distortion generates forces on you in the forward direction. At the top of the arc push on the ropes to stop your rotation about your center of mass and to launch yourself toward the rear. As you swing backward, continue to push on the ropes. Near the top rear of the arc pull on the ropes to stop the rotation about your center of mass and generate a force in the forward direction. The work you do against the ropes feeds energy into the swing.

In 1972 John T. McMullan of the New University of Ulster demonstrated that a swing can be started from rest even if it has rigid supports rather than the flexible ropes in Gore's model. The procedure is to stand on the



The initiation of swinging with rigid supports

swing with hands on the supports and arms bent, and then to fall backward until your arms are stretched out, stopping your fall. During the fall you and the swing function as a double pendulum: you rotate about the platform while the platform begins to rotate about the support bar. Once you stop your fall and your arms are rigid, you and the swing function together as a single pendulum. The energy of the motion comes from your initial fall, as potential energy is converted into kinetic energy. After you start the swing you can pump it by standing and squatting.

In 1976 the subject of stand-and-squat pumping was reexamined by Stephen M. Curry of the University of Texas at Dallas. He discovered that the rate at which energy is fed into the swinging has several surprising features. Suppose you are about to swing through an arc after having squatted. Suppose also that E_o is your energy just then, h is the height through which you shift your center of mass when you stand and L is the length of the ropes. When you again squat at the opposite end of the arc, you have an energy of about $E_o(1 + 3h/L)$. After a complete oscillation (forward and backward) you have an energy of about $E_o(1 + 3h/L)^2$, and after n complete oscillations you have an energy of about $E_o(1 + 3h/L)^{2n}$. Provided h is much smaller than L, the energy then can be restated as approximately $E_o\exp(6nh/L)$, which indicates that the energy grows exponentially with the number of oscillations.

If the amplitude of the swinging happens to be small, the time T required for a complete oscillation is a constant that depends only on the acceleration of gravity and the length of the pendulum that you and the swing constitute. Suppose you somehow begin the swinging with an energy of E_o and pump for a time t. The number of oscillations you have undergone is t/T, and your energy is then $E_o\exp[6(t/T)(h/L)]$, which shows that the energy also grows exponentially with time.

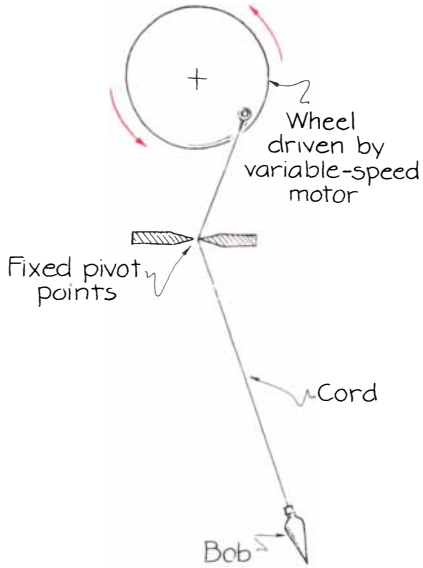
Curry noted two curious results of his derivation. First, the rate at which energy is pumped into the swing does not depend on your mass or weight. Although a tall person has an advantage over a short one because h has a larger value, there is no advantage in weight. The second result is curiouser. The calculation indicates that if the initial energy is zero, your energy after many pumps in which you stand and squat must still be zero, which certainly seems reasonable. But can the initial energy ever be truly zero? No, it cannot, because your swinging takes place in an environment of air molecules in thermal motion, which beat against you incessantly and provide some initial energy. From that supply you can build your swinging, but it takes standing and squatting for about four minutes before you see the results.

Curry carried the argument even further. Suppose you could somehow survive in an environment at a temperature of absolute zero, so that the air molecules could not help you to begin the swinging. Would that eliminate the possibility of pumping up a swing merely by standing and squatting? No, it would not, because one of the tenets of quantum mechanics disallows zero energy. Even at a temperature of absolute zero, you and the swing must still have a certain (albeit minuscule) energy, and from it you can amplify your swinging. In this case it would take about six minutes before you would see any result. The thought of pumping a swing at a temperature of absolute zero may be bizarre, but the pumping is allowed in theory.

In 1984 Juan R. Sanmartín of the Polytechnic University of Madrid observed that the mechanics of pumping a swing by standing up and squatting can be related to the dramatic pendulum motion of a censer that has been swung during rituals in the Cathedral of Santiago de Compostela in northwestern Spain for the past 700 years. The censer, which weighs as much as a slender man, hangs by a stout rope from two wood rollers about 20 meters above it. The rope wraps around the rollers and runs down to a crew of men who control it at floor level.

After the censer is pushed to begin swinging, the men pump the pendulum by timing their pull on their end of the rope. When the censer passes through the lowest point of its arc, they pull hard to shorten the length of the pendulum by almost three meters. When the censer reaches its highest point of travel, the men relax their pull and allow the length of the pendulum to increase to its original value.

The pumping they accomplish is similar to the pumping contributed by standing and squatting on a playground swing. The men add energy to the swinging when they shorten the length of the pendulum and pull against the centrifugal force on the censer. After 17 pulls and about 80 seconds, the censer swings up by almost 90 degrees, rising to within a meter of the cathedral's vaulting. Its rapid passage through the lower part of the arc greatly fans the burning coals and incense it carries. It is quite a sight to see.



A. E. Siegman's parametric amplifier

When a pendulum such as a swing is pumped twice during each complete oscillation, the pumping is said to be parametric. In 1969 A. E. Siegman of Stanford University described how parametric pumping of a swing can be demonstrated with a simple model. A pendulum bob is suspended by a long cord from a point near the rim of a small wheel that can be rotated in the vertical plane by a motor. Just below the wheel the cord passes between a pair of edges that serve as pivots.

The distance from the bob to a pivot point determines the frequency at which the pendulum would swing naturally if you gave the bob a small push. If you adjust the motor so that the frequency at which the wheel turns is twice the natural frequency of the pendulum, the pendulum will start up from rest without any push from you. Although at first the bob merely wiggles, its energy is amplified parametrically whenever the wheel pulls upward on the cord—that is, twice during each complete oscillation of the pendulum.

FURTHER READING

PUMPING ON A SWING. Peter L. Tea, Jr., and Harold Falk in *American Journal of Physics,* Vol. 36, No. 12, pages 1165–1166; December, 1968.

HOW CHILDREN SWING. Stephen M. Curry in *American Journal of Physics,* Vol. 44, No. 10, pages 924–926; October, 1976.

O BOTAFUMEIRO: PARAMETRIC PUMPING IN THE MIDDLE AGES. Juan R. Sanmartín in *American Journal of Physics,* Vol. 52, No. 10, pages 937–945; October, 1984.

COMPUTER RECREATIONS

Of worms, viruses and Core War

by A. K. Dewdney

"The only truly secure system is one that is powered off, cast in a block of concrete and sealed in a lead-lined room with armed guards—and even then I have my doubts."

—EUGENE H. SPAFFORD

The knock on the door had a palpable urgency that brought the computer-center director's head up sharply from the pile of papers before him. He grunted loudly and the computer operator entered.

"Something's gone wrong. We have some very weird processes going on. We're running out of memory. I think we've got a virus in the system."

If the center had been equipped with claxons, the director would undoubtedly have set them off.

Such a scene has been played out in one form or another all too often in recent years, and as a result computer viruses have been increasingly in the news. In fact, this department has been cited more than once in connection with the rash of virus outbreaks, probably because it is an instigator of Core War, a game in which computer programs are purposely designed to destroy one another. But, as we shall see, Core War has no direct connection with the infections.

To understand how a computer virus works one must first understand in great detail the system in which it operates. The same thing applies for understanding the operation of worms, logic bombs and other threats to computer security. This simple observation has two immediate implications. First, journalists are likely to misreport or distort virus news stories for quite innocent reasons: most reporters are more or less mystified by the internal workings of computers. Second, public descriptions of a computer virus—even fairly detailed ones—cannot be exploited to reconstruct the virus except by someone who has the requisite knowledge of the affected computer system to begin with. A knowledgeable "technopath" who is bent on destroying other people's programs and data hardly needs to read a magazine or newspaper article to begin imagining ways to construct a virus. There is consequently no harm in describing how viruses and other destructive programs work. (Indeed, such a description is probably constructive in that it might stimulate efforts to protect computer systems.)

One must distinguish from the start between the two commonest types of malignant program. A virus rides piggyback on other programs; it cannot exist by itself. A worm, on the other hand, can pursue an independent existence, more in the manner of a bacterium. Both kinds of "infection," like all programs, depend on an operating system.

Most readers know that a running computer consists of both hardware and software. In front of me at the moment, for example, is a piece of hardware: an Apple IIc. Inside the machine's memory is software: a program called the Appleworks Word Processor. The program transfers the characters I type on the keyboard into a section of memory reserved by the program for text.

But the word-processing program is not able to run by itself. The program depends on an operating system that, among other things, translates it into a special machine language that enables the hardware to carry out the program's instructions. The operating system for a personal computer normally resides on a disk. To do anything on such a machine (from writing to playing games), the disk operating system (DOS) must first be loaded into the computer's hardware memory. In a home computer the DOS is usually loaded automatically from a disk, which may or may not contain the program one wants to run, as soon as one switches on the computer.

To run a particular program on my personal computer, I must type the name of the program into the computer. The computer's DOS then searches through the disk for a program with that name, loads it into memory and runs it—instruction by instruction—as is shown in the middle illustration on the opposite page.

In loading the program the DOS sets aside part of the hardware memory not only for the program but also for the program's "work space." Here the program will store all the values assigned to its variables and its arrays, for example. In doing all of this the DOS is normally careful not to overwrite other programs or data areas, including whatever part of the DOS happens to be in memory. The DOS is equally careful in storing programs or data onto a disk.

Often a programmer may find it necessary to employ the commands the DOS itself uses, which can generally be found in the appropriate manual. Such commands make it possible to write a subprogram that can read files from disk into memory, alter the files and then write them back onto the disk—sometimes with malicious intent.

Here is a sample virus subprogram that does just that. It contains a mixture of pseudo-DOS commands and subroutines: small, internal programs (whose component instructions are kept separate from the subprogram's main body) that carry out specific missions whenever they are called.

```
this := findfile
LOAD (this)
loc := search (this)
insert (loc)
STORE (this)
```

The subroutine designated findfile opens the directory of executable files, or programs, on a disk, picks a random file name and assigns the name of that file to a variable called *this*. The next line of the program makes use of the pseudo-DOS command LOAD to copy the file into the computer's memory. Another subroutine called search then scans the program just loaded, looking for an instruction in it that can serve as a suitable insertion site for a virus. When it finds such an instruction, it determines the instruction's line number and assigns its value to the variable called *loc*.

At this point the virus subprogram is ready to infect the program it has

randomly picked. The subroutine insert replaces the selected instruction with another instruction (such as a call for a subroutine) that transfers execution to a block of code containing the basic virus subprogram, which is appended to the end of the program. It then adds the program's original instruction to the end of the appended subprogram followed by a command that transfers execution to the instruction following the insert in the host program.

In this way when the virus subprogram is executed, it also executes its host program's missing instruction. The execution of the original program can then proceed as though nothing unusual had occurred. But in fact the virus subprogram has momentarily usurped control of the DOS to replicate itself into another program on the disk. The process is illustrated graphically at the right in the illustration below. When the newly infected program is subsequently loaded by the DOS into the computer's memory and run, it will in turn infect another program on the disk while appearing to run normally.

As early as 1984 Fred S. Cohen carried out controlled infection experiments at the University of Southern California that revealed—to his surprise—that viruses similar to the one I have just described could infect an entire network of computers in a matter of minutes. In order to explain the kinds of damage such viruses can do, I shall adapt Cohen's generic virus, writing it in a pseudolanguage.

```
1234567
main program:
1. infect
2. if trigger pulled, then do damage
3. go to host program
subroutine: infect
1. get random executable file
2. if first line of file = 1234567, then
   go to 1, else prepend virus to file
subroutine: trigger pulled
subroutine: do damage
```

Cohen's generic virus is generic in all but its attachment site: instead of inserting itself in the middle or at the end of the host program, it attaches itself to the beginning. The first line of the virus program is the "recognition code" 1234567. The main program first calls up the subroutine infect, which randomly retrieves an executable file from a disk and checks whether the first line of that file happens to be 1234567. If it is, the program has already been infected and the subroutine picks another program. If the subroutine happens to find an uninfected file, it "prepends" the entire virus program to the target program. This means simply that it places itself at the head of the program and arranges to get itself executed first before transferring control back to the infected program.

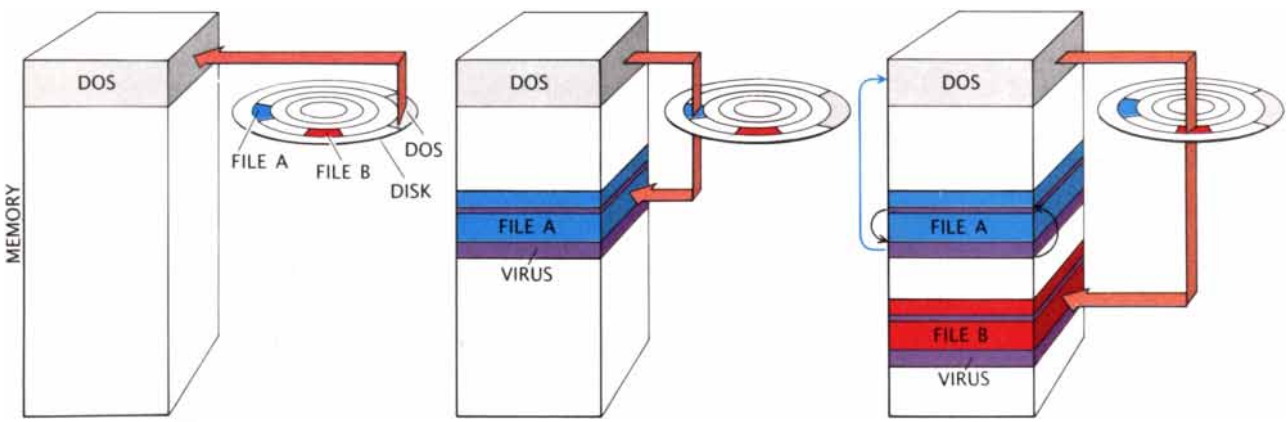
The next two subroutines call for a triggering condition and for some damage to be done. The triggering condition might be a certain date reached by the system clock, or perhaps the deletion of a certain employee's name from the payroll file. The damage to be done might be the erasure of all files or, more subtly, the random alteration of bits in just a few places. The possibilities are endless.

Triggering conditions and damage bring us to the edge of moral territory. I think there is no question that willfully perpetrating damage to computer files by means of a virus or other destructive program is a reprehensible act. Computer programs and data are not just impersonal strings of 0's and 1's but also miniature cities of thought. Why would anyone want to destroy the work of another?

Writers of virus programs do so for a variety of reasons. For example, a programmer in a large company might secretly harbor a grudge against the company's management. He might implant a virus for the day his employment is terminated. When that happens, his records will be deleted from the payroll file, triggering the virus. As a result, valuable company data and programs might either disappear or develop serious and costly errors. Replacing faulty data and programs with backups stored on other media might be of no avail, since if the backups were made recently, they too might be infected.

Of course, the kind of destruction just described would ordinarily take place in a multiple-user computer system. The operating system in this kind of computer environment is considerably more complex than a disk operating system for a personal computer. For one thing, the fact that so many users share the same facilities requires an operating system that protects users as much as possible from inadvertent or deliberate interference. Even here, however, viruses are possible, but they require much more sophistication. Usually they exploit a flaw in some part of the operating system—a bug, so to speak—as was evidenced by the "virus" (actually it was a worm) that spread throughout the Internet last fall.

During the evening of November 2, 1988, someone ran a worm program on one of the several thousand North American computers interconnected through a data-communications network called the Internet. The network connects machines at universities, businesses, Government agencies such as the National Aeronautics and



Loading the DOS *into memory* (left) *allows a program to be read* (middle) *and a virus* (right) *to replicate itself*

INSTRUCTION	EXPLANATION
DAT B	A nonexecutable data statement; B is the data value.
MOV A B	Move contents of address A to address B.
ADD A B	Add contents of address A to address B.
SUB A B	Subtract contents of address A from address B.
JMP B	Transfer control to address B.
JMZ A B	Transfer control to address A if contents of address B are zero.
JMN A B	Transfer control to address A if contents of address B are not zero.
DJN A B	Subtract 1 from contents of address B and transfer control to address A if contents of address B are not zero.
CMP A B	Compare contents of addresses A and B; if they are equal, skip the next instruction.
SPL B	Split execution between next instruction and the instruction at address B.

A summary of Core War instructions

Space Administration and even some military installations. With frightening speed the worm spread to more than 1,000 machines during that evening and the next day. As copies of the worm proliferated, operators of individual systems noticed memory utilization soaring and machine response becoming sluggish. The worm did not attack files or other programs, however. It seemed content merely to proliferate throughout as many host machines as possible. Even so, the damage in lost time was immense.

As I mentioned above, a worm is a program that pursues an independent existence within a computer; it moves from computer to computer on its own, leaving duplicates of itself in each machine. The Internet worm consisted of two parts, a vector and a body. Starting from a given host computer, the worm would seek a new host by sending its vector to another computer. Once inside the machine, the vector would establish a communication link through which the worm's body could be sent. Details of this attack were revealed by Eugene H. Spafford of Purdue University in a 40-page document a few weeks after the event. One example of the worm's operation shows the cleverness of its creator.

UNIX, the operating system of choice on many of the Internet computers, allows processes to take place in the computer that are not associated with any particular user. Such independent processes are called demons. One demon, called fingerd (pronounced fingerdee), enables users to get information about other users. Such a service is desirable in a computing environment in which users must share programs and data for research and development purposes.

The worm in its current host computer would send a message to one of the other potential host computers on its list (which was obtained illegally). In requesting the services of the fingerd demon, the worm gave it some information, just as an ordinary user might. But the worm supplied so much information to the demon that the data filled the space reserved for it in the computer's memory and overflowed into a "forbidden" area.

The area that was thus overwritten was normally reserved for instructions that fingerd consulted in deciding what to do next. Once inside such an area the worm (whose body still inhabited the original host machine) invoked a so-called command interpreter of the new machine, effectively claiming a small piece of the UNIX operating system all to itself. After the command interpreter was at its disposal, the worm transmitted some 99 lines of source code constituting the vector. On the worm's command, the unwitting potential host then compiled and ran the vector's program, virtually guaranteeing infection.

The vector program hid itself in the system by changing its name and deleting all files created during its entry into the system. After doing that it established a new communication channel to the previous host and, using standard system protocols, copied over the files making up the main body of the worm.

Once inside a new host computer, the worm's main job was to discover the names and addresses of new host machines by breaking into areas reserved for legitimate users of the system. To do so it relied on an elaborate password-guessing scheme that, owing to the carelessness with which most users choose passwords, proved rather successful. When it had a legitimate user's password, the worm could pretend to be the user in order to read what he or she may have had in the computer's memory and to discover the names of other computers in the Internet that it could also infect.

According to Spafford, most of the UNIX features (or "misfeatures," as he calls them) that allowed the worm to function as it did have been fixed. Yet the fact has not allayed his worries about computer security, as the quotation at the beginning of this article reveals. Perhaps he was thinking of Cohen's theoretical investigation of viruses, which might apply to worms just as well.

If technopaths insist on vandalizing computer systems, it may be time to form a Center for Virus Control. During the Internet worm crisis teams at the University of California at Berkeley and a few other Internet stations were able to capture copies of the worm, analyze its code and determine how it worked. It would seem reasonable to establish a national agency that would combat computer viruses and worms whenever and wherever they break out—particularly if computer infections are destined to increase. Although the Internet experience hinted at the horrors that may still come, it also showed the efficacy of an organized resistance against them.

Cohen has established that it is impossible to write a computer program that will detect every conceivable virus, even though a defense can be constructed against any given virus. On the other hand, for any such defense there are other viruses that can get around it. According to Cohen, this ominous state of affairs might subject future computing environments to a kind of evolution in which only the fittest programs would survive.

The situation is reminiscent of Core War, a computer game I have written

about in previous columns [see SCIENTIFIC AMERICAN, May, 1984, March, 1985, and January, 1987]. But a Core War program does not bully innocent systems. It picks on someone its own size: another Core War program. The two programs engage in subtle or blatant conflict in a specially reserved area of a computer's memory called the coliseum. There is no danger of a Core War program ever escaping to do damage in the real world, because no Core War program or anything like it would ever run effectively in a normal computing environment. Core War programs are written in a language called Redcode that is summarized in the table on the opposite page.

Perhaps a simple example of such a program will serve to introduce the game to readers not already familiar with it. Here is a program called DWARF that launches a 0-bomb into every fifth memory location:

```
DAT −1
ADD #5   −1
MOV #0   @−2
JMP −2
```

The memory coliseum that all Core War programs inhabit consists of several thousand addresses, or numbered memory cells, arranged in a long strip. The instructions that make up DWARF, for example, occupy four consecutive addresses in the coliseum, say 1001, 1002, 1003 and 1004.

The DAT statement serves to hold a value that will be used by the program (in this case −1) at the address 1001. The ADD statement adds the number 5 to the location that is −1 units away from the ADD statement. Since the ADD statement has address 1002, it adds 5 to the number stored at the previous address, namely 1001, changing the −1 to a 4. The MOV command moves the number 0 into the memory cell referred to by @−2. Where is that? The address is found by referring to the DAT statement two lines in front of the MOV command. There one finds the address where the program will put the number 0. The final command, JMP, causes execution of the DWARF program to jump back two lines, to the ADD command. This begins the process all over again.

The second time around, DWARF will change the contents of the DAT cell to 9 and then deliver a 0 to that memory address. If an enemy program happens to have an instruction at that address, it will be rendered inoperable and the program will perhaps "die" as a result.

In this manner DWARF goes on dropping 0-bombs on every fifth location until it reaches the end of memory—but memory never ends, because the last address is contiguous to the first. Consequently DWARF's bombs eventually begin to fall nearer and nearer to itself. Yet because DWARF is only four instructions long and the number of memory cells is normally a multiple of 10, DWARF avoids hitting itself and lives to fight on—albeit blindly and rather stupidly.

Over the past few years Core War has evolved into a rather sophisticated game with numerous strategies and counterstrategies. There are programs that spawn copies of themselves, that launch hordes of mindless one-line battle programs and that even repair themselves when they are hit.

The International Core Wars Society, which currently has its headquarters in Long Beach, Calif., and branches in Italy, Japan, Poland, the Soviet Union and West Germany, organizes annual tournaments in which a programmer's skills are put to the test. Readers interested in joining a Core War chapter should contact William R. Buckley at 5712 Kern Drive, Huntington Beach, Calif. 92649.

In the 1987 tournament the Japanese entries gave the North American warrior programs a run for their money. The winner of the most recent tournament, held last December in Los Angeles, was a program from the Soviet Union called, oddly enough, COWBOY. Written by Eugene P. Lilitko of Pereslavl-Zalesky, a small city northeast of Moscow, COWBOY appeared to watch for "bombing runs" by enemy programs, to move itself out of harm's way and then to retaliate massively. Lilitko won the first prize of $250. The second prize of $100 went to Luca Crosara of Pistoia, Italy. The third prize of $50 was won by Douglas McDaniels of Alexandria, Va.

In closing I quote Spafford once again: "Writing and running a virus is not the act of a computer professional but a computer vandal." Let those who would even contemplate such an act try Core War instead.

I should like to thank Cohen, Spafford and John Carroll, a computer-security expert at the University of Western Ontario, for help with this article.

FURTHER READING

COMPUTER VIRUSES. In *Computers and Security,* Vol. 7, No. 2, pages 117–125, 139–184; April, 1988.

COMPUTER VIRUSES. Peter J. Denning in *American Scientist,* Vol. 76, No. 3, pages 236–238; May–June, 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.

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. 51593. 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, on 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

BOOKS

The music of physics, the coal of China, grains of truth, the citrus fix

by Philip Morrison

LONGING FOR THE HARMONIES: THEMES AND VARIATIONS FROM MODERN PHYSICS, by Frank Wilczek and Betsy Devine. W. W. Norton & Company ($19.95).

At human time scales, a chair endures. It remains a chair; its form is conserved. Vaporize the stuff, though, and you cannot tell even whether it had been one chair or two, or anything at all about its shape. Only the molecule count would remain, strictly unchanged. Heated still more, the chair might as well have been a matched mix of atoms from a chemistry set. "And so on.... A grand dream of physics is that ... at a high enough temperature a pound of chair parts or of gold, Gutenberg Bibles, or positron plasma would all turn out to be the same thing.... The more extreme the conditions, the fewer the conserved structures we need to worry about ... we understand the interior of the sun better than the interior of the earth, and the early stages of the big bang best of all."

It would be hard to find a deeper view of the nature of thermal equilibrium. All mere technicalities have been stirred into the smooth mixture; we are left not in colorless abstraction but instead with a sense of great gain in breadth of vision. Not forgotten is the fact that equilibrium is only approximate.

In a carefully ordered array of 32 concise and graceful essays of like quality, all in cultivated voice and with no hint of the oracular, Frank Wilczek and Betsy Devine have set out the harmonies as physics now sees them. True to the musical model, the short chapters fall into thematic groups on the high themes of modularity, uniformity, transformation and inevitability. A second round of variations on the same themes follows, now arranged for a more remote context. Add two less-controlled movements, one on quantal reality and another on the power of symmetry, and the 10-part work is outlined.

Of course, this engaging and original book is not merely a harmonious score. The insights and pleasures invoked here are rendered as concretely as any composer could hope to render the sonorities of brass and woodwind. Among very few numbers and a handful of diagrams, the general reader will find an artful account of many of the concepts and results on which our view of macrocosm, microcosm and their new-found relation now rests.

Consider a few examples of topics that are given simple but explicit explanations. The fuel of the sun is the unburned remnant of the incompleteness of the big bang: the sun uses what was earlier rejected, as the peasant kindles flame from the boon of cow dung. Similarly, "all of Earth glows in the dark." That is, the radioactivity that heats and mobilizes the interior of our planet to drive the slow drift of continents is the lingering fluorescence of star explosions long past.

There is human conviction in a personal narrative of Wilczek's part in discovering the unexpected disappearance of all forces between quarks when quarks come close together. "Very early one morning, I got a very excited and mysterious phone call at home from him. (Actually, it was around noon but in those days I worked late into the night....) ... he was eerily calm: 'We'd better check this. I think we have asymptotically free theories.'" The caller was David Gross of Princeton, Wilczek's young thesis supervisor, and their grand joint result was exactly what Gross had intended to prove no quantum field theory could ever produce: a predicted force between two interacting particles that is vanishingly weak when the particles are close enough but grows stronger and stronger as the two move apart. To suggest tests, there followed "quite a bit of calcula-

tion, but it was pretty straightforward .... We made a game of it, racing ... to calculate the graphs faster. David was usually faster, but when our answers disagreed mine was usually right .... I'm eternally grateful that asymptotic freedom discovered me."

Not everything here is so agreeable, even though there is coherence. Some witty turns of phrase will fall flat for any reader; this reviewer was not won by the neologism contrived to describe the particles of this quantal world as laves, made from lumps and waves. Less idiosyncratic but no more persuasive are the treatment of the famous paradox of Schrödinger's unhappy cat and a metaphysical view of the world that regards each quantum transition as generating new universes, endlessly branching. Those scholars of old Paris and Salamanca, whom we ridicule for counting hypothetical angels dancing on their pin stage, were less darkly metaphysical than that. Professor Wilczek is a justly celebrated leader in his brilliant generation of theorists, and he is fully entitled to his philosophies, but it might have been better to soften issues whose tentative resolution is so breathlessly farfetched.

Among many recent summations of both the established and the hoped-for triumphs of field theory and its cosmological impact, this is the most imaginatively written and therefore possibly the most instructive for attentive readers at an introductory level. The two authors (Betsy Devine is a writer with engineering training) have by their art caught us up in the music, even if so far we have only a grand unified puzzle. They remind us that the joy is in a searching dialogue with nature, not in wishful thinking about the unread message. "Success is expected, but never guaranteed."

ENERGY IN CHINA'S MODERNIZATION: ADVANCES AND LIMITATIONS, by Vaclav Smil. M. E. Sharpe, Inc., 80 Business Park Drive, Armonk, N.Y. 10504 ($37.50).

Not so long ago any penetrating appraisal of current economic and environmental affairs in China was something like archaeology—based, as Professor Smil puts it, on the piecing together of many "shards of information" into shadowy larger shapes of reality. No longer. This penetrating, friendly if critical book by a geographer from the University of Manitoba draws on a wealth of detailed Chinese publications, largely from the 1980's. The task is now to find pattern in the

flood of consistent new data, and he succeeds admirably.

The daily necessities for a billion people are dauntingly diverse. Widespread changes in their energy supply and consumption are of necessity intricate, often marked by surprises, missteps, even paradox. China is in part one of the large, modernizing industrial states, with some 200 million people living more or less by city ways. But around them are 800 million villagers, more than a third of them adults hard at work on the land. It is no small achievement that China is able to feed its people—one-fifth of humankind with access to only one-fifteenth of all cultivated land—with a calorie supply per person not much below Japan's. Of course, there are plenty of ill-fed people still: a tenth of the population, mainly in interior provinces.

What is most striking is that the Chinese countryside lives by coal as well as by sun-grown grain. Coal provides two-thirds of the country's ammonia-based fertilizer, the nitrogen fixed in a large number of small and inefficient synthesis plants. The rest of the essential commercial nitrogen is produced in large, efficient new plants (provided by the U.S. firm M. W. Kellogg) fueled by natural gas. The famous recycling of organic wastes, including night soil, still goes cannily on as it always has, but now that turnaround can supply at most a fourth of all the nitrogen applied to the crops. (Much comes as well from the symbiotic microorganisms of the rice paddy.)

The overall outcome is unexpected: the energy supplied for agricultural production to each hectare of cultivated land in China is double the amount supplied in the U.S. It is about the same as that of prosperous France, although only a fourth of what prevails in the land-poor Netherlands. The reasons are plain: fertilizer is heavily used in China, and mechanical power too is essential for extensive irrigation by pumped water. Mechanization of fieldwork is still minimal. Only about a fifth of all the energy brought to the farm goes to build or to fuel farm machinery, tractors or trucks. Yet the populous countryside is seriously energy-short. About half a billion people lack fuel to stay warm or to cook three hot meals a day during the cold seasons. When in winter the housewives of one county in cold Gansu "face the stoves, tears come to their eyes."

These hardworking people endure the bitter energy crisis of the world's poor: not the lack of gasoline but the scarcity of fuelwood. For rural China even the term fuelwood is a kind of euphemism. Four-fifths of household energy in the Chinese village is gained by burning biomass, all right. But about half of that fuel is from crop residues, the cereal straws, stalks and vines from the fields. The other half is from the woods: branches, twigs, roots, needles and bark raked from the woodland floor, even dried sod and grass tufts lugged home in the back baskets of women and children, some 20 pounds per household per day. (Biogas from wastes and cattle dung furnishes a few percent.)

It is here that a simplistic evaluation of energy efficiency on the farm as the ratio of fossil-energy subsidy to photosynthetic gain loses touch with reality. Even the frugal winter-pinched households of village China consume three or four times more fuel energy for daily living than is spent on their fields, the fuel burned at low efficiency in traditional stoves. Long-range hope rests in growing food with less human labor on the land. What country people most need in the near term, along with the extensive spread of high-efficiency household stoves (some clever new models are shown here), is above all "the greatest possible extension of private fuelwood lots."

China's abundant coal is mined by the largest corps of underground workers in the world, two or 2.5 million miners at the coal face, working mostly in small, local, shallow mines and open pits with shovel, pickaxe and barrow—a simple handcraft, arduous and often risky, with open flames instead of enclosed safety lamps. It is a big mine of this class that can manage to ship a single truckload of coal per day. The mean yield per shift belowground is about what it was in late-Victorian Britain. Only the U.S. now rivals China for overall coal production, and our tonnage is won with less than a tenth as many miners. Half of Chinese coal is the quality product of a few large mechanized state mines, but the rest of it is the raw, dusty and unsorted yield of 50,000 country mines, more of which are being opened every day. For it is coal that drives Chinese railroads and Chinese industry, rural and centralized production alike; two-thirds of all the energy used in China comes from coal.

The rivers of China plunge from the roof of the world down to the crowded coastal valleys and lowlands. No other country has so much hydropower potential. Here too experience has brought surprise, in some ways the reverse of the story of coal. Some 70,000 small hydro plants dot the countryside. This appropriate technology is based on cheap, effective turbines, made available in 80 varieties for placement in any irrigation ditch or stream. This development is an internationally acclaimed success that surely rests in part on China's "unmatched ancient tradition of hydraulic engineering."

On the other hand, the grandest of the Sino-Soviet hydro schemes of the 1960's, a giant dam in the rocky gorge of the Three Gates, below the big bend where at last the Yellow River leaves the easily eroded loess, has miscarried. Its basin was silt-filled within a few years, and the headwaters backed up to dangerous flood stage. The dam had to be pierced to let the river flow free; the high hopes for Sanmenxia ended in "a low head run-of-the-river plant," its power capacity cut sixfold, with no potential for flood control. The fearful passage of the Gates has had no pity on human effort since grain first traveled the river in the days of Han.

There are lofty plans afoot: coal-slurry pipelines to take some of the heavy burden of coal off the railway system, where it is the bulk of all cargo; modern open-cast coal mines to fuel power plants right at the mine mouth, sited along the rich coal seams of Shanxi at a density that will match that of the Ruhr or the valleys of the Ohio-Pennsylvania border. Expensive French-licensed nuclear reactors will soon be coming on-line, and new major hydropower schemes, including the largest power plant in the world, are rising now within the Three Gorges of the Yangtze, China's largest river.

Nothing comes easy. The health of land and air will be at multiple risk, both in the shadow of the big projects and in the path of the undirected "coal rush" of ambitious peasant miners. China must and will grow richer, although more slowly than hopeful planners reckon. Efficiency in every detail is needed—say more diesel trucks, and cheap sedans for road-bound officials instead of expensive four-wheel-drive vehicles. Frugality, productivity and prompt attention to externalities are as important as ambitious plans.

The thematic maps, the careful tables and graphs, rough estimates from first principles used to explain and test the data, and a lively sense of each activity in its context give this

model of a book its authority and interest. Over the years the author has prepared half a dozen expert volumes, several of them studies of China, others ranging beyond that land. (His provocative worldwide survey, *Energy-Food-Environment,* appeared in 1987 under the imprint of the Oxford University Press.) Certainly no one's prophecy in these high matters will always prove right, but Smil's sober treatments seem a mark against which other views should be tested. He does not usually share the highest hopes or cry the loudest alarms; for him Rome was not built, nor did it fall, in a day.

The epigraph he cites—and then caps—is from Master Meng, eloquent apostle of Confucius. It hints at Smil's unfailingly humane concern for China's prospects and our own: "'If you do not interfere with the busy seasons in the fields, then there will be more grain than the people can eat; ... if hatchets and axes are permitted in the forests ... only in the proper seasons, then there will be more timber than they can use.... When the aged wear silk and eat meat and the masses are neither cold nor hungry, it is impossible for their prince not to be a true king.' Would Deng Xiaoping disagree?"

SAND, by Raymond Siever. W. H. Freeman and Company ($32.95).

That countryside New England poet with an ear for expressive metaphor reported two options for the world's end: fire or ice. This satisfying introduction to modern sedimentology forecasts something quite different. Look ahead a few billion years: the internal heat of the earth will be spent and the currents within will be inadequate to shuffle the plates of the crust. The ocean ridges will cool, all drifts will stop, volcanic activity and mountain building will come to an end. The only plate boundaries will be as passive as the present Atlantic Coast of North America.

Yet the sun will still drive the circulation of wind and water, so that the rocks will weather until all the continents have eroded to broad, gently rolling plains. Over all the dry land, and brimming out to the continental margins, will lie the final stable formation of all: a ubiquitous bed of quartz-rich white sand. Its surface will weather too, until in time a baked bauxite-rich mud, leached of nutrients, will thinly cover the sand-strewn lands. The curtain will have fallen on all geologic change. What will happen to climate, air and sea we do not know; the effects on life we may leave to those "imaginative guides to the future, science fiction writers." But the earth ball itself will have reached a state close to a dusty end.

Ray Siever writes with a vivid imagination under careful control. His long Harvard career has spanned the geologists' revolution of plate tectonics. He was among a group of students of sediments who were led during the 1950's by an odd scientific couple. One of the two leaders was Francis Pettijohn; "always kind, he could be a quiet tease, especially at an outcrop." He was a low-key but deep questioner in the style of the small towns of the American Middle West. The other, Paul Krynine, was a different sort: an aristocrat, a young White Russian émigré, "subtle and witty, ... an ebullient agent-provocateur.... These two men, one stormy, the other quiet, remade the field."

Their group read the sandstones, light, dark and red, pure and mixed, marked everywhere with signs of winds and streams and volcanic outpourings, to map out most of the continents throughout the past. "In hindsight, I imagine we were honing our tools, getting ready for the big event." It arrived, plate tectonics, in the mid-1960's. It was quickly embraced by these sedimentologists, whose disparate jigsaw puzzles it soon brilliantly united. Sand grains imply the rise and decline of mountain ranges as the great plates clash, part and slip.

The tools for studying sand are many. The most direct is field geology. From visible textural contrasts called cross-bedding, familiar in showy photographs of the bare red sandstone cliffs of our West, ancient stream directions can be inferred. They gave a single purpose to field geology. The customary mapping, recording bit by bit all that is there to be seen in the rocks, takes years of patient walks, hammer in hand. "Our plan was simple ... ignore all the other interesting geology; just measure the cross-bedding" over six states.

"[Paul E.] Potter and I," the author recalls, measured the cross-bedding direction of "two cross-bedded units at each of two outcrops in each of two ... sections in each of [the] survey townships along the entire outcrop belt" in a matter of weeks. The two young men drove all day at breakneck speed from one sight of an outcrop to the next. One jumped out to measure at each outcrop, the other recorded the data in the notebook; they hardly paused even for a cold drink in the hot Middle Western afternoon. Such commando sand mapping soon disclosed the paleocurrents of a vast river system hundreds of millions of years old. From many such recoveries the drainage of entire continents began to appear, epoch by epoch.

Yet fieldwork is not everything. Geologists require laboratories too. They may wash away sands to simulate submarine currents or study the microprocesses that part, work, sort, carry and reunite sand grains. The usual chemical handbooks list various minerals as "insoluble," but the geochemist knows better. It is of course the near-insolubility of quartz that allows pure quartz sands; the off-white feldspars of the granite have dissolved and the memory of the entire rock from which the sands were made grows dim. In cold mountainous Alaska, freshly made sands retain the full feldspar-rich composition of their parent granites; in the humid flatlands the slowly formed grains that endure are nearly all quartz. In this way the sands bear witness both to landform and to rock type.

Sand can be reworked, and new sandstones arise out of old ones; such reworked grains are common, often the result of many cycles. The surface texture of the grains is another clue. A frosted surface is easy to see on sand grains, even with the unaided eye. It has been shown that such a surface cannot physically arise from the rather feeble collisions of grain with grain, although in general it does appear to be the sign of windblown sand. Frosting seems to originate mainly in the slow chemical etching of quartz by water, a tedious version of the etching of glass by hydrofluoric acid. "I once monitored the dissolution of quartz in distilled water at room temperature for two years and watched the silica in solution climb to only 1 ppm (parts per million) in all that time."

Continental granites are the source of most sands, since rocks are dissected most easily along the grain boundaries of their constituent crystals. Granites are usually made of grains of sandlike size, both of durable quartz and of less enduring feldspars. Oceanic and volcanic rocks, dark basalts, are not frequent sources of sand: they are low-lying and hence little weathered, the grain texture is so fine that the loosened grains soon dissolve and many of the rocks are glassy, without grains. It is the continents that bear and transport sand, and the history of sand discloses the changes of major landmasses.

Chapter by chapter Siever leads us

to examine sand, analyze its travels and consider its places of origin and deposit, to look at mobile beaches, visit sea canyons and shelves and follow sand's long evolutionary round through the epochs. His tables and photographs are both striking and germane. We go among the sedimentologists as well, mainly by way of the author's own experiences. Inevitably, the book begins and ends by citing William Blake, who saw "a World in a Grain of Sand." This scientist fully justifies that challenging fancy by giving us a look into his subdiscipline that is unexpectedly personal, fresh and compact with the poetic reward of unity.

THE HISTORY OF SCURVY AND VITAMIN C, by Kenneth J. Carpenter. Cambridge University Press ($12.95).

"The year 1907 saw the publication of what I judge to have been the most important single paper in the whole history of this subject." Two barely remembered Oslo medical researchers, investigating beriberi, found that guinea pigs fed only on grain developed a disease analogous to scurvy in humans. They had happened on an animal model for the study and assay of antiscorbutics; the way was open to the concentration and finally the isolation of vitamin C, ascorbic acid.

Ascorbic acid is found in the tissues of all land animals. Most mammals even make it on their own in the liver or kidney and do not need any in the diet, but the enzyme for the last step in the assembly of the decorated sugarlike ring has been lost by fruit bats, primates and guinea pigs. Ascorbic acid was isolated in 1932 in a dead heat between two groups, one moving from scurvy by the guinea-pig route, the other taking a biochemical path that was entirely different until the final tests.

The story of vitamin C opens long before modern biochemistry. In 1498 the navigator Vasco da Gama led the first four European ships around the Cape of Good Hope. After 10 weeks at sea "many of our men fell ill... their feet and hands swelling and their gums growing over their teeth so they could not eat." In the East African ports they ate oranges, and in a week they had recovered. On the way back the illness came again, and again fresh oranges, "much desired by our sick," proved curative. They had endured a new specific disease (close to half of the men died at sea) and at the same time had stumbled on its all but magical cure. Over the next three centuries some two million sailors on long voyages under sail (and many an African slave as well) would die of scurvy, the worst of all occupational illnesses.

In 1747 Dr. James Lind, naval surgeon on the man-of-war H.M.S. *Salisbury,* tried what may have been the first of all controlled clinical experiments. Twelve sailors, all with scurvy as "similar as I could have them," were brought to the same quarters and fed the same diet, save that for 14 days the men were allocated in pairs to one of six regimes esteemed for treating the disease. One treatment was two oranges and a lemon each day; the others varied widely, from a pint of seawater daily to vinegar with each meal. The oranges-and-lemon patients improved, and they nursed the others. "Their experienced virtues have stood the test of near 200 years," Lind said of the citrus.

It took 40 years for the Royal Navy (under an influential Physician to the Fleet who was also personal physician to the powerful Admiral Rodney) to begin the regular issue of lemon or lime juice. Scurvy ceased to be a problem to the limeys of the Royal Navy, who were issued three-quarters of an ounce of juice a day, and on the East India Company's vessels, from Nelson's time onward. By mid-century steam had shortened all routine sea voyages; fresher food was served, and scurvy left the fleet.

On land scurvy claimed convicts, forty-niners in California, prisoners of war, Irish farmers deprived of fresh potatoes by the blight, Arctic explorers (but not the Inuit hunters around them) and the people of more than one besieged city. During late-Victorian years many infants of affluent London and New York families whose parents bottle-fed them on condensed milk, rich and free from infection—but from vitamin C as well—succumbed to "Barlow's disease." The condition mimicked adult scurvy, except that the infants had no teeth to lose. Internal hemorrhages seen at autopsy revealed the cause: frank scurvy. Breast-fed infants escaped; although czarist Russia, ill-fed and old-fashioned, was still an endemic center of land scurvy in adults, "Barlow's disease did not seem to occur there." (Mother's milk, we now know, can contain up to seven times the ascorbic acid concentration of maternal blood plasma.)

By the mid-19th century, when the mariners did not much need their lime juice, the lime regime seems to have lost its efficacy; several Arctic naval parties well provided with regulation

lime juice suffered heavily from scurvy. The reason is not certain, but it is probable that the heating and copper plumbing involved in packaging the juice severely reduced its vitamin content. The soluble vitamin, present only to the weight of a pill in a week's ration of juice, is unstable against heat, exposure to air and copper salts, and is leached away by boiling.

The cure for scurvy had been quickly found by simple experience. Yet theory and experience, both essential to science, are always in creative but dangerous tension. In the case of scurvy that antagonism persisted through four centuries. Every incomplete theory reigned over its harmful day. Was it the fruit or was it the good air of green land that had cured the seamen (as bad night air brought malaria)? Voyagers to new places often found scurvy; to be sure, those were longer voyages than usual. Once chemistry had developed some strength, theorists argued for the beneficent effect of sulfuric acid and other acids that might be more practical than fresh lemon juice. Once Justus von Liebig had proclaimed the key role of proteins, it seemed clear that milk and grain products were the real center of health: how could a diet adequate in flour, butter, eggs, cheese, well-cooked meat or fish be at fault? Once the role of oxygen was grasped, forecastles were redesigned for fresher air; it did not help scurvy. Was an active, cheerful, "dancing" crew healthier? Maybe citric acid was the real answer? Or was the cause the tainted meat, the ptomaines, that dogged all long journeys and military campaigns?

By now 50 million Americans eat some of the pure vitamin C crystals daily, and many take megadoses of a gram or two, which have been recommended against colds, cancer—even mental illness—by one or two of the finest chemists of our time (a proposal at least prima facie unsupported by careful trials). We know that vitamin C is an essential coenzyme in the formation of the structural fiber collagen; without it the connective tissues are weakened. Why, though, is it stored in fruit? (Plants make no collagen at all.) Ascorbic acid is much used in food processing as an antioxidant. Perhaps it is part of the preservative packaging of fragrance, color and sweetness that some plants prepare to encourage the transport of their seeds. This volume by a senior nutritionist at Berkeley is an absorbing and scholarly history, a prime bargain in paperback.

ESSAY

Litigation thwarts innovation in the U.S.

by Peter Huber

American innovation is in trouble in the courts. Burt Rutan, the pioneering designer of the *Voyager,* used to sell construction plans for novel airplanes to do-it-yourselfers. In 1985, concerned about the lawsuits that would follow if a home-built plane crashed, he took the plans off the market.

The Monsanto Company decided in 1987 not to market a promising new filler and insulator made of calcium sodium metaphosphate. The material is almost certainly safer than asbestos, which it could help to replace in brakes and gaskets. But safer is not good enough in today's climate of infectious litigation.

Liability fears have caused the withdrawal of exotic drugs that the Food and Drug Administration considers safe and effective, including some for which no close substitute is known.

In the past 15 years most companies have halted U.S. research on contraceptives and drugs to combat infertility and morning sickness. "Who in his right mind," the president of one pharmaceutical company asked in 1986, "would work on a product today that would be used by pregnant women?"

Liability is supposed to fall on "defective," unduly dangerous products and services. What has gone wrong?

The old rules of negligence, which lasted until the 1960's in most states, looked closely at the human actors on the scene. Were they careful? Had they been well trained? Thus tested, engineers, surgeons, chemists and pharmacologists at the leading edge of their professions fared well.

The new rules of "strict liability," invented by U.S. courts in the 1960's and 1970's, place technology itself in the dock. After an accident jurors are given a few days to evaluate the design of a mass-vaccination program, a power plant or an advanced military aircraft. Sympathy for the victims clouds the analysis, and if finding a design defect is what it takes to help out the unfortunate claimant, then that is what many juries find.

Moreover, human nature is predisposed to accept the old and familiar risk while rejecting the novel and the exotic. Cigarette makers usually win before the jury, whereas pioneers in gene splicing or laser surgery are at constant peril. By the same token, consulting engineers favor older design options in their specifications, fearing that new ones will carry greater risk—not physical risk but legal.

The various elements of liability in the courts today all join to thwart innovation. Take the duty to warn of hazards, great and small, common and bizarre, in staggering detail. It is a game that sellers master only by playing for a long time. The warnings on birth-control pills have been honed for 30 years and now run on for several pages of dense detail. No equivalent warning can be offered for a next-generation mode of contraception, even if on balance it is safer.

Modern law also demands that risky products come, in effect, with their own insurance contract attached, underwritten by some producer's liability insurer. Insurance, by design, spreads costs broadly and somewhat indiscriminately; when one product comes under intense liability attack, an entire industry may lose its coverage. For the prudent business no insurance usually means no product.

The most regressive effects are felt precisely where fruitful innovation is most urgently needed. Liability today is highly—and often indiscriminately—contagious. Progress is undercut the most in the markets already battered by a hurricane of litigation: contraceptives, vaccines, obstetrical services and light aircraft, for example.

More often than not the best anticipatory defense in the modern legal environment is to sit still. Age, familiarity and ubiquity provide the surest legal protection. When it encourages improvement at all, today's liability system promotes the trivial and marginal change. The drug manufacturer endlessly fine-tunes the warning or microscopically adjusts the dosage in the capsule. The doctor orders more tests and X rays in order to pile up a protective paper trail. Companies hire squadrons of risk managers, industrial hygienists and consumer psychologists. Liability-driven safety management becomes a mirror image of the legal process itself—fussy, cumulative, bureaucratic and preoccupied with paper.

Meanwhile the threat of liability impedes or prevents the sharp break with technological tradition, the profound change in method or material, design or manufacture. Over the long term, however, the bold leap forward is all-important in the quest for safety, and it is precisely in the riskiest areas of life, where the litigation climate is worst, that such change is most urgently needed.

Today's U.S. liability system, unique in the world in its reach and impact, is all too adept at condemning services and technologies deemed unacceptable for one reason or another. What it lacks is a reliable way to say yes.

What is the solution? When we deal with essentially private risks (in transportation or personal consumption, for example), fair warning and conscious choice by the consumer must be made to count for much more than they do today—not because individual choices will always be wise (they surely will not be) but because such a system at least allows positive choice and the acceptance of change.

Informed consent by the individual is not, however, going to take care of such complex or far-reaching safety issues as chemical-waste disposal, mass vaccination or central power generation. Those are, and obviously must continue to be, delegated to expert agencies acting for the collective good. But if they are to be useful at all, agents must be able to buy as well as sell. For safety agencies this means not only rejecting bad safety choices but also embracing good ones. Yet the long-standing rule, vigorously applied by our courts, is that even the most complete conformity to applicable regulations is no shield against liability.

The courts should be strongly encouraged, instead, to respect the risk and safety choices made by expert agencies. It may be politically unrealistic to propose that liability should be entirely foreclosed even in cases where activities are conducted with the express approval of qualified regulatory agencies, but surely it could be firmly curtailed in such circumstances. At the very least, full compliance with a comprehensive licensing order should provide liability protection against punitive, if not compensatory, damages. It has always been true that ignorance of the law is no excuse. Today knowledge of the law is no excuse either. It should be.

PETER HUBER, a senior fellow at the Manhattan Institute for Policy Research, is the author of *Liability: The Legal Revolution and Its Consequences,* published last fall by Basic Books.

WHEN EVERY NANOSECOND COUNTS.

The 1989 Toyota Camry V6 achieves impressive acceleration without apparent regard for the laws of physics. Its hyper-powerful four-cam, 24-valve, 153-hp, V6 engine is a formula guaranteed to be your equation for sporty driving. And despite Camry's responsive stance, it'll amaze you with its luxury-car room and comfort. Toyota Camry V6. It only takes a nanosecond to fall in love.

A **36-month/36,000-mile** basic new vehicle limited warranty with no deductible and no transfer fee applies to all components other than normal wear and maintenance items.

Call 1-800-GO-TOYOTA for more information and the location of your nearest dealer. Get More From Life... Buckle Up!

TOYOTA QUALITY

WHO COULD ASK FOR ANYTHING MORE!