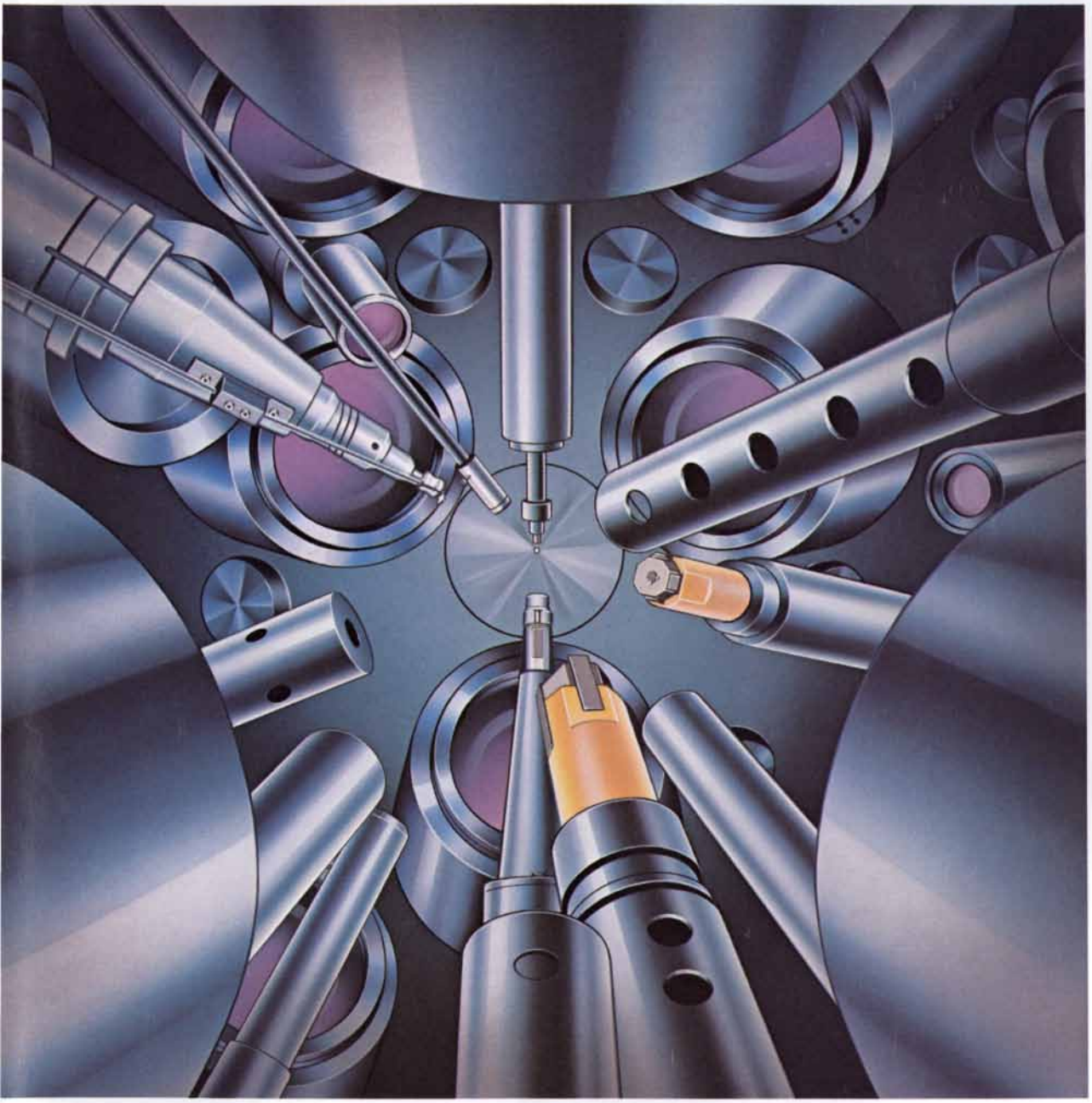
SCIENTIFIC AMERICAN



LASER FUSION

$2.50

August 1986

Hitachi's wide-ranging technologies in communication (from left to right): optical fibers, image signal processor, advanced telephone exchange system, satellite communication, and machine translation system.

COMMUNICATION

Communication is not simply sending a message...it is creating true understanding—swiftly, clearly and precisely.

"I know he's trying to tell me something: but what does he really mean?" In our world of proliferating technologies and new terminology, this kind of question is asked a lot. Here is what we are doing about it.

Hitachi's scientists and technicians' long-term goal is to break the language barrier. They are diligently at work today on an array of projects that will vastly improve the communications of tomorrow.

For example, we've made tremendous progress on a system to translate Japanese into English.

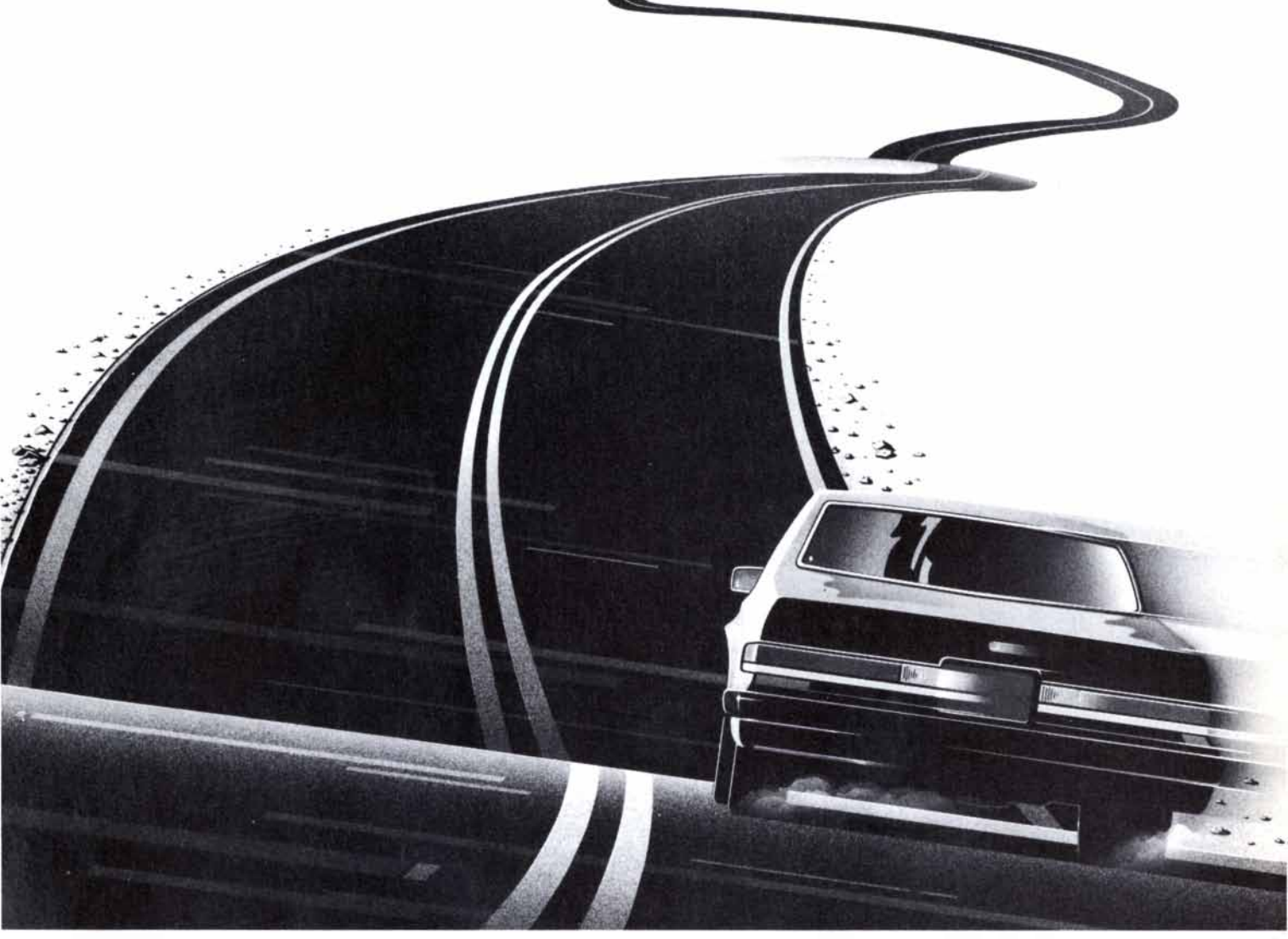
This system can be used to translate various scientific/technical papers and machinery/equipment manuals. Special "glossaries" can be developed to adapt it for fields as diverse as medicine, electronics and aeronautics. Further development could lead to automatic telephone translation or even portable verbal translators for travelers.

In addition to the machine translation system, Hitachi's research specialists are also developing advanced transmission systems that send your phone calls or business data across great distances using hair-thin optical fibers and laser beams. They are also working on other new methods of communications, such as advanced telephone exchange systems, satellite communication systems, TV conferences, and so forth.

At the root of much of this is our highly advanced computer technology: because Hitachi is producing some of the fastest, largest-capacity systems available today.

We link technology to human needs. We believe that Hitachi's advanced technologies will result in systems and products that are functionally sophisticated but easy to use. Our goal in communications—and transportation, energy and consumer electronics as well—is to build products and systems that will improve the quality of life the world around.

HITACHI



All you asked was to go faster and farther. With more power. And less fuel.

In many of the finest American cars, Bosch fuel injection components help deliver precisely the amount of fuel needed for any driving condition.

And those American cars are going faster and farther. Producing up to 20 percent more horsepower. Using less fuel. And reducing exhaust emissions.

Power as you need it.

Bosch multi-point systems inject fuel right at the individual cylinders.

A microcomputer reads thousands of engine operations and driving conditions every second, and dictates the most efficient mix.

So Bosch fuel injection not only gives you more power when you need it. It can boost mileage by up to 15 percent.

Traveling in the best of companies.

Bosch fuel injection arrived in the U.S. in imported cars over 30 years ago. It's since evolved from a curiosity to a high-technology must for today's best performing machines.

Today, every U.S. manufacturer uses Bosch fuel injection systems or components.

Virtually every European model sold in the U.S. is equipped with Bosch fuel injection. And every fuel-injected Japanese model uses systems or components produced under Bosch license or using Bosch patents.

Bosch breaks ground.

Sophisticated problem solving is what Bosch is all about.

We supply every major U.S. and European automaker with an array of advanced products, from high-performance spark plugs to anti-lock braking systems. But our expertise isn't limited to automotive.

We're a leading manufacturer of high-technology products, such as Bosch assembly systems, power tools, packaging machinery, TV studio equipment and video graphics, home appliances, and medical equipment.

In fact, we've pioneered literally thousands of the world's most sophisticated electronic and electro-mechanical advances.

As a result, we're an international corporation with manufacturing facilities in 16 countries. And with sales and service in 137.

But more important are 144,000 Bosch people—9,000 in R&D alone—who are capable of finding the right answer to the toughest challenges.

Including yours.

BOSCH

Bringing high technology down to earth.

Established 1845 **SCIENTIFIC AMERICAN** August 1986 Volume 255 Number 2

ARTICLES

DEPARTMENTS

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 RATE: $24 PER YEAR IN THE U.S. AND ITS POSSESSIONS; $33 PER YEAR IN ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017.

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, UT 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, NY 10017. Manuscripts are submitted at the authors' risk and will not be returned unless they are accompanied by postage.

Advertising correspondence should be addressed to C. John Kirby, Advertising Director, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, NY 10017.

Subscription correspondence should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 5969, New York, NY 10017. The date of the last issue on your subscription is shown in the upper right-hand corner of 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

THE COVER

The painting on the cover depicts the interior of the target chamber of the OMEGA laser system at the University of Rochester, where 24 laser beams simultaneously focus their energy onto a fuel pellet (see "Progress in Laser Fusion," by R. Stephen Craxton, Robert L. McCrory and John M. Soures, page 68). The fuel pellet is the small, bright sphere at the center of the painting; it is essentially a hollow glass shell, about a millimeter in diameter, that contains deuterium and tritium gas. The atomic nuclei of the gas fuse when the shell and the gas are compressed and heated by the laser; the fusion reactions could give rise to useful energy. The pellet is suspended by a filament, sometimes made of spider webbing, attached to a positioning device. Each laser beam passes through a lens and a blast shield that protects the lens from debris created when the pellet is vaporized. Antireflective coating on the blast shields appears purple. The pellet is surrounded by diagnostic instruments: for example, the orange-tipped devices at three and five o'clock are X-ray microscopes; the narrow instrument at six o'clock is an X-ray pinhole camera, and the conical-tipped device at 10 o'clock is a streak camera.

THE ILLUSTRATIONS

Cover painting by Hank Iken

Page	Source
18–20	Ilil Arbel
34	Ministry of Defence, U.K.
36–41	Ian Worpole
42	Saxe Patterson
44–50	George V. Kelvin
51	Leonid A. Bendersky and Robert J. Schaefer
53	Anders Björklund (*top*), Carol Donner (*bottom*)
54	Alan Fine
55–58B	Carol Donner
69	James Kilkelly
70–73	Gabor Kiss
74	Gabor Kiss (*top*); Frederic J. Marshall, University of Rochester (*bottom*)
75	Groot Gregory, University of Rochester (*top*); Johnny Johnson (*bottom*)
76	Eugene Kowaluk, University of Rochester
77	Gabor Kiss
78	Bruce LaFontaine, University of Rochester
81	Quesada/Burke
82	Paul C. Mangelsdorf (*left and middle*), Tom Prentiss (*right*)
83–85	Tom Prentiss
86	Bobbe Baker
89	George E. Williams
90	Andrew Christie
91	George E. Williams
92–94	Andrew Christie
95	Ronald N. Bracewell, Stanford University
98–99	Enid Kotschnig
100–104	Patricia J. Wynne
107	Norman Hammond
108–109	Alan D. Iselin
110–113	Norman Hammond
114	Lowell Georgia, © National Geographic Society
115	Norman Hammond (*top*), Alan D. Iselin (*bottom*)
116–118	Michael Goodman

"I can't, I'm wearing magnetic underwear."

—Computer backup excuse #685

Doing computer backup really gets under people's skin.

They'll do anything to get out of doing it, because backing up is so utterly boring.

But not doing it can definitely be a matter of getting your education at the college of hard knocks.

We're reminded of the PhD candidate in biology who did hunger research for his doctoral thesis without backing it up.

He had data from a year's worth of injecting and weighing rats stored on the disk. The computer crashed, the rat data was erased from memory, and he had it to do all over.

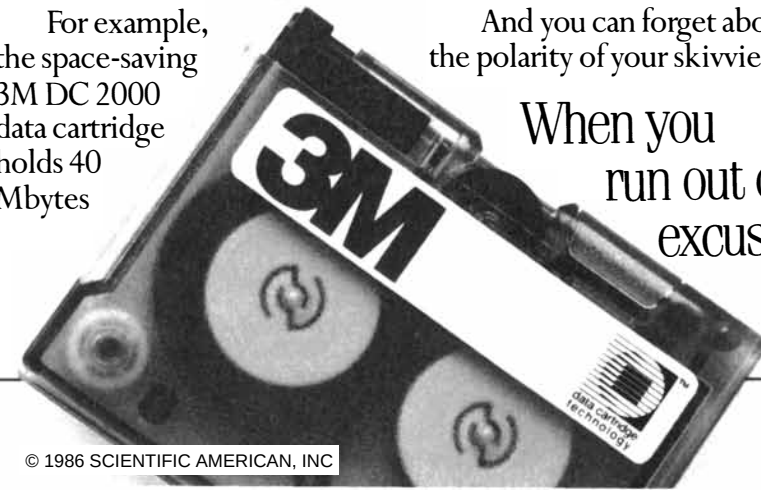
Back to the rat race.

So do your backup. Floppies if you don't have much memory. Or, if you have 5 to 10 Mbytes or over, on data cartridge – a 3M developed technology whose time has come.

For example, the space-saving 3M DC 2000 data cartridge holds 40 Mbytes of information. You'd need 80 floppies to store that much. And it can fit in your breast pocket.

To read more true-to-life "horror stories" of people who didn't back up, ask your computer products dealer for our leaflet "Horror Stories."

And you can forget about the polarity of your skivvies.



When you run out of excuses.™

LETTERS

To the Editors:

I should like to comment on the article by Miroslav Nincic, "Can the U.S. Trust the U.S.S.R.?" [SCIENTIFIC AMERICAN, April]. The first issue I should like to address is his incorrect assumption that Soviet arms-control violations are merely the result of treaty ambiguity. I should also like to turn to his implication not only that the U.S. and the U.S.S.R. are both guilty of pushing the limits of agreements or even of engaging in noncompliance but also that this behavior is an acceptable part of competition between superpowers.

Treaty ambiguity is, of course, a problem that the U.S. is working hard to avoid in its current arms-control negotiations. Regarding current agreements, where "loopholes" exist no violation has been charged. Soviet behavior with respect to these agreements, however, is characterized by a pattern of numerous verified, and hence verifiable, violations. Nincic acknowledges as much in the case of the Krasnoyarsk radar. Other examples include the Soviet violations of the 1972 Biological Weapons Convention (BWC) and the 1925 Geneva Protocol. The BWC is significant in that it is multilateral and bans the development, production and stockpiling of an entire class of weapons. The Geneva Protocol explicitly prohibits the use of chemical and biological weapons. The terms of these treaties are absolutely clear. The Soviets have nonetheless maintained an illegal offensive biological-warfare program and capability since they signed the BWC. They have been involved in the illegal production, transfer and use of lethal chemical and toxin weapons in Laos, Kampuchea and Afghanistan.

These cases constitute violations of the spirit, intent and provisions of key arms-control agreements. The judgments that these are violations were not made lightly. They were the result of exhaustive analysis of all the available evidence by the relevant agencies of the U.S. Government, including the Arms Control and Disarmament Agency, the Department of State and the Department of Defense. An extremely strict standard of evidence was applied before a violation was charged.

Nincic's second theme revolves around the notion that since the U.S. and the U.S.S.R. have both charged the other with noncompliance, both are equally culpable. This argument suffers from several flaws. First, the U.S. is strict in its treaty compliance—by law. Its defense programs are thoroughly examined to ensure their compliance with all U.S. treaty obligations. (The U.S. Arms Control Impact Statement, submitted to Congress annually by the Administration, does exactly that.) In our open society pressures from the public, the press and Congress would make U.S. violations impracticable—assuming they were ever attempted!

Moreover, while the Soviets have often used invalid countercharges of U.S. noncompliance to deflect sanction for their violations, the U.S. has gone far beyond its strict obligations to address any stated Soviet concerns. Unfortunately the Soviets have not been as forthcoming. U.S. statements of concerns and requests for corrective action have been made to the Soviets in the Standing Consultative Commission (SCC) and even by President Reagan during his talks with Mr. Gorbachev. The Soviets have not provided explanations sufficient to alleviate U.S. concerns, nor have they taken the actions necessary to correct existing violations. They simply make implausible arguments (as Nincic notes) and assert that they are in complete compliance.

Nincic missed a final point. Arms-control violations cannot be written off as understandable factors in superpower competition. Such an approach to arms-control compliance is unacceptable and somewhat cavalier. Soviet violations endanger the viability of the arms-control process, since compliance cannot be unilateral. For one side (the U.S.) to adhere and for the other side (the U.S.S.R.) not to adhere does not constitute real arms control. Rather, it constitutes a dangerous form of unilateral disarmament in the guise of bilateral arms control. This is no basis on which to build a foundation of real arms control that can contribute to national security and international stability.

DAVID F. EMERY

Deputy Director
U.S. Arms Control
and Disarmament Agency
Washington, D.C.

To the Editors:

Let me briefly address the main points raised in David F. Emery's critique of my article on Soviet SALT compliance. He and I obviously agree about the Krasnoyarsk radar, but Emery raises three additional issues. Two are relevant to the article, but his first point is simply a statement about Soviet adherence (or lack thereof) to the Geneva Protocol on chemical and biological weapons. Since I dealt exclusively with matters of SALT compliance, that is, with issues related to strategic weaponry, chemical and biological weapons were outside the scope of the article. I have not personally examined charges in this area, but I can refer the reader to the work on this subject by Matthew Meselson of Harvard University [see "Yellow Rain," by Thomas D. Seeley, Joan W. Nowicke, Matthew Meselson, Jeanne Guillemin and Pongthep Akratanakul; SCIENTIFIC AMERICAN, September, 1985].

The two subsequent points do not involve any new material or information, nor do they deal with *specific* issues raised in the article. Rather, they are broad indictments of the article stemming in part from an insufficiently careful reading. Both points concern the question of whether the two superpowers share some culpability for noncompliance. I did not state that both countries are equally culpable but rather that while the U.S. has adhered more rigidly to a strict interpretation of treaty obligations, it has not let a narrow interpretation of those obligations stand in the way of what were deemed to be defense needs. Nor did I suggest that violations were in any ethical sense acceptable, or that they would occur in an ideal world. I indicated only that behavior skirting the edges of the permissible is, given the depth of the U.S.-Soviet rivalry, probably inevitable, and that this does not necessarily negate the value of arms control. Neither side appears willing to subordinate perceived national-security imperatives to its treaty obligations, although neither seems willing to violate treaty provisions openly (except in the case of the Soviet radar at Krasnoyarsk).

Simple formulas provide easy, predictable and quite often psychologically comforting interpretations of the world but rarely an adequate guide to constructive policy. In the realm of arms control it is important to remember that the benefits of treaties are contingent not only on an absence of ambiguous compliance behavior but also on the security and the economic benefits that such agreements may provide.

MIROSLAV NINCIC

Department of Politics
New York University
New York, N.Y.

INCREDIBLE VACATIONS FOR INQUISITIVE MINDS.

Close encounters with sunken wrecks.

There's more of Moore in Ontario.

Say hello to a fossil.

Alexander Graham Bell lived and worked here.

Past, present and future mingle in Ontario.

Ontario is timeless. Search for 135 million year old fossils one day, take a round trip to Jupiter in 2019 A.D. the next.

Or, enjoy a superb meal in 1986. The ages mingle in Ontario. The sciences, the arts and history live side by side. Ontario is the vacation spot made for time travellers.

With the great exchange rate, your U.S. travel dollar goes a lot farther. Visit the ages in incredible Ontario, Canada.

For more information about Ontario vacations, call us TOLL FREE at 1-800-268-3735.

ONTARIO *Incredible!*

The personal computer that continues to raise high performance to new heights.

If you work with high volumes of information, you need answers fast.

Which is why IBM created the Personal Computer AT.® The PC designed to push high performance even higher.

The power of Advanced Technology.

Turn on the power, and you'll notice the advances right away.

To begin with, the Personal Computer AT can compute with astonishing speed. That's something you'll appreciate every time you recalculate a spreadsheet. Or search through a data base.

It can store mountains of information—up to 15,000 pages' worth—with a 30-megabyte "hard file" (fixed disk). And when business gets bigger, you can double your capacity to 60MB with a *second* hard file.

Advanced Technology advances again.

Fast as it is, new models of the Personal Computer AT run up to 33% faster.

Plus, you can now select an enhanced keyboard. It has separate cursor keys, a separate numeric keypad (for easier data entry) and twelve function keys. It also has a main typing section patterned after the classic IBM Selectric® keyboard, which makes word processing easier than ever.

In addition, you can also choose a new option that lets you use 3.5-inch diskettes with your Personal Computer AT.

Family ties.

The Personal Computer AT is compatible with the IBM PC and PC/XT. So it can run many of the thousands of programs written for the IBM PC family.

And with new IBM products, the Personal Computer AT can more easily communicate with other PCs.

IBM's networking options, for example, let you share files from a variety of popular programs.

While the IBM PC 3270 Emulation programs let you retrieve information from a mainframe.

You can also use the Personal Computer AT to run IBM PC XENIX® —an enhanced multi-user, multi-tasking operating system.

A powerful value.

Only the Personal computer AT offers these capabilities *and* IBM's commitment to quality, service and support. A combination that can't be cloned.

Better still, you'll find that the new models offer higher performance without a higher price. And if you qualify, you can conveniently charge your Personal Computer AT on IBM's Credit Card. Or lease one with the IBM Commercial Lease Agreement.

See the Personal Computer AT at an Authorized IBM PC Dealer. Or call your IBM marketing representative.

For the name of a store near you, call 1-800-447-4700. In Alaska, call 1-800-447-0890.

The IBM Personal Computer AT, for Advanced Technology.

IBM®

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

AUGUST, 1936: "One of the most significant developments of air transport is the increasing size of units and the definite certainty that still larger planes will shortly find their place in the air lanes, both over land and over sea. Having fledged their wings to a point which makes possible a high degree of reliability of operation coupled with speed at least three times that of surface transportation, passenger liners for the overland route and flying boats for the overseas services, which, before the year is out, will bring both major oceans into the scheduled airline map, are ready to expand in dimensions for increased comfort and increased economy."

"The day of miracles in aeronautical research is passing. In the future we must expect steady improvement and refinement without such radical departures as in previous years, at least until we come to rocket flight, or flight at 50,000-feet altitude."

"Science and mathematics professors in the colleges now regard the present crops of high school graduates as inferior in basic training to those students who entered college only 10 years ago. This is the unfortunate result of a recent trend in education wherein more stress is placed on the mechanics of teaching than on knowledge of the subject matter. These facts have been brought out by a committee of the American Chemical Society who charge that the 'professional teacher training' group among educators has secured a monopoly of the education machinery."

"It is fashionable in certain quarters to attack the modern streamlined car as unsafe. The commonest criticism is that these efficient cars are too highly powered and therefore tempt drivers to excessive speeds. Another criticism is that streamlining and lower passenger seating decrease visibility. It is also claimed that modern cars have poor roadability. High steering ratios are said to produce sluggish control. Low ground clearances are held to be a source of danger to the transmission and exhaust systems. Poor ventilation and exhaust gas fumes are stated to be the direct cause of frequent accidents. A careful statistical study made by the Travelers Insurance Company seems to refute these views, and to place greater blame on the driving public than on the design or condition of automobiles in service."

"On June 10th radio communication hailed the approach of new services by which businessmen will send one another entire letters by radio instead of terse 10-word telegrams and by which social notes will speed through space to be received and delivered in the exact handwriting of the senders. The occasion was the first demonstration of RCA's new ultra-short-wave radio circuit connecting New York and Philadelphia. The circuit makes possible the transmission of drawings, type matter and handwriting in facsimile."

SCIENTIFIC AMERICAN

AUGUST, 1886: "The carefully conducted experiments of M. Marcel Deprez, on the transmission of electricity over long distances, have finally resulted in success. After many trials and difficulties, the conductors established between Creil and Paris began to work satisfactorily. The power transmitted, and rendered available at the receiving station, was found by measure to be 50 horse power, an efficiency of 47 per cent. As the distance between Creil and Paris is almost 32 miles, this result is not unsatisfactory."

"The introduction of the pneumatic movement for organs was one of the great steps in the development of this instrument. By means of the movement the strain of opening the pipe valves was removed from the fingers of the performer. Now there is another improvement. It calls electricity into play, and the pneumatic movement is controlled by the electric current."

"During the last thirty days all the colleges, high schools and other advanced institutions of learning have held their commencements and thrust their graduates out upon the world. The number of these young persons probably reaches tens of thousands, but of them all, perhaps not two per cent have learned how to *do* anything. The education has been of the head alone, and not at all of the hand. As a consequence we import from Europe every year thousands of skilled workmen, while our own young people are driven into poorly paid clerkships or persuaded to attempt success in the overcrowded professions."

"The volume of the United States Census Reports of 1880, which has been issued recently, furnishes an interesting mass of facts bearing on the length of life. During the census year it appears that of a hundred deaths reported, forty were of persons under five years of age, fifty-two were of persons under twenty, and only twenty-two were of persons over fifty. Only about ten per cent survive their threescore years and ten."

"We have received a large number of communications discussing the problem of horizontal curve pitching in baseball. From an examination of the almost unanimous testimony thus submitted, and from an independent consideration of the problem on its own merits, it appears the horizontal curve is in the same direction that the ball is rotating and not in a contrary direction, as stated by our contributor, Mr. Chadwick, in the original article."

"Mr. Verbeck, who was deputed by the Dutch Indian Government to report on the origin and character of the volcanic outbreak in the Sunda Straits in August, 1883, has published his report. He calculates that the amount of ejected matter from Krakatoa must have been at least 10 cubic miles."

"The framing for the Statue of Liberty has been finished with the exception of two small parts—that supporting the right hand and that of the head. The shell of the statue has been carried up only a short way. It is extremely doubtful whether the statue can be finished by the 3d of next month, which is the date set for what we may term the unveiling."

Liberty's face from the inside

TOYOTA VANS

WIDE OPEN SPACES
Toyota's WonderWagon has enough room to pack in the troops, seven passengers, or even a load of antiques. The quick-release rear and middle seats are a snap to remove. It also packs in a 2.2 liter engine with plenty of get-up-and-go for your dashes from home, to school, to lessons. Yet it still packs in great mileage.* Maneuver yourself into Toyota's WonderWagon—you and the troops deserve its style.

PLENTY OF ROOM TO MANEUVER THE TROOPS.

GETS INTO TIGHT PLACES
Toyota WonderWagon's 15'1" turning radius makes it a cinch to handle in traffic and tight parking spots.

Get More From Life... Buckle Up!

*EPA estimated 21 City MPG/EPA estimated 23 Hwy MPG with 4-speed automatic overdrive transmission.

WHO COULD ASK TOYOTA FOR ANYTHING MORE!

The Most Important Natural

The 50th Anniversary Edition

Roger Tory Peterson FIELD GUIDES

Bound in rugged, genuine leather and accented with 22kt gold.

Each Guide is a convenient 4¾" x 7½" size. The rugged leather covers are flexible — easy to handle and use.

In the 1930's, Roger Tory Peterson saw a dream come true ... with a single book he turned bird watching and the appreciation of nature into one of the most popular pastimes in America.

That book? **A Field Guide to the Birds.**

Since then, many more books have followed ... a breathtaking array of Roger Tory Peterson Field Guides — written by Peterson and leading experts — covering the full spectrum of nature: **Birds, Butterflies, Wildflowers, Sea Shells, Stars and Planets, Weather and Atmosphere, Trees and Shrubs, Coral, Rocks and Minerals, Freshwater Fish** — some 35 titles in all. Together, they unlock the whole world of nature — enabling you to see, know and identify all of nature's wonders immediately by sight and by name.

A 50th anniversary "lifetime" edition.

Now, for the first time ever, and in honor of their 50th anniversary, this complete collection of Roger Tory Peterson Field Guides — will be bound in rugged, beautiful, genuine leather, housed in a hardwood display case and contain a volume autographed by Dr. Peterson.

The Easton Press has created an edition of superb quality. Each Guide is bound in beautiful, ruggedly supple leather. Each weather resistant cover is deeply embossed with a golden design. A ribbon marker is permanently affixed into each binding. And the papers used are groundwood-free, made to capture the full-color fidelity of illustrations. Each is thread-sewn to last for generations.

An autographed edition.

Roger Tory Peterson, a man who has earned numerous awards, including The Presidential Medal of Freedom, is generally acclaimed as the best known bird artist since John James Audubon. Throughout the collection, key illustrations chosen by Peterson from his archives will be newly added to the Guides. You will also receive the classic **A Field Guide to the Birds** signed personally by Dr. Peterson.

Attractive hardwood display case free of additional charge.

This collection will create an exciting presence in your home. And your collection will have a proper setting! At no additional charge, you will receive an impressive hardwood display case, allowing you to show your books to their best advantage, while lending warmth and dignity to your home.

Available exclusively from The Easton Press.

The authorized 50th anniversary leather-bound edition will be available exclusively from The Easton Press. None will be sold in book stores.

You need send no money now. Simply complete the Preferred Subscription Reservation and mail it today.

History Library Ever Created.

Prompt action required.
You are urged to mail your reservation promptly. This special edition may not be offered again.

This hardwood display case will be provided at no additional charge.

Sponsored By The National Audubon Society And The National Wildlife Federation

Preferred Subscription Reservation 354

ROGER TORY PETERSON FIELD GUIDES

The Easton Press
47 Richards Avenue
Norwalk, Conn. 06857

No Payment Required.
We will bill you.

Please send me the first volume of the Roger Tory Peterson Field Guides bound in genuine leather and reserve a subscription in my name. Further editions will be sent at the rate of one book per month at $29.75* each. This price will be guaranteed for the next two full years. I understand that I may return my book within 30 days for a refund, and that either party may cancel this subscription agreement at any time.

*Plus $2.50 per book for shipping and handling.

☐ As a convenience, I wish to pay for my first book now. I enclose $29.75 (plus $2.50 shipping and handling — a total of $32.25**).
☐ Please charge each monthly shipment to my credit card:
☐ VISA ☐ MasterCard

Credit Card No. _____ Expiration Date _____

Name _____
Address _____
City _____
State _____ Zip _____
Signature _____

**ude sales tax. Please allow 8–12 weeks for shipment.

THE AUTHORS

JOHN PRADOS, JOEL S. WIT and MICHAEL J. ZAGUREK, JR. ("The Strategic Nuclear Forces of Britain and France"), share an interest in arms-control and defense policy. Prados, who holds a Ph.D. in political science from Columbia University, has authored several books, including *The Soviet Estimate: U.S. Intelligence Analysis and Russian Military Strength* (Princeton University Press, 1986) and *The President's Secret Wars* (William Morrow and Company, forthcoming). He has designed a number of commercial war-simulation games. Wit, an analyst in the Bureau of Intelligence and Research of the U.S. Department of State, studied at Columbia, receiving a master's degree in 1979 from the School of Public and International Affairs. He has held several positions dealing with national-security issues, most recently as a consultant for the Congressional Research Service of the Library of Congress. Zagurek, who works for the Digital Equipment Corporation, got his master's degree from Columbia's School of Public and International Affairs in 1979. He has spent a number of years designing and programming computer simulations of nuclear war. The views in the article are the authors' own and are not meant to represent those of their employers.

DAVID R. NELSON ("Quasicrystals") is professor of physics at Harvard University. He studied at Cornell University, where he earned an A.B. (1972), an M.S. (1974) and a Ph.D. (1975). After completing his doctorate he worked as a research associate at Cornell and then moved to Harvard as a junior fellow. In 1978 he joined the faculty at Harvard, and he was made full professor in 1980. He was awarded a five-year MacArthur fellowship in 1984. Nelson's research interests include phase transitions, turbulence, liquid crystals and the statistical mechanics of glass.

ALAN FINE ("Transplantation in the Central Nervous System") works in the Molecular Neurobiology Unit of the Medical Research Council in Cambridge, England. His undergraduate studies were done at Harvard University and were punctuated by years spent in France and Israel. He went on to pursue graduate studies in physiology and veterinary medicine at the University of Pennsylvania. After graduate school he held postdoctoral fellowships at both the National Institute of Mental Health and the Weizmann Institute of Science in Israel. Fine has been in Cambridge since 1983.

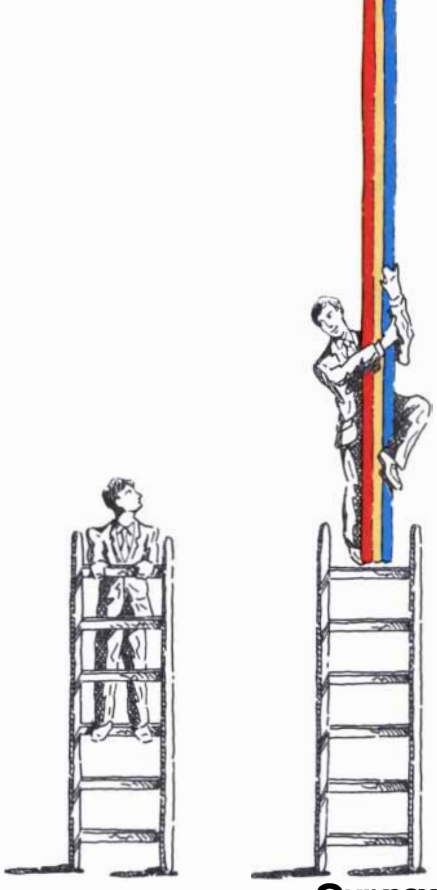
R. STEPHEN CRAXTON, ROBERT L. McCRORY and JOHN M. SOURES ("Progress in Laser Fusion") work at the Laboratory for Laser Energetics of the University of Rochester. Craxton is senior scientist in the laboratory's theoretical division. He got his B.A. (1970) in mathematics from the University of Cambridge and his Ph.D. (1976) in physics from the Imperial College of Science and Technology in London. After finishing his graduate education Craxton moved to the University of Rochester, where his main research interest is the computer simulation of laser-produced plasmas. McCrory, who holds a B.S. and a Ph.D. from the Massachusetts Institute of Technology, is professor of mechanical engineering at Rochester and director of the Laboratory for Laser Energetics. He was staff physicist at the Los Alamos National Laboratory before going to Rochester. McCrory has served as vice-chairman and trustee of the Consortium for Scientific Computing, Inc., in Princeton, and he is a consultant to several private and governmental organizations. Soures is deputy director of the Laboratory for Laser Energetics and director of its experimental implementation division. He studied at the University of Rochester, receiving three degrees: a B.S. (1965) in physics and an M.S. (1967) and Ph.D. (1970) in mechanical and aerospace sciences. After getting his doctorate Soures joined the laboratory, where he has been designing and developing high-power lasers.

PAUL C. MANGELSDORF ("The Origin of Corn"), Fisher Professor Emeritus of Natural History at Harvard University, is a leading authority on corn genetics, having devoted most of his career to the study of corn. The son of a Kansas florist and seedsman, Mangelsdorf developed an early interest in plant cultivation and studied plant breeding at Kansas State College and at Harvard, where he got an Sc.D. in 1925. After 13 years of breeding new varieties of corn, wheat, oats and barley at the Texas Agricultural Experiment Station, he returned to Harvard to teach in 1940, and he later became the director of its botanical museum. For many years he also served as an adviser to agricultural programs sponsored by the Rockefeller Foundation in Mexico and South America. In 1968 he retired to Chapel Hill, accepting an appointment as a lecturer at the University of North Carolina. His chief interests now, aside from corn, are his grandchildren (who have provided material "for a fascinating study in human genetics"), bird watching and philanthropy.

GEORGE E. WILLIAMS ("The Solar Cycle in Precambrian Time") is principal research geologist in the minerals-exploration department of the Broken Hill Proprietary Company Ltd. in Adelaide, Australia; he is also an adjunct professor of planetary sciences at the University of Arizona. After receiving an M.Sc. in geology from the University of Melbourne, he went to England on a scholarship and got a Ph.D. in sedimentology from the University of Reading in 1966. He has worked as a prospector for minerals in many countries besides Australia, including Britain, Algeria, Turkey, Canada and the U.S. At the same time Williams has pursued his other main research interest, which is the subject of his article: reconstructing solar and geophysical history through the study of ancient sedimentary rocks.

CHARLES C. COUTANT ("Thermal Niches of Striped Bass") is manager of the U.S. Department of Energy's global carbon-cycle program at the Oak Ridge National Laboratory. He was both an undergraduate and a graduate student at Lehigh University, where he received his Ph.D. in 1965. After working as a research scientist at the Pacific Northwest division of the Battelle Memorial Institute, he moved to Oak Ridge, where he has held various positions in the environmental-sciences division. For most of his career Coutant has studied the impact of human activity on the natural environment, and in particular the effects of dams and power plants on the ecology of rivers and streams.

NORMAN HAMMOND ("The Emergence of Maya Civilization") is professor of archaeology, anthropology and classics at Rutgers University; he is also director of the university's archaeological research program. After getting his Ph.D. in 1972 from the University of Cambridge, he stayed on as fellow at the Centre of Latin American Studies. He then taught at the University of Bradford in England and at the University of California at Berkeley before joining the faculty of Rutgers in 1977. He has directed investigations of Maya sites in Belize since 1970. In addition to the Maya, he is studying early settlement and trans-Andean communication in southern Ecuador.



Why wait for the digital future when New York Telephone can connect you to it now?

Our newest services can start moving your business towards the Integrated Services Digital Network.

The future of business communications lies in an international, end-to-end digital communications network called ISDN. The Integrated Services Digital Network.

This network, now in development, will dramatically improve the quality and efficiency of daily business communications. Soon ISDN will carry video, voice, and data worldwide with speed, accuracy, and economy never before seen.

And you can benefit from ISDN technology right now. Because New York Telephone has designed a family of digital communications services called NYNEX Pathways℠ services as the first steps towards ISDN.

Six Pathways services are already available. And as new ISDN technology develops, New York Telephone will update and expand the network.

Through Pathways services, you can take the first steps towards ISDN and the digital future now. For more information, call one of the sales agents below or New York Telephone at **1 800 942-1212, ext. 25.**

New York works on New York Telephone.

New York Telephone

A NYNEX Company

AIM Telephones Inc. 1 212 684-9000
ALL-Mode Communications, Inc. 1 607 347-4164
DIAL Communications, Inc. 1 212 868-6677
ETCO Communications Inc. 1 212 242-9135
Executone Long Island 1 516 582-3200
Gaffney Communications, Inc. 1 315 797-8222
IPC Communications, Inc. 1 212 825-9060
NYNEX Business Information Systems 1 800 346-9X9X
Ronco Communications & Electronics, Inc. 1 716 879-8100
Standard Telecommunications, Inc. 1 212 532-5440
Tel Plus, Inc. 1 800 TEL-PLUS

Introducing NonStop VLX.™

Tandem technology sets the new standard for large applications in on-line transaction processing.

More transactions per second at a lower cost per transaction than any system in the world.

THE CIRCUITRY'S FAST.

We designed the system in our own laboratory, right down to our own unique VLSI chips. The result is more circuitry in less space. With fewer components than our next largest system, the VLX delivers twice the performance and three times the reliability.

PROCESSORS WITH LARGE APPETITES.

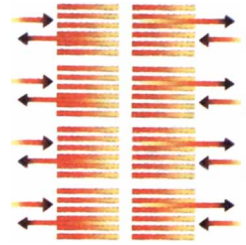
The VLX processors move transactions in 32-bit chunks. They reach into main memory in 64-bit chunks. Because this happens in parallel, more work gets done in less time at a lower cost per transaction.

THE SERVICE IS EASY.

All critical components are field replaceable. When service is required, it's faster. You don't even have to stop an operation to add or replace components.

THE DATA EXPRESSWAY.

In a conventional database, I/O requests must be handled sequentially. This creates queues that slow response time. In the VLX system, there are multiple paths to multiple disks. Data enters and leaves the database simultaneously. No time is wasted, and all disk space gets used.



DIAGNOSTICS FROM A DISTANCE.

An integrated microprocessor allows us to monitor the system environment from anywhere in the world. We can even run stress tests remotely. If a failure does occur, the VLX has the capability to automatically dial out to remote centers anywhere in our worldwide network.

THE SYSTEM KNOWS THE SYMPTOMS.

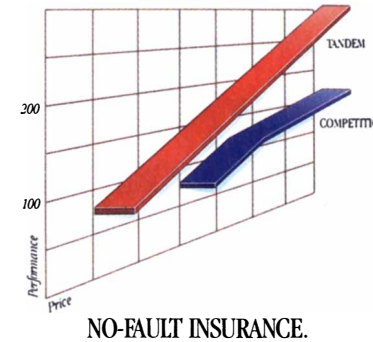
Expert systems software, using fault analysis, directs the problem diagnosis systematically. It also allows us to analyze it and shorten service time even more.

SECRETS ARE SAFE.

We offer software that will protect the security of your data whether it's in the VLX, in another Tandem system or in transmission.

NO GROWING PAINS.

To add power, just add processors. You can grow from a base four-processor system to 16. From there, you can expand in whatever increments you choose, all the way to 255 systems. You never buy more than you need, and you'll never have to rewrite a line of applications code.



NO-FAULT INSURANCE.

Tandem systems achieve fault-tolerance with a unique, parallel processing architecture. There are no idle back-up components. Instead, multiple components share the workload. If one goes down, the others pick up the slack, and application processing is uninterrupted.

HERE TODAY. HERE TOMORROW.

The VLX is compatible with any Tandem system and with all major communications standards — SNA, X.25, MAP and O.S.I. And by acting as a gateway to other vendors' systems, the VLX can link them and enhance their value as well.

WE HAVE EXCELLENT REFERENCES.

Tandem systems are already at work for Fortune 500 companies in banking, telecommunications, manufacturing, transportation, retailing and energy, as well as several branches of the U.S. Government.

To find out what we can do for you, call 800-482-6336 or write to us. Corporate Headquarters: Tandem Computers, Incorporated, 19191 Vallco Parkway, Loc. 4-31, Cupertino, CA 95014

TANDEMCOMPUTERS

COMPUTER RECREATIONS

Digital prestidigitation: the fine art of magic and illusion by computer

by A. K. Dewdney

Put the words "magic" and "computer" in the same sentence and you are likely to throw a rational person into a state of cognitive dissonance. How could magic, the essence of a shadowy world we chain off with links of logic, invade the computer, the primary expression of our rationality? The dissonance lies in the word "magic" as a symbol of the supernatural. In reality magic is a piece of the natural world, and magicians are logical folk.

To bring off a feat of magic three things are sufficient, if not always necessary: audience, props and presentation. I have seen magic done without props but never without an audience; any audience will do, but I must leave the hornswoggling to the aspiring magician. A computer, however, makes a very clever prop, and I shall assume here that the magician can materialize at least a microcomputer. Simple programming enables the machine to guess a card drawn at random, to engage in numerical precognition and to determine the identity of a card by telepathy, all to the consternation of a hapless volunteer. In fact, even a hand calculator is enough to put you into the business of mind reading.

So much for audience and props. Presentation is everything else. It converts the inner logic of a trick into an outer appearance. Real magic holds the stage only for as long as the magician can sustain and manipulate certain beliefs in the minds of the audience. Each belief seems to lead inexorably to the next one, until there is an astonishing denouement. Sometimes it helps to weave a spell before beginning a trick, and there is plenty of mumbo jumbo for that purpose available from computer science. You are advised to use it liberally with the uninitiate and judiciously with the knowledgeable. Here is how the first feat of magic might be introduced:

"Ladies and gentlemen, sooner or later it was bound to happen. A breakthrough in artificial intelligence has made it possible for computers to read the human mind. The theory of recursive nondeterminism guarantees that a certain nonpolynomial algorithm, first described by the famous Tibetan computer scientist Professor Yan Kee, will always emerge from infinite loops. Step by step the program cycles through a semi-infinite stochastic data structure, retrieving logical primitives and compiling them into data-base queries.... [One may ramble in this way for some time.] May I have a volunteer from the audience, please."

The unfortunate volunteer is introduced and then asked to pick a card from a deck at random. Any card will do. The magician holds up the card for all to see. Suppose it is the six of clubs. This is a good time to make an even bigger fool of the volunteer. The magician says: "Please press the card against your forehead and concentrate hard. Your mind and the programmed mind of the computer must reach a state of simulated simultaneity."

The magician then goes to the computer keyboard and starts the program. Immediately a message appears on the screen. The magician reads the following dialogue aloud for the benefit of the audience:

"IS THE SUBJECT CONCENTRATING?" (The computer speaks in capital letters.) "Yes." (The magician types the answer.)

"IS THE CARD BLACK?" "Yes."

"IS IT CLUBS?" "Yes."

"IS IT AN 8 OR LESS?" "Yes."

"IS IT A 5 OR MORE?" "Yes."

"IS IT A 5 OR 6?" "Yes."

"IS IT A 6?" "Yes."

Time after time the program zeros in on the selected card by asking seven questions. Amazingly, the answer to every question is yes, as though the program merely wanted confirmation. Readers might well wonder how the trick is accomplished. Ordinarily I would hesitate to satisfy that curiosity: there is an unwritten code among magicians that forbids revealing sleights of hand. The magician's livelihood depends on an air of impenetrable mystery, and so I cannot disclose the basis of all the tricks presented here. For this clever card trick, though, I have permission to go public from my source Christopher Morgan, a former practicing magician. Morgan is now an executive at the Lotus Development Corporation in Cambridge; presumably his livelihood is no longer based on magic.

When the volunteer draws the card from the deck, the magician notes it carefully, then strides with a flourish to the keyboard and types "yes" in answer to each question. Each "yes," however, may be accompanied by an unobtrusive stroke of the space bar. The presence or absence of the space bar is not really the answer to the preceding question; instead it determines the content of the next one. Thus when Professor Yan Kee's fabulous program asks, "IS THE SUBJECT CONCENTRATING?" the magician types y e s space bar only if the selected card is black. If the card is red, the magician types y e s only. The program then knows enough to ask the right question at the next stage. If the magician typed a space, the program asks, "IS THE CARD BLACK?" Otherwise it asks, "IS THE CARD RED?"

In this way the magician guides the program down the branches of an implicit binary tree; the hidden logic encodes all the choices needed to identify any card in the deck [*see illustration on next page*]. For example, after the color of the selected card is determined, the next choice governs its suit. Once the computer "learns" the card is black, the magician can secretly tell it whether the card is clubs or spades. If the card is red, the next step is to distinguish hearts from diamonds. All the remaining choices govern the value of the card: from ace, 2 and 3 up to jack, queen and king. Since there are 13 possible values, four questions must be asked to distinguish among them.

It is essential that the magician misdirect the audience each time the space bar is touched. The simplest misdirection is based on the carriage-return key (called variously "return" or "enter"). Since it must be pressed each time the magician's answer is entered, a great flourish can be made of pressing it with the right hand while the left hand quietly trips the space bar. Practice makes perfect.

The program I call YAN KEE is longish but quite easy to write. It is divided into 31 small sections, corresponding to the total number of distinct questions it must be able to ask. All the sec-

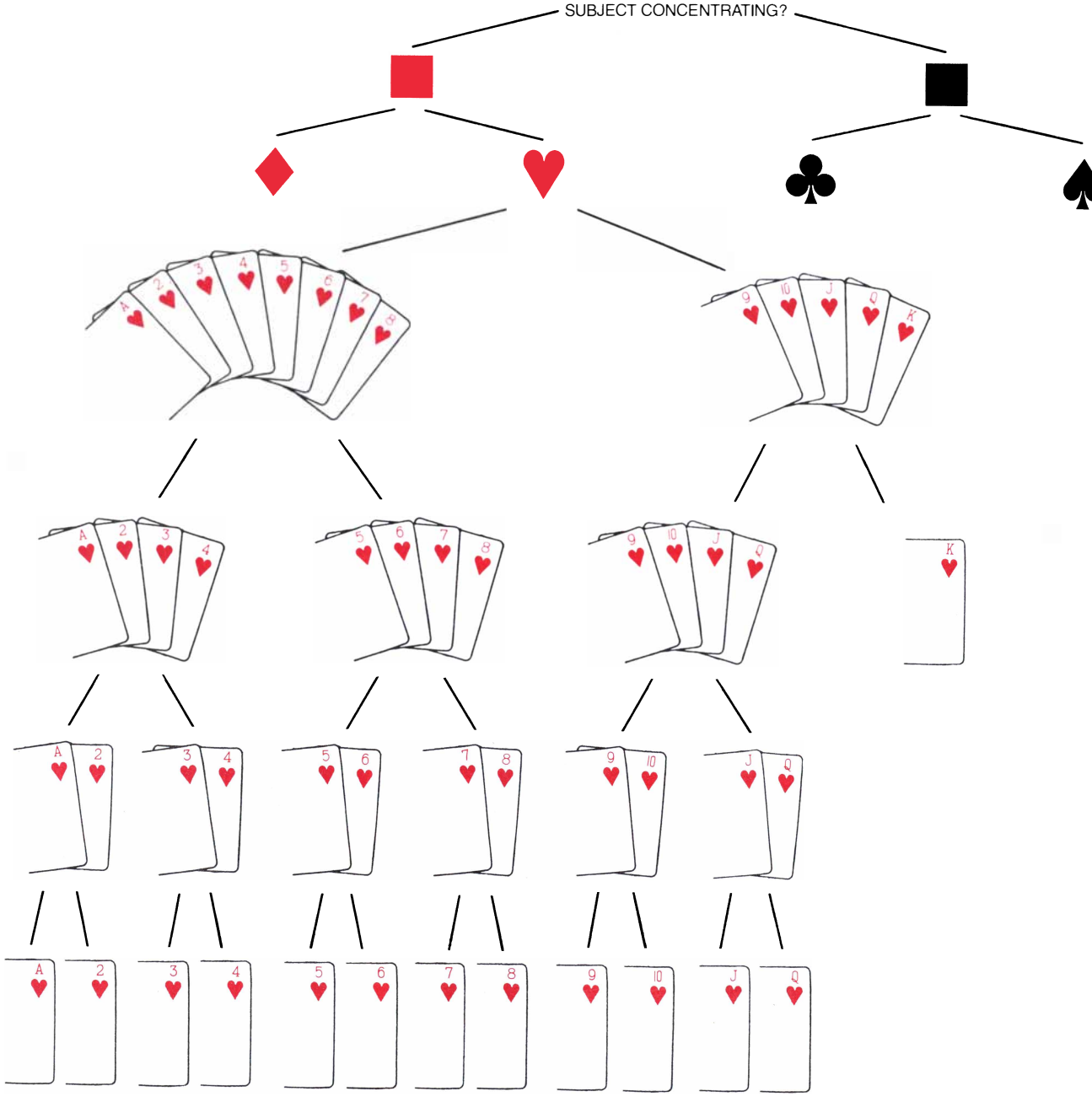
tions have the same basic algorithmic structure:

```
XX output "IS THE CARD BLACK?"
   input characters
   compute number of characters
   if number of characters = 4
      then go to YY
      else go to ZZ
```

Here XX stands for a line number or label to which the execution of the program has passed after the first input y e s space bar in my example. After printing the question appropriate in this section, the program accepts the magician's input characters and simply counts them. If four characters were typed, the magician must have hit the space bar. In this case execution passes to the program section labeled YY, which begins with the question "IS IT CLUBS?" If, on the other hand, the magician did not press the space bar, the program branches to the section labeled ZZ, which begins with the question "IS IT SPADES?" All would-be magicians must memorize the decision process inherent in the space-bar option at each level of questioning.

The next feat of programmed legerdemain was suggested by Michael Rohregger, a reader in Linz, Austria. The computer magician, having appealed to the same questionable theories outlined earlier, now advances the claim that a second program is able to divine a person's thoughts even before they occur to the thinker! Unfortunately, the magician goes on, development in this area is not yet complete, and so the program succeeds only most of the time. A volunteer is summoned from the audience, seated in a chair and asked to choose between the two bits, or numbers, 0 and 1. To keep the volunteer from being distracted by the computer monitor it may be best if both the volunteer and the monitor face the audience.

The volunteer calls out a number



The secret decision tree for Christopher Morgan's computer card trick, with selected continuations

Dr. Donald P. Ames

One Fellow who is making the world we'll live in tomorrow.

Dr. Donald P. Ames is an exceptional achiever in technology—as such he has been named *McDonnell Douglas Distinguished Fellow.* Dr. Ames is the first to earn this honor, the highest the company bestows upon its scientists. Dr. Ames is general manager of the McDonnell Douglas Research Laboratories, a position he has held for the past 15 years.

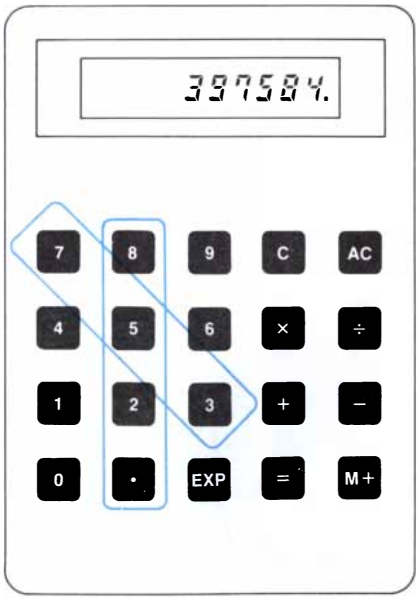
A man of many talents.
Dr. Ames has marked his passage across a spectrum of scientific fields, including nuclear fission, instrumentation, low-temperature physics, materials, lasers and particle beams. His research in these fields has earned him numerous honors, and has led to assignments on national scientific committees.

McDonnell Douglas searches out and develops men and women who are creative leaders in their work. It surrounds them with challenges to be met and provides the facilities to fulfill their potential. It offers opportunities for advancement and recognizes their accomplishments.

As the Distinguished Fellow, Dr. Ames joins 25 previously appointed Fellows who bring progress and honor to the company. These remarkable individuals represent in the world of technology what the Albin Polasek sculpture shown here represents in art: Mankind carving its own destiny.

"Man Carving Out His Destiny"

MCDONNELL DOUGLAS



Magic product of two "colinear" numbers

and the magician and the audience compare it with a digit already displayed on the monitor. If the digits are the same, the magician nods sagely. If they are different, the magician explains that it may take some time for the program to tune in to the volunteer's mental processes. (Indeed, the volunteer can always be blamed for causing "statistical static.") The magician presses a key and the program predicts the next bit to be chosen by the volunteer. The volunteer is asked to choose again. The experiment is repeated as many times as it takes for the program "to synchronize or nearly synchronize with the volunteer's space-time thought framework." Every time the volunteer calls out a number the magician nods and smiles if the machine is right and frowns patiently if there is a failure.

At first the program appears to score no better than chance: it is right only about 50 percent of the time. As the experiment is repeated, however, there is a marked improvement: the program tends to score at least 60 percent, and in some cases it may do considerably better. Here is how it works.

The program, called PREDICTABIT, maintains a 16-by-2 array called *subject*. Each of the 16 rows is indexed by one of 16 possible sequences of four consecutive bits. The columns of the array represent the two possible bit choices, 0 and 1. The first entry in a row is the number of times the volunteer has chosen a 0 immediately after choosing the sequence of bits corresponding to the row index. The second entry is the number of times the volunteer has followed the sequence by a 1.

For example, suppose the volunteer has chosen the bits 0, 1, 1, 1 in the preceding four trials. The program treats the bits as a single binary number and converts them into its decimal equivalent, in this case 7. If *subject*(7,0) is equal to 4 and *subject*(7,1) is equal to 2, the volunteer has followed the sequence 0, 1, 1, 1 by a 0 four times in the past but only twice by another 1. Given such past performance, the program predicts the volunteer's next choice will be a 0.

The program implicit in this description is probably already fairly clear. In its simplest form PREDICTABIT cycles through six simple sections:

Input latest bit
Update array at old sequence index
Shift variables
Form new sequence index
Compare *subject* entries
Make prediction

As soon as the volunteer calls out a binary digit, the magician types it into the computer. Four variables called *first, second, third* and *fourth* hold the volunteer's preceding four choices. A variable called *index* stores the row number computed on the basis of the current values of the variables *first* through *fourth*. If the latest bit called out is 0, the program adds 1 to *subject*(*index*,0). Otherwise it adds 1 to *subject*(*index*,1).

The third section of the program then shifts the contents of the four variables as follows:

fourth ← latest bit (input)
third ← *fourth*
second ← *third*
first ← *second*

PREDICTABIT forms the new sequence index number by computing a sum of products: $1 \times first + 2 \times second + 4 \times third + 8 \times fourth$. Next the program compares *subject*(*index*,0) with *subject*(*index*,1) for the newly computed value of *index*. If *subject*(*index*,0) exceeds *subject*(*index*,1), PREDICTABIT prints a 0 on the screen as its next prediction. Otherwise it prints a 1. To operate successfully the program must also predict the volunteer's first four guesses, just to fill in dummy values of the first four variables. The dummy values might as well all be 0's.

For the magician who wants to announce PREDICTABIT's rate of success there is one more feature worth including in the program. Declare a new variable called *score*, and just before the shift section insert additional instructions that compare *fourth* with the latest bit called out by the volunteer. If they are equal, add 1 to *score*. When the demonstration is finished, the program divides *score* by the number of trials, multiplies by 100 and displays the result.

The trick works because people tend to adopt rules unconsciously when they try picking one of two alternatives at random. For example, after choosing two 0's in a row, it is tempting to make the next digit a 1. In fighting this tendency many people may add a third 0, but then the temptation to make the next digit a 1 becomes even greater. The probability that the next digit will be a 1 increases with each 0 added to the sequence. In a truly random selection, of course, the previous choices have no bearing at all on the choice of each succeeding digit: the probability that the next digit will be 1 is always the same, namely .5.

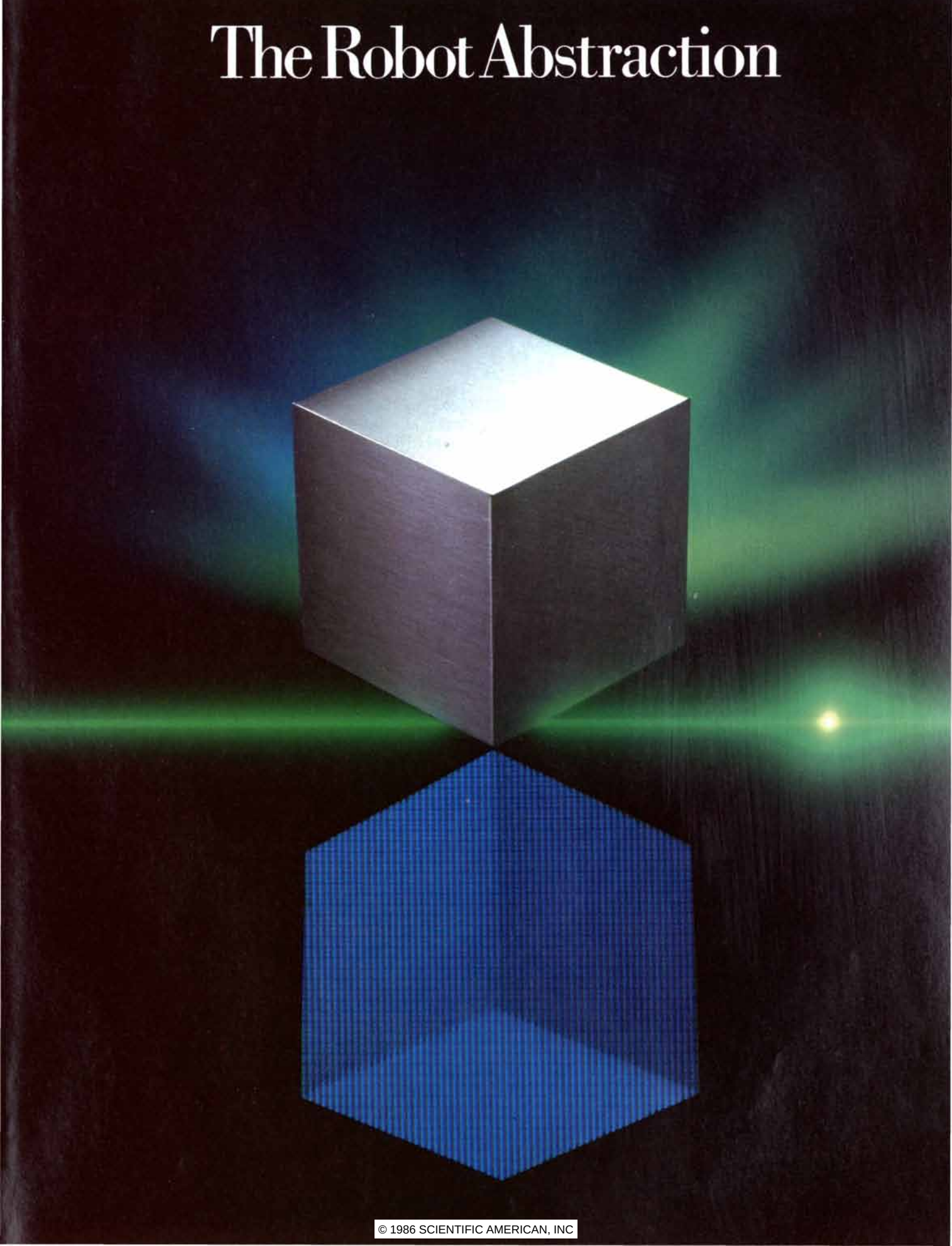
Perhaps even subtler factors operate when people try to choose a random bit. Again, however, if such factors somehow depend on recent history, PREDICTABIT should be successful. I have found, incidentally, that it is crucial to avoid subjects who think too creatively. Such people tend to adopt rules consciously and then change them in the middle of the demonstration, with disastrous results. To keep a volunteer from becoming too unpredictable it might be wise to add some misdirecting patter, such as: "When I press this key, the program will predict the number you will think of next. It is crucial that you *not* think of your number yet—otherwise the machine will have the unfair advantage of being able to read your mind directly." Thereafter the magician must continually caution the volunteer to wait until the key is struck. Remember that the key struck is the volunteer's guess on the preceding round.

In my next feat of magic the computer is both prop and magician. The (human) magician hands the volunteer a deck of ordinary playing cards and intones: "Please have a seat before the computer incarnation of a great magician, now departed this life." To further bamboozle the volunteer, one can make glib analogies between spirit and software. If necessary, one can go on with the theories of Professor Yan Kee. Finally the magician asks the volunteer to press the space bar. The following instructions emerge from the Beyond. When one instruction has been carried out, a tap on the space bar brings up the next:

SHUFFLE THE CARDS THOROUGHLY, SQUARE THEM UP AND PLACE THEM FACEDOWN ON THE TABLE [space bar].

CUT THE DECK ROUGHLY IN THE MIDDLE, NOTE THE CARD EXPOSED AND RESTORE THE DECK [space bar].

The Robot Abstraction



The Robot Abstraction

Until now, matching robots to specific industrial tasks has been done by trial and error, requiring the creation of expensive prototypes. Recent advances at the General Motors Research Laboratories have produced a computer system that can be used not only to select the right robot, but also to program it to perform the task in the most efficient way.

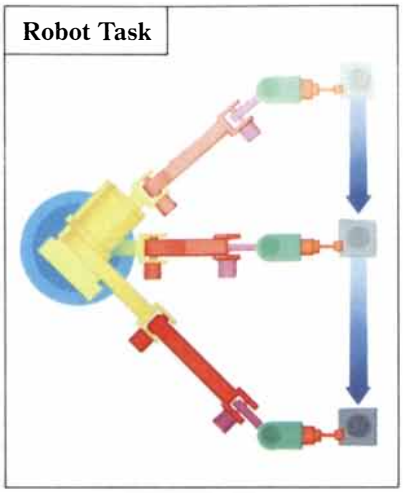


Figure 1: Two-dimensional overhead view of a robot task—the straight path trajectory of a solid.

Figure 2: Three-dimensional illustration of the robot work cell layout, showing reach capability for the task in Figure 1. Areas of color show total reach as well as the joint limits stored in the robot model.

THE DECISION to use robots to automate a manufacturing facility introduces the need for more decisions. There are several dozen kinds of robots, each with different capabilities. Thus far, choosing the right robot for a given set of tasks has been largely a manual process, involving great expenditures of time and money. By combining previously separate disciplines in a single computer system, two General Motors researchers have made the introduction of robots to the factory floor a more rational, less costly undertaking.

RoboTeach is the first computer system which integrates robotics, solid modeling, and simulation. It was designed and developed by Dr. Robert Tilove and Mary Pickett, both members of the Computer Science Department.

The use of powerful programming languages for manipulating robots is a major new development in the discipline of robotics. The languages specify desired robot motions, but they have no way of describing the robot's environment. Hence, they cannot automatically take into account physical obstacles or anticipate collisions. With only robot programming languages at one's disposal, assuring proper interaction with the environment requires testing with actual robots and parts.

Solid modeling, on the other hand, provides geometrically complete representations of environmental components and their spatial relations. But solid modeling cannot represent processes, because it has no way of representing temporal relations. Traditional solid modeling deals only with static relationships. While robot programming is without physical context, solid modeling is nothing but physical context. Neither by itself is adequate.

Nor are they satisfactory together. Only by simulation of both the robot and its environment can the sequence of discrete steps in a robot task be converted into the continuous motion of a process. Also without simulation, there is no way to represent accurately the robotic process as it unfolds in its environment.

RoboTeach, by combining all three disciplines, provides computer representations of the environment, the robot, and the task. Consequently, it helps users reach high-

level decisions about the real world without the investment of time or money in actual robots, actual parts, or the factory setting.

One key RoboTeach abstraction is a mathematical robot model. Solid modeling techniques represent the geometric form of each link of the robot. Then constraints are imposed on the relative positions of mating links to produce a mathematical abstraction of a mechanical joint. By insisting that the joint constraints always be satisfied, RoboTeach insures that the abstract robot model corresponds to a physically realizable geometric configuration.

OTHER representational facilities in RoboTeach handle robot task definitions. The representation of any task can be matched with the representation of any robot. In this way, RoboTeach helps users to determine the optimal robot for the task. Once a robot has been selected, RoboTeach can be used to program the robot off-line.

Not only are robots proliferating, but the tasks assigned to them are becoming more complex, making the need for off-line programming more urgent. When there are only a half dozen robots in a factory, the prospect of reprogramming them all by conventional show-and-teach methods for every new task is not overwhelming. But when there are hundreds of robots, the value of being able to reprogram without interaction with each robot becomes more apparent. Without off-line programming, the savings which justified the initial robot investment may quickly vanish.

RoboTeach distinguishes between two kinds of off-line programming: at the task level (what to do) and at the robot level (how to do it). For example, in the creation of a mechanical assembly, task-level instructions would include what components to assemble, the alignment of the components for the assembly process, and criteria for verifying that the final assembly is correct. Typically, there is a one-to-many relationship between task-level instructions and robot-level instructions.

"RoboTeach is currently in use," says Robert Tilove, "to study robot reach capabilities and to simulate simple robot-level tasks."

"Future research," adds Mary Pickett, "will explore the possibility of using RoboTeach to approach problems from the more abstract task level, with the user defining the task at a high level and RoboTeach filling in the details."

THE PEOPLE BEHIND THE WORK

Dr. Robert Tilove and Mary Pickett are Staff Research Scientists in the Computer Science Department at the General Motors Research Laboratories.

Mary Pickett received her B.S. in mathematics from Iowa State University and her Master's in computer science from Purdue University. She was a member of the team that developed GMSOLID, an interactive geometric modeling system. Her research at GM has also included the design of real-time programming languages. She joined GM in 1971.

Robert Tilove received his undergraduate and graduate degrees in electrical engineering from the University of Rochester. His Ph.D. thesis concerned the design and analysis of geometric algorithms for solid modeling. His current research interests also include the application of geometric modeling to computer vision and robot control. He joined GM in 1981.

DEAL THE CARDS FACEDOWN INTO SIX SMALLER HANDS [space bar].

TAKE A MOMENT TO LOOK THROUGH EACH PILE UNTIL YOU FIND THE ONE CONTAINING THE CUT CARD [space bar].

NOW DEAL ONE CARD AT A TIME FROM THE PILE CONTAINING THE CUT CARD; DEAL EACH CARD FACEUP, CONCENTRATE ON IT AND PRESS THE SPACE BAR WHEN YOU ARE READY FOR ME TO DECIDE WHETHER IT WAS THE CARD YOU PICKED [space bar].

As the volunteer deals the cards one at a time, the program echoes the decision of the departed magician. "NO" is the first response. A tap on the space bar tells the magician's spirit that another card has been dealt. Again "NO." After a few more cards the spirit suddenly cries out (so to speak): "STOP, HERE IT IS!"

The third program in this series is the simplest. I call it HOUDINI for reasons I hope will be obvious. HOUDINI simply displays the instructions above, one set at a time, until the volunteer starts dealing the cards from the pile that includes the cut card. The program says "STOP" when the fourth card is dealt. That card or the next one is always the cut card.

HOUDINI is so simple it needs little or no description. When it runs, however, the temporarily unemployed human magician might hover near the volunteer, sharing in the growing amazement and generally making sure things go right. The cut, for example, must be made within a few cards of the center of the deck. The order of the cards in the six piles must also be maintained while they are searched. Finally, the human magician must interpret the STOP command. If the fourth card is the one sought, the silicon spirit of the machine has read the volunteer's mind. If the fifth card is magical, however, the magician advises the volunteer to turn it over. "Leaping LISP," says the magician, "I might have known. The great Houdini has read the cards by direct clairvoyance!"

True to my promise not to reveal the basis of all the magic presented here, I leave an explanation of this last trick to the reader. The answer will be revealed next month, and so please do not send me your thoughts. They will arrive much too late. I am indebted for the trick to Harry Lorayne, a New York magician and memory expert. HOUDINI is a computer adaptation of Lorayne's well-known "Stop!" trick.

The last bit of mind reading comes by way of Martin Gardner, and it requires only the humble hand calculator. The magician calls a volunteer

from the audience. After giving the calculator to the volunteer, the magician turns away, touches his brow and vamps on about some kind of electronic communion with the calculator's circuitry. At last he explains the following procedure to the volunteer:

"Choose any row, column or main diagonal of the numeric key pad. This is the little square of numbers from 1 to 9. Enter the digits in that line in any order you like. O.K.? You should have a three-digit number in the display. Now press the multiplication key and select another line—that is, another row, column or main diagonal of the little square. Got one? Good. Again, enter those three numbers in any order you like. Now you should have another three-digit number in the display. When you press the equal sign, the product of the two three-digit numbers will appear in the display. Do you see it? That is a magic number. Choose any nonzero digit from the number, but do not tell me what it is. Tell me the other numbers instead. No, the order is not important; I just want to clear them out of your mind so that you can concentrate on the number you did not tell me. Very good. Now concentrate on that all-important number and with a little luck I should be able to pick it up."

Readers also have a month to discover how the magician determines the missing number. Look closely at the illustration on page 20. In a future column I shall be happy to describe the adventures and misadventures of readers who attempt the magic programs presented here.

NEOTREE, a program that mimics the extinction of family names, caught the fancy of more readers than did PALEOTREE, a program that models the evolution of genera and species. Both programs were described in this department in the May issue.

Many readers were content to follow the evolution of a single genus without a computer, merely by tossing a pair of dice: a total of 2, 3 or 4 on the dice would make a species extinct, a 9 or a 10 would give birth to a new species and any other total would cause no change. The probabilities corresponding to these outcomes are respectively 6/36, 7/36 and 23/36. In my own experiments I had noticed the branching trees that trace the evolution often die out. But how many of the survivors, I wondered, would go on living forever? I. Jack Good of Virginia Polytechnic Institute and State University in Blacksburg, Va., pointed out that the single root of the equation $1/36(6 + 23x + 7x^2)$ lying between 0 and 1, namely the root 6/7, is the probability that such a tree is finite. Hence a seventh of my trees are destined for immortality. Robert M. Solovay of the University of California at Berkeley reached the same conclusion by an elementary argument. Meanwhile in Park Ridge, Ill., Edgar F. Coudal repeatedly ran the program needed to trace the complete evolution of a tree. Thinking that all the trees were destined for finitude, Coudal began generating them one evening in order to determine their average size. He writes that occasionally a tree would fill his screen, and an out-of-bounds condition built into the program would terminate the run. Not suspecting that infinity was the culprit, his thoughts turned to the consequences of the screen limit for theoretical paleontology: "Wouldn't [David M.] Raup [of the University of Chicago] be interested to find that...Nemesis cycles are really only coincidental out-of-bounds errors in the eternal CPU?"

Readers who tried NEOTREE all report the same results from the experiment with 1,000 family names. Given typical male birth patterns, the number of names decreased by three-fourths in just over 20 generations. For example, James W. Cox of Ottawa, Ont., found that half of the family names disappear in only three generations and two-thirds of them are extinct in 10 generations. There is a kind of stability thereafter. Matthew M. Cammen of Painted Post, N.Y., explains it this way: "After, say, 50 generations most of the names have over 100 living males who marry." Will the number of family names never shrink to one? Craig J. Albert of New York, N.Y., is skeptical. Experience with NEOTREE suggests that one must be "extraordinarily tenacious" in waiting for the event. The answer, as always, depends on the model adopted. For an abstract model in which the population grows continually perhaps 200 names will go on forever. No computer, however, can house more than a finite number of human tokens. When the limit is reached, the number of names will start to decline. After the year googolplex A.D. ($10^{10^{100}}$) either Smith or Chan will be lost forever.

I remind readers that ruthless programs can be seen battling in Boston during the last weekend of August at the first Core War tournament. Anyone traveling in that area might like to visit the Computer Museum, where the tournament will be held.

Rollo Silver of Box 111, San Cristobal, N.M. 87564, has proposed a newsletter devoted to the Mandelbrot set. He would be happy to hear from readers who would like to contribute or subscribe.

BOOKS

The continuing mystery of what kills apparently well infants as they sleep

by Philip Morrison

SUDDEN INFANT DEATH: PATTERNS, PUZZLES AND PROBLEMS, by Jean Golding, Sylvia Limerick and Aidan Macfarlane. University of Washington Press ($25). MELLONI'S ILLUSTRATED MEDICAL DICTIONARY, SECOND EDITION, by Ida Dox, Biagio John Melloni and Gilbert M. Eisner. Williams & Wilkins ($23.95). Three-month-old David's mother is a nurse in her early 30's. He is her third child; happily married to a physician, she does not smoke. Through normal birth she has just mothered a normal, contented baby. At two months he was screened at the health clinic: no abnormality at all. At three months, not long before David was due for his first immunization, she put him in his crib for his afternoon sleep. Warm but not encumbered, he went right to sleep. She heard no sound, but in 30 minutes she went up to check on him. He was very pale; when she picked him up, he was floppy and around his mouth was a blood-tinged froth. She cleared the tiny mouth, phoned for an ambulance and tried resuscitation. Within 10 minutes David was in the hospital, still warm but without breath or heartbeat. The emergency staff gave up after an hour's effort. The postmortem was entirely normal except for some vomit in the infant's upper mouth and airways. Certificate: Sudden Infant Death (SID).

Even during the late 1970's (the most recent data on hand) parents in Bangladesh had to face the fact that there one baby in six would die somehow during the first year. That was the hard experience in the U.S. or the U.K. just a century ago. Today in developed lands total infant mortality has fallen by more than an order of magnitude, to perhaps one infant death per 100 live births. Change continues as health care improves; since 1971 infant mortality in England and Wales has decreased by a factor of almost two.

There lies beneath that falling curve one small component that has remained quite flat over decades. Of each 1,000 live births in modernized lands, 990 infants can be expected to live to their first birthday. Of the 10 who die before that, six do not survive the first month of life, victims of gross errors of growth, mostly either low birth weight (prematurity) or a clear congenital defect. Two or three of the remaining four will die in the hospital, taken there during the first year for some late-detected birth abnormality or with an acute infection that resists treatment. The rest, two or three apparently well babies out of the 1,000, will die unexpectedly nonetheless, mostly in their own cribs, too suddenly to have reached the hospital or even a doctor, bearing no marks or signs, infants "in whom a thorough necropsy examination fails to demonstrate an adequate cause of death."

Taking the U.S. and the U.K. together, about 10,000 families endure such a blow each year. The definition of SID is plainly operational; it implies an autopsy, a degree of expert medical attention hardly to be expected where this faint signal is swamped in the grim noise of abundant infant death, whether in our own past or where the old regime still rules and ill-fed mothers take water from unprotected wells for pining infants. Recognition rests on some 40 studies in conditions of low total infant mortality, some local and some national. They have been carried out in Britain, Canada, the U.S., Australia and New Zealand, in a few countries of western and eastern Europe and in Japan and Israel. Wherever the statistics seem reliable the results are broadly similar; the one clear exception is in Sweden, where two big cities showed a fourfold lower rate for sudden infant death.

Fifteen years ago this column first noted the report of an international conference on the problem. Five such gatherings have convened in the past 25 years to unravel the causes of sudden infant death. The physicians have taken counsel, and yet the mystery remains unsolved.

This recent book by two British epidemiologists and a research pediatrician sums up clearly what is known in a fascinating and humane account, a fine case study in epidemiology and its problems. The small volume presents new data from two recent English surveys, one following up in detail the record of every infant death for a decade in a medically well-tended regional population with 14,000 births per year. The clearest result is the age distribution of these enigmatic deaths. They follow a curve that rises to a peak between eight and 12 weeks and thereafter trails off into a longish tail, a few cases occurring even at the age of two.

Some 600 or 800 papers that seek a cause are referenced here. Of course no single cause may dominate; the problem is after all a form of contemporary ignorance, diagnosis by failure to diagnose. Yet Table 13 lists no fewer than some 30 characteristics of mother and dead infant that show considerable consistency worldwide. These include such factors as the age distribution and the season of the year (deaths increase in winter). Risk is relatively high among infants born to young mothers, particularly those in the lower social classes and those who smoke or use drugs. Male babies and babies whose birth weight is low are also unusually susceptible. It is not hard to conclude that sudden-death infants often have endured "an adverse period *in utero*."

Yet for half of the cases the families were not under economic or social stress, a third of the mothers did not smoke, more than 80 percent of the babies who died were of normal weight; nothing amiss is known.

The failed hypotheses are legion. Breast-feeding does not protect statistically; inhalation of traces of cow's milk by an allergic infant was a popular theory of the 1960's. No antibody differences were found between sudden-death infants and the controls. The same story can be told of protein from mites in house dust. "Gentle battering," infanticide in some form, was a cause ascribed by the Bible, put delicately as accidental overlaying. There are cases of infanticide here; they hit the headlines, but the evidence indicates that they are a numerically unimportant fraction of the whole.

Siblings of dead infants do not seem to be at much greater risk, allowing for social and maternal biases, although there are certainly exceptional cases: one mother of an SID case in the big survey reported here had three other children who died suddenly and unexpectedly. Surviving twins are in fact at higher risk, mainly during the four weeks after the first twin dies; identical twins are not more at risk. Genetics seems to offer little.

The sudden arrest of breathing was suspect—some failure in bioelectronic development?—and gave rise to a special campaign around near-misses, children who had been found alive but blue from lack of breath. Consequently electronic alarms have been devised and installed. The results do not suggest that the scheme is an answer to SID, although parents who must raise siblings after one sudden death often value this kind of support. None of the 30 infants who eventually died among the more than 9,000 whose breathing movements were studied during the first weeks of life had shown abnormalities. The birth-control pill is not at fault, nor are the usual triple vaccines given infants, nor any other medical drugs, except possibly barbiturates taken during pregnancy.

There are some surprising hypotheses formed by analogy, unlikely and ingenious. Selenium deficiency leads to fatal heart disease among children in China and a kind of sudden infant death among piglets. In New Zealand selenium is low, yet SID is not high. It is true that there the city of Auckland has both more selenium and less SID, but the correlation did not hold elsewhere. Nor is selenium lower in the measured blood of SID infants than it is in the controls. Measurements of a dozen other trace elements, along with a variety of vitamins and hormones, have given no firm results. Overheating (perhaps augmented by fever) has been suspected of somehow causing death by heat stroke. Although febrile convulsions are common in young children, they occur only at ages far beyond the peak of SID.

One father of a sudden-infant-death victim had a "relatively rare" condition that causes a sharp rise of muscle temperature under anesthesia; the disease can cause sudden death among pigs. In 1982 an Australian team reported that five of the 15 parents of sudden infant deaths whose muscle tissue they sampled had the muscular disorder. The second shoe of confirmation has not yet fallen.

The age distribution among babies admitted to the hospital with respiratory syncytial virus (the common virus infection leading to bronchitis and pneumonia) matches the age distribution of SID cases. That virus infects boys more than girls in the same ratio found for SID; it is winter-prevalent, more likely in cities, lowest in better-off families. The match is striking; it is not a negative point that other and less common viruses might show similar patterns. The immune system of the child is in transition, shedding its endowment of maternal immunoglobulins in favor of its own defenses. This event transpires between the age of two and three months, the age peak of SID. Perhaps anaphylactic shock to the sensitized infant in reaction to respiratory virus infection is the pathway to "a proportion" of SID cases, the authors cautiously write.

About half of a modest sample of SID cases do yield small amounts of common viruses. This is a larger fraction than the controls show. Of course, there is no massive viral infection; such a clear finding would have excluded the infant from classification as a sudden infant death. The hypothesis must include some kind of hypersensitivity.

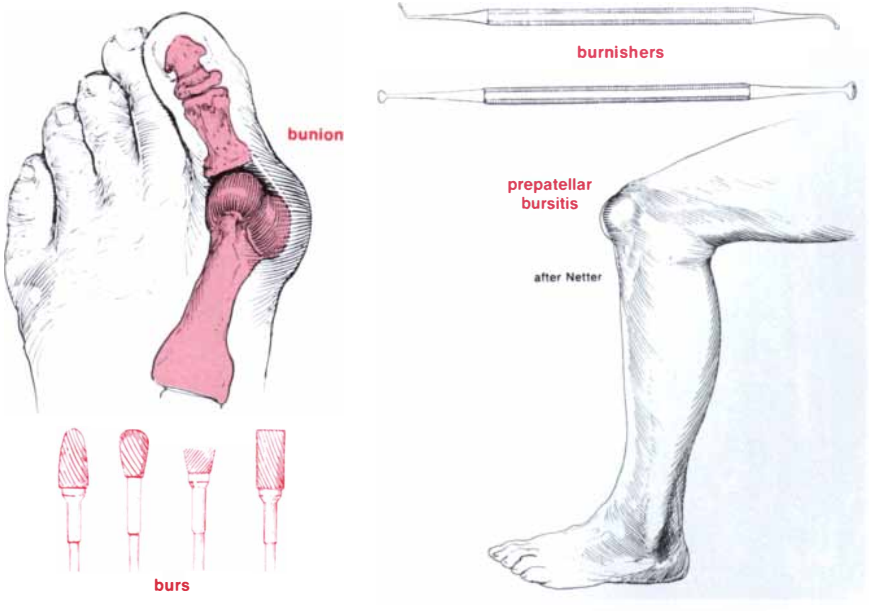
The riddle is still unanswered. Controls are difficult: not many normal children die at a comparable age. In the absence of a standard of comparison erroneous yet plausible hypotheses easily persist. The cause of sudden death in the 1900's was held to be a grossly enlarged thymus. Actually the thymus is normally large in infants, but the "controls" of the day were babies who had died of malnutrition or infection; their thymus glands had shrunk. It took 50 years for those facts to be generally recognized. Misunderstanding and fashion enter today as well. Current tests center on biochemical studies, cases of near-misses and the follow-up of infant siblings.

The logical outcome of advances in medical knowledge is certainly a steadily declining incidence of death "from unknown causes." SID is not as complex as the eventual residual mix of everything unknown. But the age distribution of sudden deaths does differ strongly between those in summer months and those in winter. That suggests we may be on the verge of resolving SID into two differing patterns, perhaps with distinct causes.

Even more than SID's intellectual challenge, even more than the human suffering it brings—both well described in this study—the circumstances pose a deeper riddle yet in our contradictory epoch. If we can struggle so steadily against that half percent of SID, can we not act to relieve the infants of the Third World, who are dying at a dozen times the SID rate from familiar microbial infections we know very well how to prevent? If not we, who? If not now, when? Those are old questions, but they still have bite.

Biagio John Melloni's medical dictionary is recent, legible, neither overpriced nor uncomfortably bulky. Its 26,000 entries are made much more usable by a richness of about a tenth as many apt and well-arranged tables and diagrams, and by illustrations of anatomy at every scale. Size is in general not explicitly marked, although the entries contain the data for objects, such as red cells, not visible without the microscope. The work is for nonspecialists; it tabulates, for instance, all arteries, bones, muscles, nerves and veins by name. The generally useful glossary of the SID book omitted a definition of the suspect respiratory syncytial virus, but that definition and a little context were quickly found in the attractive volume by Melloni—the image expert—and his friends.

BUILDINGS FOR MUSIC: THE ARCHITECT, THE MUSICIAN, AND THE LISTENER FROM THE SEVENTEENTH CENTURY TO THE PRESENT DAY, by Michael Forsyth. The MIT Press



Drawings from a page of **Melloni's Illustrated Medical Dictionary**

($35). The first page presents a fine colored engraving of the most celebrated of musical Lutheran churches, the Thomaskirche, in its square in Leipzig, taken from a prospectus issued when the cantor there was Johann Sebastian Bach. The color illustrations offer a look at more than a dozen of the opera theaters and halls that have echoed splendidly to the sounds of music in Europe since Bach's day, to end with the underground Espace de Projection at the Centre Georges Pompidou in today's Paris, a versatile experimental chamber for 400 listeners, "virtually a musical instrument in itself," controlled by the composer.

Two ancient acoustic prototypes appear to span the auditoriums of the world. There are those that resemble the open air, where the spoken voice belongs. They are unreverberant, for any echo is a risk to clarity. Around the other pole are grouped the spaces that share the acoustics of the cave (or of the shower stall), chambers fit to encourage the sounds of music through harmonious reverberation. The harmony that is characteristic of Western formal music may indeed have begun not by intention but by the unplanned overlap of resonant echoes within lofty Gothic enclosures of hard and massive stone.

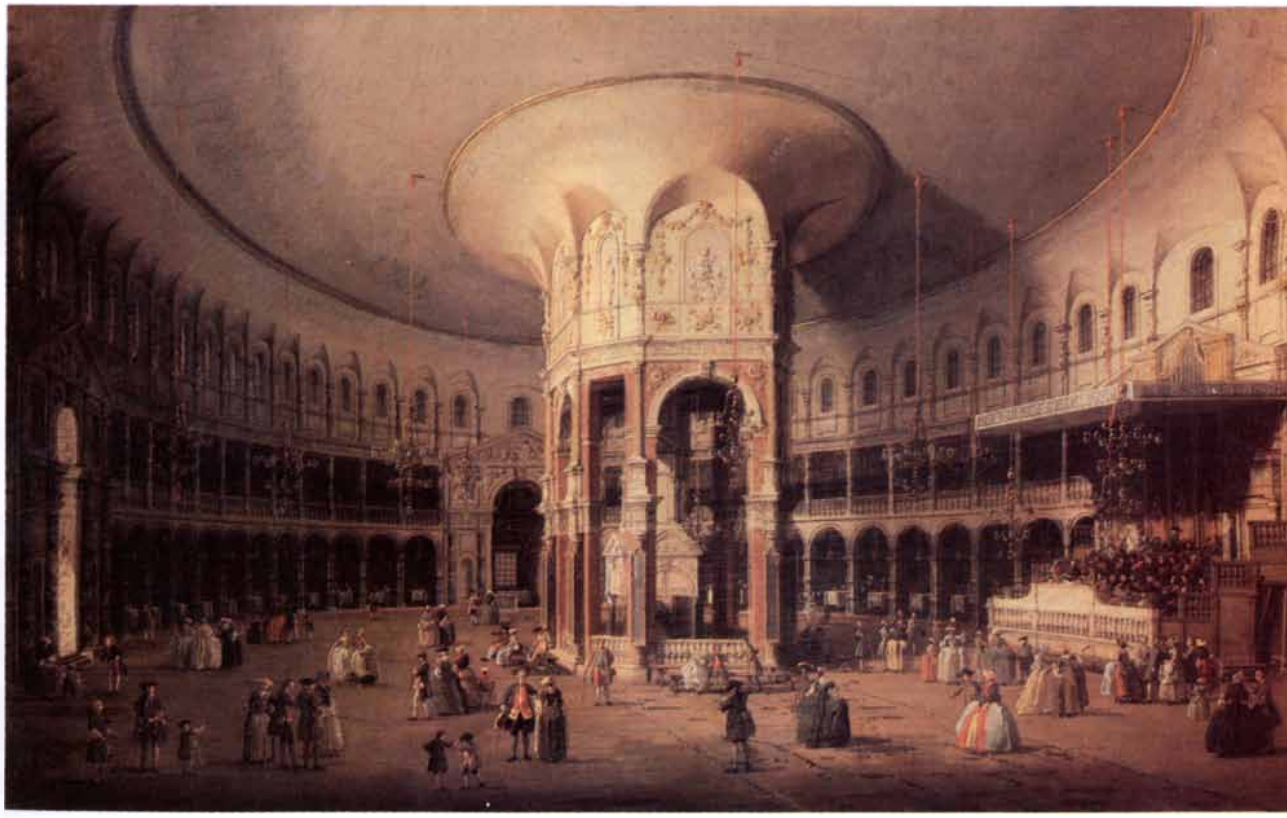
Undoubtedly musicians have long understood something about acoustic spaces. Bach's Thomaskirche, even before his day, was remodeled like many another Protestant church by introducing draperies and new balconies nearer the pulpit to help public understanding of the preacher's words, given new weight by the Reformation. Bach composed choral works suited to the rather short reverberation time characteristic of the Thomaskirche when it is filled with worshipers; his swift rates of change of harmony and his brisker tempos would not have worked in the original church.

Rational acoustic design was not yet possible; that began only at the turn of this century, and its maturity is arguably still to come. The main ingredient of architectural acoustic success was luck. Behind the fine musical reputation of many old spaces lies natural selection and a touch of myth. "The even greater posthumous reputation that halls tend to acquire can rarely be matched, in buildings old or new!"

Michael Forsyth (of the University of Bristol) takes a reader on a cultivated and nontechnical tour of the places built over the past three centuries that embody the interaction of social structure, architectural design and formal musical style in western Europe and America. Musical professionals had by 1650 come to be at home as much in the privacy of the court as in the echoing public cathedral. The princely music room was not specially designed, merely chosen among convenient halls or salons, although perhaps suitably decorated. The public concert hall began in England; the earliest one we still have is the Holywell Music Room, built by private subscription and opened at Oxford in 1748. The little hall is a rectangle except for a rounded end; its strong early reflections give it a bright and powerful tone. When "the little German boy WOLFGANG MOZART," eight years old, and his sister played daily for the Londoners, they used the Great Room at the Swan and Hoop. Taverns remained favorite concert sites in England throughout the 18th century, as coffeehouses were in Germany.

The Continental nobility began to build purposeful concert rooms in their residences; Prince Esterházy's music makers, led by Papa Haydn, enjoyed and conformed to the intimate acoustics of his great music salon. Haydn scored many of his symphonies of that period twice: one version with trumpet and kettledrum for performance out of doors, the other without their intensity for the good-size but not enormous concert hall.

It was in Venice that the first opera house was built in 1637. That form of musical drama, expensive to produce in view of its multiple artistry and



A painting by Canaletto of the Rotunda at London's Ranelagh Garden

elaborate stage tradition, had worked its way from the lavish entertainments of Roman nobility to the well-to-do mercantile public along the Rialto. Its horseshoe theaters were the earliest buildings intended for musical performance. The architects grasped something about their task. They recognized that minimal dimensions were essential. The crowded audience in full costume provided a helpful absorptive surface. The thin wood panels that formed most interior walls pass and hence eliminate the low frequencies but reflect the higher ones. The result is a quick dying out of sound, after an early crisp reflection of voice detail. The opera houses, even such grand ones as the Teatro alla Scala in Milan, show short midfrequency reverberation times, like the much smaller concert halls that spread from London.

Large orchestras, larger audiences and still larger conceptions of music followed; romanticism flowered in music and architecture at once. London's Royal Albert Hall is so large—eight times the volume of La Scala, although with less than three times the seating capacity—that its echoes are spaced by an easily perceptible interval, one-sixth of a second (a difference in path length of 180 feet). Its troubles are compounded by an elliptical ceiling vault, whose focus happens to be near ground level. At another acoustic extreme, the vast and wonderful Crystal Palace of greenhouse architect Joseph Paxton, its 10,000 modular tons of glass panes and iron framing moved and reconstructed in the London suburbs, was as transparent to sound as it was to light. Its properties were indeed those of open air; it nurtured massive music by British composers for nearly a century, but it always diminished the sound expected from the monster assemblies, orchestras 500 strong and choirs by the thousands, that served its gala performances.

Science approached the acoustics of halls from two directions. In 1838 John Scott Russell proposed ray tracing. Sight lines determined an optimal seating geometry, both for viewing and for sound, along an "isacoustic curve." The method remains widely used. The pioneering Chicago architects Dankmar Adler and Louis Sullivan built in 1889 the immense and daringly advanced Auditorium (during the summer its air intakes were ice-cooled) on Russell's acoustic lines. Unfortunately Adler contrived the big, complicated ceiling along similar curves. The size and the vagaries of focus made the hall quirkily reverberant, almost unusable by the orchestra, an exciting building but, Forsyth reports, an acoustic disaster.

In the spring of 1896 the Harvard physicist Wallace Sabine measured the times for decay of sound as he systematically varied the absorption of his test hall by using a large supply of movable seat cushions borrowed from Sanders Theater. His source was a short blast on a single organ pipe; his detector was his own ear; he controlled ambient noise by working in the small hours of the Cambridge night, its stillness broken only by the rattle of an infrequent trolley. He opened the door to the basically statistical portion of the modern theory of room acoustics, the phenomena of sound paths so manifold that ray tracing can hardly help. Boston's Symphony Hall, no daring innovation, had the benefit of Sabine's consultancy in its design; he matched it to the successful hall at Leipzig, rescaling by his own theory to suit the demanded change in size. Its shoe-box shape remains a favorite; we understand the reasons better nowadays than Sabine could with his single-parameter theory of reverberation time. Yet his good sense and well-based partial hold on the problem opened a new era.

Real mastery of room acoustics is not yet quite here. Modern halls have their failings too, often because the architect has programmatic priorities higher than fulfilling the specifications of his acoustic consultant. Remedial acoustic repairs are not uncommon: New York's Avery Fisher Hall, the Royal Festival Hall in London and in some sense the Sydney Opera House are notorious examples.

For more than 20 years acoustic success has been the norm. One solution is more powerful than it is elegant: it has centered on variable acoustics, achieved by heroically movable architecture. In the experimental studio of Pierre Boulez at the Pompidou Center the ceiling and walls are all motorized and mobile; the volume can be reduced fourfold, and all six paneled surfaces are changeable. Another way is that of electronically assisted resonance: microphones coupled to amplifiers and speakers are installed in large numbers all over the hall. By the use of filters, delays and adjustable gain a hall can be given about any reverberation characteristics one might want; the characteristics can even be changed during the performance. The dozen recent halls that use this system are so far more prudent: they modify built-in times by a modest fixed factor.

The high-fidelity world of audio is new and strange; we are barely across the threshold of the musical era of the computer. This interesting study is fine preparation for what is beginning to happen in public and has already transformed our private musicality. Clearly and agreeably written, richly and authentically illustrated, informed by an uncommon sense of period, the volume is a necessity for readers caught up in the concert world, and a pleasure for the others.

THE WORLD FOOD PROBLEM 1950–1980, by David Grigg. Basil Blackwell ($39.95). One morning 25 years ago a well-fed traveler along the Malabar Coast offered the lunch his hotel had provided, fresh chicken sandwiches on crustless white bread, to the slight, wiry men whose work it was to pole the heavy car ferry across a narrow arm of the sunstruck sea. His experienced companions assured him such strange food would not find acceptance. They were wrong; the men were hungry, and they wolfed the stuff down, unfamiliar or not.

One way to test the adequacy of diet is to observe the response to an offered increment. A page of graphs in this book shows a more representative version of just such a test: the steady climb of calories consumed, geared to rising consumer income, as sampled during the 1970's in four countries around the Third World. "Such data support the assumption that poverty is the main cause of hunger."

David Grigg, a geographer at the University of Sheffield, summarizes what we know of food worldwide in this cool, factual and critical book, the fourth he has published on the broad topic. He has chosen not to remain strictly within the time confines of his title; rather, he provides a valuable short history of hunger over the centuries. He also directs attention to the uncertain methods that bring us the copious noisy statistics forming what quantitative knowledge we have of such a complex problem.

By now the literature is large and understandably tendentious, not to say heated. Sir William Crookes addressed the wheat problem as early as 1899, but his concern for the exhaustion of natural supplies of nitrates was elicited by what he saw of European needs, not of world ones. Their proudly inclusive titles notwithstanding, most of the other studies published before World War II were in fact focused on Europe. Only since 1943 has the rest of the world received its share of attention, notably through the four World Food Surveys that were issued by the Food and Agriculture Organization of the United Nations between 1946 and 1977. (There is a one-page table summarizing those four reports.)

The present world food problem can be tersely stated. A "large proportion of the population of Africa, Asia and

Latin America is either undernourished or malnourished or both." The kernel of the matter is not the acute but transient state we know as famine; it is a chronic condition. While we now fashionably label this a world problem, it is not genuinely worldwide: malnutrition is no longer a public-health problem in western Europe. The food supplies of those lands have been adequate ever since the mid-19th century. True, the aggregate was misleading: the total supply was not justly shared. A table shows the relation between calorie intake and household income in late Victorian England, a marketplace version of test by increment. Families with servants had twice the calorie intake and enjoyed more than twice the protein per head of families in the lowest income group.

It was a full century from the first achievement of adequate overall national food supply before such telltales as army rejection rates, adult height and infant mortality were able to testify as one voice to widespread adequacy of diet. That condition has been reached in the developed world mostly within the past 30 years (although not yet for members of some minority groups in America the Bountiful). The difference between the first two worlds and the third is that "in the developed countries the lowest income groups have, generally, sufficient money to buy an adequate diet." (In some countries rationed supplies in one or another way attempt to guarantee a minimum for all.)

There is no agreed census of the hungry. No longer is their problem seen as a problem of protein shortage alone. Calories count. Vital statistics give the firmest data, but they are indirect. Food-intake studies are closer to what we seek, but they necessarily involve detailed and expensive study, and that implies small samples. Estimates of overall food supply country by country are useful; unfortunately they require supplementary income data to assay comparisons between richer classes and poorer. Even more detail is needed if differences in availability between parents and children are to be discerned. No wonder we end up with a number for the hungry that varies widely with source and method, estimates that range from some .5 billion to almost 1.5 billion of all 4.5 billion of us travelers in solar orbit.

Is it supply that is wanting, too many mouths for the crops spread out under the sun? Or is it economic demand that fails, too many hungry people unable to buy what they need? We can even ask whether those two states are really distinct. Three chapters take up the question of world supply. Certainly food output has grown steadily, doubling since 1950. World food production now outruns world population; better perhaps to say that it outwalks population, for its annual growth leads only by a narrow 1 percent.

Three regions brand all the maps with the hatchings of undernutrition. Of these, tropical Africa has the poorest farm performance. Until about 1950 traditional methods of farming based on fallowing systems produced a "poor but adequate" diet. The population has doubled since then, and output has increased—lagging a little behind the population growth—mostly by reducing the time in fallow. Shifting agriculture, with 20 years in fallow, so that even forest regrows, is now very rare. Bush fallow, too short in duration to reestablish anything but scrub, is widespread; permanent agriculture, its fields left fallow for only a year or two, occupies little area, but an area that has to support most of the population. Modern inputs, even the ox-drawn plow, have not much entered this continent. Little of the land is found in large farms. Most African peasants remain subsistence farmers in the old way, smallholders by communal tradition, producing yields that change slowly through the long march of the fat years and the lean.

Latin America is the second region of hunger. There two agricultures coexist. Poor peasants with small subdivided farms on steep eroding slopes watch their families migrate to the mushrooming city slums or go as labor to the big haciendas, the large farmsteads that hold two-thirds of the continent's arable land. Once such landowners were content to take the rents of many poor tenants, and paid little heed to increases in output. They enjoyed plenty in any case. Since 1950 the grand indifferent style has dwindled as the large holdings have been industrialized to produce cotton, sugar, bananas or soybeans in quantity by modern techniques, for the cities and for export. Production has increased mainly by expansion of the area sown in crops; few countries have food supplies that are less than the national requirement. The problem of hunger from the Rio Grande to the Andes is lack of enough income to buy food or the land to grow it on.

Most of the hungry are Asian. In Asia agriculture resembles that of the old European tradition more than it does that of Africa or Latin America. Farms tend to be small, and they occupy most of the land. They are worked by families; the power source is still mostly human labor. Mechanization has risen rapidly since the 1960's; it is estimated that in 1981 almost half of China's sown area was tractor-plowed at least once "compared with 1.6 percent in 1960," a remarkable change. In Asia too food output has kept up with population growth. There are few new lands, and not much fallow. The increase in output has been achieved by conventional means: more irrigation, intensive cultivation, multiple cropping, improved seed strains and increased fertilizer. Yields per hectare have risen over the past three decades for all major crops; Asian wheat yields rose 50 percent outside China and Japan, and within China by more than 100 percent.

Professor Grigg casts his survey net wide, yet the work remains broadly centered on agronomy and agricultural economics, on crops and land use, on inputs and markets. The larger issues of war and inequality, of foreign exploitation and domestic depredations, are sketched, but they could hardly be treated in the quantitative and comparative tone of the tables and graphs that crowd these pages.

There are many who seek to change our hungry world; they will usually need first to understand it. This cool historical account of the facts can help anyone; it does so less by teasing out simplicities than by providing a documented and analytic roundness. The illuminating volume makes no predictions and claims no one overriding cause of hunger. It closes with a few sentences of modest hope, welcome next to the daily headlines of war and injustice. "In an age of unprecedented population growth food output has also increased at hitherto unknown rates, and the proportion of those undernourished has declined. This may not be good enough, but it is a considerable achievement."

COMMUNICATIONS SATELLITES, by Larry Van Horn, edited by Robert B. Grove. Grove Enterprises, Inc., Brasstown, N.C. 28902 (paperbound, $14.95, plus $1.50 shipping charge in the U.S.). Hardly an American roadside hotel or well-established tavern nowadays lacks its microwave dish staring loyally up at the celestial equator whence the video signals come, from a couple of dozen geosynchronous satellites. If it is far enough north, the seeking dish must point almost to the horizon, where the desired signal is endangered by obstructions and mirages of every kind.

The engineers whose task it was to supply a reliable video link to television stations across Siberia found another way. Instead of the popular 24-hour orbit they chose a 12-hour one and placed powerful birds into a long elliptical loop. At the far end one Mol-

niya satellite (class 3) stands over Tomsk, 10 earth radii above the turning planet; half a day later it travels over Hudson Bay. For five or six hours during each orbit it remains in easy view of the big, slow tracking earth dishes of Siberia until it begins to set southward. By then another one of the four satellites marching along nearby orbits stands high enough to take its place; two minutes of color bars and the logo ORBITA mark each normal switchover. Here is a photograph of the evening news anchorman in Moscow taken from the television screen by an American amateur who monitored the Canadian apogees with his four-meter dish, tuning for the broad beam and strong signals just below U.S. television Channel 9.

The Molniya format is the international standard SECAM; any U.S. television set not too digitally finicky can convert it into adequate black-and-white pictures. Color would demand a better-matched system. There are several video channels accompanied by their audio bands, as well as lots of other services, including telephony, the national radio network in Russian, and even the famous "hot line" (carried now on an older class of Molniya satellites). The television programs range "from humdrum to dull, demonstrating a startling poverty of imagination, much like American television," according to our jaded informant, who adds, however, that "the glimpse of Soviet society as they see it is quite extraordinary." A considerable number of geosynchronous satellites also broadcast Soviet television to the entire country and to all the ships at sea; seen from this hemisphere, "selected reception varies widely."

This unusual book is a work of devotion by Larry Van Horn, a knowing amateur, and his friends. It intends to provide a comprehensive introduction to the art of monitoring the new artificial ionosphere, the orbital domain of 1,000 listed frequency channels for every flag and purpose. It is personal, tentative in organization, typography and layout; its wealth of material includes official lists and summaries, widely culled magazine reports (not always consistent), firsthand results and reasoned conjecture.

Here are details of the dozen satellites, American, English and Soviet, that were built by and launched for amateur radio operators around the world, who use them as intermediaries for relayed conversation. Radio sputniks 5 and 7 offer robot contacts; the five-watt satellite calls for anyone to answer, responds to the first acceptable Morse-code reply, seeks repeats until it has your identity straight, and later on politely mails a confirmation. Van Horn describes the wonderful weather watchers, reproducing a photograph of clouds over Cuba and the Gulf taken from one Soviet orbiter, *Meteor,* which uses a system much like that of the NOAA polar-orbiting satellites. The nature of the geosynchronous weather services and access to them are treated in some detail; they offer the amateur meteorologist "the ultimate toy," frequent and current weather images and data from U.S., Japanese and European satellites.

The morass of domestic and international satellite video, telephone and data channels is surveyed heroically if rather confusingly. If you want Telebet "Racing Alive" (scrambled) or Country–Coast to Coast or the Playboy Channel or the Ottawa Parliamentary Channel, it is listed here. The international services are of course even more tentative, but the entire polyglot world is found: Argentine television in color, Norway radio, Sahel television from Niger and more. The beamed footprints that serve antipodal Perth and Kuala Lumpur are listed but are hardly visible in these parts.

The latter half of the book treats the largest sector of the COMSAT world, space data, military and surveillance satellites, as well as navigation, search and rescue, links to manned spacecraft and other specialized uses. Here the firsthand experience, although it is not quite absent, falls in importance; the data are less complete, the signals are encrypted or at least unannounced, the technology is more advanced. There are plenty of brief official descriptions, drawings and photographs, savvy guesses, launch histories and a summary of past successes in teasing out what is going on, particularly the work of the Kettering monitoring group in England. It makes fine reading for anyone who likes timely little puzzles of technique and purpose. As usual, there is much more information easily available than the uninitiate might expect. There are also pleas for readers to share what they have learned. At orbital distance the intelligence community today still evokes the "Great Game" of Rudyard Kipling's high-Victorian fiction.

The closing appendix lists by longitude more than 200 satellites in geostationary orbit, all those registered with the International Telecommunications Union in Geneva, both civil and military—Indonesian Palapa, American FLEETSATCOM and Soviet Raduga—with their assigned stations along the orbit and their frequency bands. Eighteen flags circle there, including those of three international consortia. The earth too now has its ring.

SCIENCE/SCOPE®

A U.S. spacecraft orbiting Venus made the first close-up views of Halley's Comet, giving scientists valuable insights into the comet at a time when it was on the far side of the sun and direct observations from Earth were impossible. NASA's Pioneer Venus Orbiter, built by Hughes Aircraft Company and circling Venus since 1978, conducted its investigation a month before five other spacecraft flew by the comet. The Orbiter was delicately repositioned with precise commands from Earth to observe Halley's at its closest point to the sun, a distance of about 55 million miles. The spacecraft measured changes in the comet caused by intense solar heating. It also provided an ultraviolet image of Halley's and its large surrounding hydrogen cloud. Data gathered by the Orbiter helped scientists determine the gas composition of the comet, the rate at which water vaporized, and the ratio of gas to dust in the comet.

The AMRAAM missile may become the next-generation weapon for protecting U.S. Navy surface ships against threats that have slipped through the outer defense shields. Sea AMRAAM, under study for ship self-defense, would be essentially the same as the Advanced Medium-Range Air-to-Air Missile in full-scale development by Hughes for the U.S. Air Force and Navy. However, compared with existing missiles, Sea AMRAAM would increase a ship's firepower because the missile's guidance system is much less dependent on the ship's radars. Many missiles could be fired at different targets simultaneously, and they could home in even if the targets were outside the field of the ship's radar systems. Sea AMRAAM is also faster, more maneuverable, and can fly farther than current ship self-defense systems.

An innovative digital receiver is being developed to alert military aircraft when they are approaching enemy radars and electronic warfare systems, thereby putting them at less risk while on a mission. The device, designed for electronic support measures (ESM), will be approximately 1/20 the weight and substantially smaller than current receivers. It will search for, intercept, record, analyze, and locate sources of radiated electromagnetic energy. The receiver can store this information. Or, if an enemy signal poses a threat, it can pass this information along to another type of electronic warfare system, such as a jamming device. Hughes is developing the receiver with independent research and development funds.

Cellular telephones may take a back seat to a proposed satellite system when it comes to making long-distance calls. The mobile satellite network, consisting of two Hughes HS 393 spacecraft, would relay two-way voice and data communications services directly from airplanes, cars, trains, or remote locations. While cellular telephone systems are limited to areas equipped with fixed antenna networks, mobile satellites would cover the continental U.S. and Canada, and possibly Mexico. Users would have their own mobile ground terminals. Hughes Communications Mobile Satellite Services, Inc. is seeking authorization from the Federal Communications Commission to operate the system.

Hughes needs engineers, scientists, and programmers to forge new frontiers in aerospace radars, weapon control systems and avionics, airborne displays, aerovehicle data links, and airborne countermeasures. Current openings are for people experienced in design, development, test and manufacture for systems engineering, project/program management, design of circuits and mechanisms, and bringing these to reality through the application of advanced manufacturing techniques. Send your resume to Hughes Radar Systems Group, Engineering Employment, Dept. S3, 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
AIRCRAFT COMPANY

Subsidiary of GM Hughes Electronics

Established 1845 SCIENTIFIC AMERICAN August 1986 Volume 255 Number 2

The Strategic Nuclear Forces of Britain and France

They could more than quadruple in size by 1995. So far they have been exempt from any arms-limitation treaty, but it may be in the interests of both superpowers to include them in future arms talks

by John Prados, Joel S. Wit and Michael J. Zagurek, Jr.

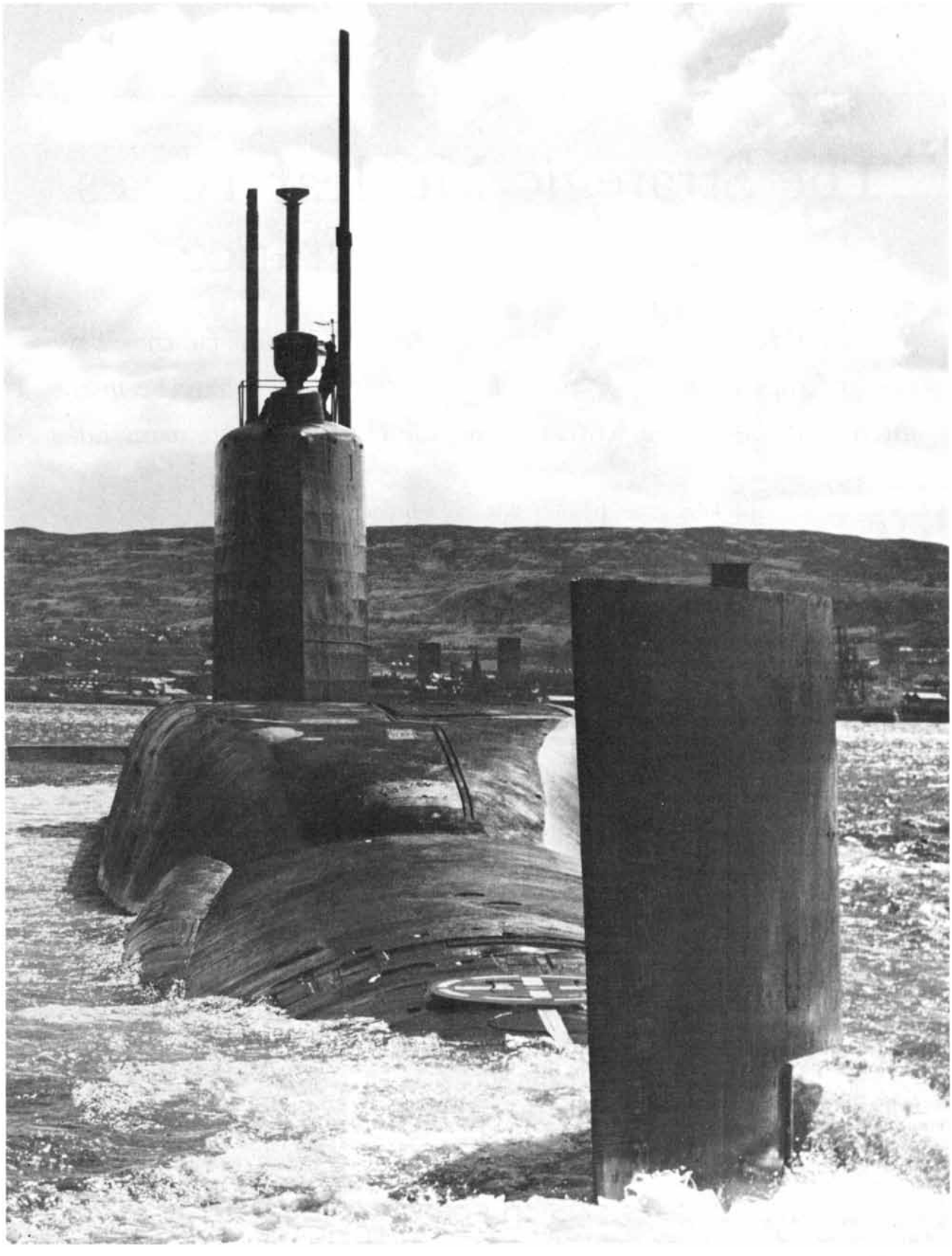
When people speak of nuclear weapons, they are generally referring to those in the arsenals of the U.S. or the U.S.S.R. The weapon inventories of these two countries account for about 97 percent of the nuclear warheads in the world today; Britain and France have most of the remainder. Yet even the comparatively modest claims the two Western European nations can make in this regard represent an awesome military capability that the U.S.S.R. (and, to a lesser extent, the U.S.) cannot afford to ignore. Indeed, the recent proposals both superpowers have made to dismantle all their European-based intermediate-range missiles differ sharply in particulars but raise the same fundamental question: What role will the "other" nuclear forces in Europe have in the future military balance?

The American proposal excludes any third party from whatever limits may ultimately be negotiated in Geneva. The Soviets, however, insist that in order for such a "zero option" to be viable British and French forces must be frozen at their current level. The Soviet demand is intended to prevent the expansion and upgrading of nuclear forces that both Britain and France have planned. These so-called force-modernization programs, if they were fully implemented, would make the combined Anglo-French force equal to a significant percentage of the Soviet strategic force—particularly if the Soviet arsenal were to be reduced through an arms-control agreement. If nuclear-arms control is to proceed further, the European strategic forces may soon have to be included within its purview.

Yet the Europeans themselves are unlikely to place their limited strategic forces on the bargaining table along with those of the U.S.S.R. and the U.S. unless the superpowers impose severe constraints on their own strategic-weapon programs, and in particular on their strategic-defense programs. The deployment of such defense systems poses a worrisome dilemma for the Europeans: while in the near term a Soviet anti-ballistic-missile (ABM) system might have only marginal military impact, over the longer term a novel, "exotic" defense could conceivably nullify the independent deterrent Britain and France have so arduously sought. At the same time an American defense system (such as those being considered under the Strategic Defense Initiative, or SDI) could, by decoupling the defense of Europe from that of the U.S., increase the need for precisely such an independent deterrent in the minds of the British and the French.

The fear that at the moment of truth the U.S. might not commit its nuclear forces to battle in defense of Europe has always provided the main impetus for the autonomous European nuclear-weapon programs. Britain, however, has worked out a compromise between its perceived need for an independent nuclear force and its unique position of trust with the U.S. In exchange for maintaining at least a semblance of subordination to the nuclear-war plans drawn up by the U.S. along with the other allies in the North Atlantic Treaty Organization, it has received much valuable (and classified) technical support. Indeed, from 1958 to 1982 every nuclear warhead fielded by the British armed forces was either copied from a U.S. weapon or designed and built in the U.S.

Britain's own contribution to its strategic nuclear force (defined as the nuclear force that could attack, from its normal stations, targets in the territory of the Soviet Union) has largely been limited to the construction of four *Resolution*-class nuclear-powered submarines. These aging submarines are still the mainstay of the British strategic force. Each is capable of firing 16 American Polaris submarine-launched ballistic missiles (SLBM's). Britain has replaced the missile's original warhead-dispensing system with a more advanced system of British design known as Chevaline. It is intended to overcome ABM defenses (in particular the Galosh defense currently protecting Moscow) with a combination of decoys, obscuring chaff and maneuvering reentry vehicles (RV's). From three to six RV's, each containing a 40-kiloton warhead (with a yield equivalent to the explosive energy of 40,000 tons of TNT), are said to be carried by Chevaline-outfitted SLBM's. (They are not multiple independently targetable reentry vehicles, or MIRV's, al-



H.M.S. *REPULSE* is one of four *Resolution*-class nuclear-powered submarines in Britain's Royal Navy. Each submarine can serve as an underwater launch platform for 16 U.S.-supplied Polaris ballistic missiles that have been outfitted with British nuclear warheads. Although Britain's ballistic-missile-submarine fleet is nominally a part of the allied forces of the North Atlantic Treaty Organization, Britain has reserved the right to assume unilateral command of the fleet if its supreme national interests are imperiled. The *Resolution*-class submarines are to be replaced in a decade by four new submarines that will carry more capable U.S. Trident II missiles.

though they can strike more widely separated targets than the earlier Polaris RV's could.)

Further modernization of the submarine-based deterrent is considered a pressing problem in Britain. By the mid-1990's the *Resolution*-class submarines will be almost three decades old and ready for retirement. Moreover, by that time keeping a ready stock of operational Polaris missiles would be difficult and costly. (Britain has already been forced to replace defective solid-fuel Polaris motors at great expense.)

In July of 1980 the administration of Prime Minister Margaret Thatcher announced its decision to buy the Trident I SLBM from the U.S. and to install the missile in four new British-built submarines beginning in the mid-1990's. In March of 1982 the British government revised its decision, choosing instead to purchase the more capable Trident II SLBM (which is not due to be operational until 1989 in U.S. submarines). The new British submarine would carry 16 SLBM's, each outfitted with eight MIRV's. The British SLBM force would then reach a total of 512 warheads.

Whether Britain will be able to carry through its Trident modernization program fully (or, for that matter, whether it will even continue to field an independent nuclear force) is an open question. A strong disarmament movement has taken hold in Britain, and the opposition Labour Party advocates unilaterally abandoning the nuclear deterrent. Some Conservative members of Parliament as well as members of the Social Democratic Party have also attacked the nuclear program on financial grounds. They believe the money might be better spent bolstering Britain's conventional military forces.

The unsettled issue of modernization will almost certainly continue until the next election, possibly as late as 1988. By that time so much money may have been spent on research, development and production (the total cost of which is currently estimated to be about nine billion pounds, or about $15 billion) that it may be politically difficult to cancel the entire Trident program regardless of which party wins the parliamentary majority.

In contrast to its neighbor across the English Channel, France has remained staunchly aloof from the U.S. while pursuing its own nuclear-weapon program, which includes aircraft, land-based missiles and nuclear submarines. The primary aircraft of the French strategic air force is the obsolescent Mirage IVA, introduced in 1964. These bombers, based in northeastern France, currently number about 22. With airborne refueling the Mirage IVA has a range of 3,200 kilometers, enough to reach targets in the western Soviet Union. Each aircraft is reported to carry a 70-kiloton nuclear bomb. France is equipping 18 of the bombers with a medium-range air-to-surface nuclear missile to preserve the aircraft's effectiveness as a delivery system for another decade.

France's S-3 missile, based in underground silos, is also able to strike targets in the Soviet Union. Each missile carries a single one-megaton warhead (equivalent to a million tons of TNT). Only 18 have been deployed. Given the vulnerability of such fixed, land-based missiles to a nuclear attack, the French government has studied the possibility of deploying a mobile intermediate-range missile, designated the S-X. The deployment of such a missile is uncertain, given fiscal constraints and doubts as to the military need for a larger land-based strategic force.

Like the British, the French regard a fleet of ballistic-missile submarines as the cornerstone of their strategic nuclear forces. Half of France's total investment in nuclear-weapon systems has been devoted to the strategic submarine force, the Force Océanique Stratégique (FOST), since its establishment in 1971.

FOST consists of five *Redoutable*-class submarines and one newly commissioned *Inflexible*-class submarine. The older submarines are armed with 16 single-warhead M-20 SLBM's. The new submarine is armed with the M-4 SLBM, the first MIRVed French missile. Since an M-4 carries six RV's and there are 16 of these missiles on board the *Inflexible,* the number of SLBM warheads available to the French more than doubled when the submarine joined the force last year. By the early 1990's five of the six submarines will carry the M-4 missile. As a result France's sea-based nuclear force will have a total of 496 warheads. A seventh submarine, to be equipped with a new MIRVed SLBM (designated M-5), is scheduled for delivery in 1994, further increasing the total.

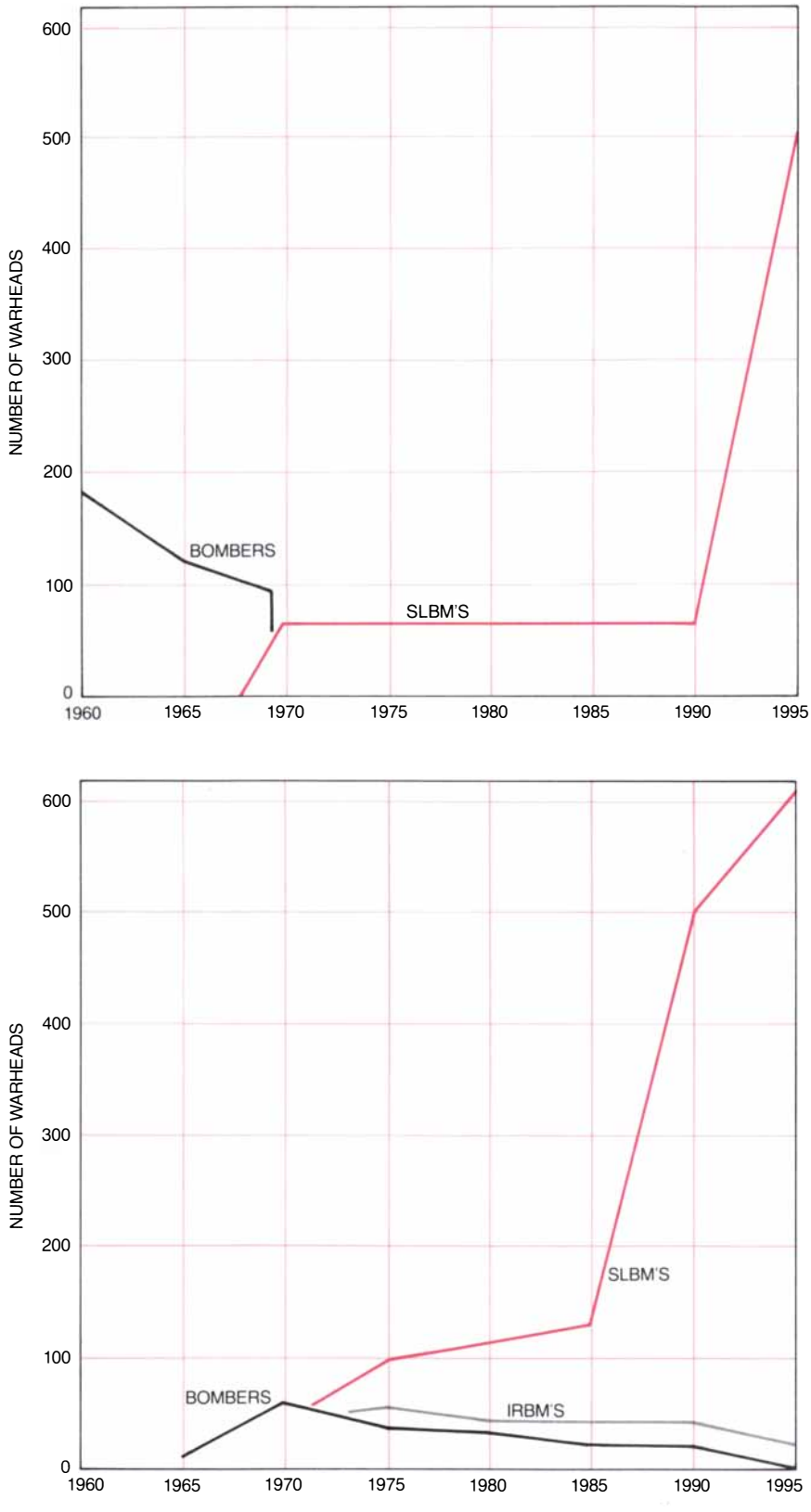
The French modernization plans seem unlikely to be upset by the vagaries of domestic politics. There appears to be a remarkable political consensus in France in support of its nuclear program. Although the Socialist and Communist parties were initially opposed to the idea of a national strategic nuclear force, since the late 1970's they have joined the parties of the center and the right in supporting it. Indeed, the Socialist government of President François Mitterrand has pushed for modernization of the nuclear force at the expense of France's conventional forces (particularly as reflected in manpower levels).

Along with doubts about the reliability of the American guarantee to engage its nuclear forces in defense of Europe, arguments emphasizing the intrinsic political and military value of nuclear weapons have been advanced by British and French government officials to justify nuclear-weapon programs. Advocates of an independent, albeit modest, nuclear arsenal argue that an ever present threat to detonate nuclear warheads over cities in the enemy homeland is enough to effectively deter any potential aggressor—including as formidable an opponent as the U.S.S.R. Such reasoning has from the start underlain the strategic doctrine of Britain and France. Yet each country has behaved differently in planning for the commitment of its nuclear weapons in the event of a war that, theoretically at least, no one would be foolish enough to start.

In keeping with NATO's command structure, potential targets for nuclear destruction are assigned to Britain by the organization's supreme allied commander in Europe, who has always been an American officer. Because the Polaris missile is not very accurate, the target lists currently handed to the British probably include such vulnerable military installations as airfields, supply depots and troop-marshaling areas. Nevertheless, Britain retains final authority over its own sea-based strategic force and has reserved the right to employ it as the "ultimate guarantor of national security." Toward this end separate target lists are drawn up at Whitehall. In keeping with the notion that the threat of substantial national destruction is the best deterrent, presumably the targets on this British list are key Soviet urban and industrial centers rather than military installations.

Because France's array of nuclear weapons is somewhat more varied than Britain's, it is not surprising that France has opted for a slightly different targeting doctrine. French doctrine stresses the "neutralization of the adversary's administrative, economic and social structures"; it would entail a large number of civilian casualties. Furthermore, France officially refuses to allow inclusion of any of its nuclear forces in NATO war plans, although rumors persist that at least informal consultations to coordinate forces in the event of war have taken place between France and its allies.

Assuming that Britain and France continue to abide by a doctrine that as-



GROWTH OF EUROPEAN STRATEGIC FORCES (those European military forces that are capable of attacking the Soviet homeland from their normal stations) is characterized by a pronounced shift in the vehicles that would deliver nuclear warheads. The forces of Britain (*top*) and France (*bottom*) were initially composed of bomber aircraft, but by 1975 submarine-launched ballistic missiles (SLBM's), which are better able to survive a nuclear strike, dominated the European strategic arsenals. France's force is further distinguished by the deployment of a few land-based intermediate-range ballistic missiles (IRBM's). Both countries intend to augment their forces drastically in the next decade by deploying new ballistic-missile submarines and increasing the number of warheads carried by each SLBM.

signs available RV's or nuclear bombs to major Soviet urban and industrial centers, how would each country's force-modernization program affect the mission of its strategic force? Because the effectiveness of the British and French nuclear deterrents is based on a "credible" threat of massive destruction, we constructed a computer model for calculating the levels of urban and industrial damage that the Anglo-French strategic forces could inflict on the Soviet Union.

Our results indicate that in 1985 France's strategic force could have caused between 23 and 34 million fatalities while destroying between 16 and 25 percent of Soviet production capacity (which includes such facilities as factories, steel mills, oil refineries and chemical plants). By the end of 1990 these figures will increase significantly: potential fatalities could range from 38 to 55 million, and between 25 and 40 percent of Soviet production capacity could be destroyed. The damage would rise to even higher levels in 1995, when France's seventh ballistic-missile submarine is to be commissioned, at which time up to 81 million Soviet citizens could be killed and up to two-thirds of its production capacity could be ruined by a nuclear attack.

The capability of Britain's strategic force to inflict damage on the Soviet Union will also increase over the next decade, although not at the same pace as that of France's force. Applying the same analysis to the British force as was applied to the French (and assuming that all SLBM's carry a six-warhead Chevaline payload), we estimate that the potential casualties from a British nuclear attack would range anywhere from six to 21 million dead; concomitant losses in the production base would range from 5 to 15 percent. Although there will be no changes in the composition of the British force until 1990, these damage levels will probably decline slightly in the meantime owing to improvements in the Moscow ABM system. Assuming each Polaris SLBM carries only three RV's, damage levels would be much lower, amounting to between three and eight million fatalities and about 5 percent of the production base.

With the introduction of Trident II SLBM's into its nuclear force in the 1990's, however, Britain would be able to inflict between 24 and 68 million Soviet fatalities and to incapacitate up to half of the Soviet production base. Given the overwhelming number of warheads in the planned British Trident force, the effectiveness of the Moscow ABM system would have little impact on overall damage levels.

If the U.S.S.R. is dissuaded from at-

tacking Britain or France by the magnitude of the potential toll in human lives and production capability that a nuclear war with either country could exact, then the proliferation of warheads entailed by both countries' force-modernization programs would clearly increase the deterrent effectiveness of their strategic forces. Indeed, the levels of damage approach, if they do not exceed, the "assured destruction" criteria set for U.S. strategic forces in the 1960's.

The increased numbers of warheads available to both nations could also allow either to adopt a revised targeting doctrine. The 1980 British White Paper (equivalent to the U.S. Secretary of Defense's annual Posture Statement) admitted that "there might with changing conditions be more than one way of posing a threat to Soviet state power" and that "some flexibility in contingency planning is appropriate." In other words, warheads could be devoted to targets other than cities, provided enough warheads have been set aside to ensure a deterrent of last recourse based on massive urban and industrial destruction.

In accordance with such a doctrine the British might willingly make available to NATO's limited-target options any warheads beyond those deemed necessary to sustain a credible independent nuclear deterrent. These options present the NATO supreme commander with alternative attack plans for scenarios short of urban-industrial annihilation. In such plans British Trident SLBM's could be launched against the same kinds of targets in the European theater of operations as are assigned to the U.S. Pershing II: military command and control centers, storage sites and bases.

For France the question of limited-target options is a more difficult one. Many French observers have maintained that an increase in the number of warheads would enhance deterrence by making it possible to target a wide spectrum of military installations in addition to urban and industrial centers. They argue that as things now stand the French government has no choice other than to threaten massive destruction in the Soviet homeland—a recourse that most certainly would subject France to a comparable retaliatory attack. Because France would be inviting national suicide if deterrence fails, these observers have argued that war plans should include options far short of an all-out attack on the social fabric of the U.S.S.R.

The French government has consistently rejected these arguments and affirmed the deterrent effect of an "anti-

	DELIVERY VEHICLE	NUMBER DEPLOYED	NUMBER OF WARHEADS CARRIED	WARHEAD YIELD (KILOTONS)	YEAR FIRST DEPLOYED
BRITAIN	POLARIS (CHEVALINE) SLBM	64	3-6	40	1982
	TRIDENT II SLBM	64	8*	475?	early 1990's
FRANCE	S-3 IRBM	18	1	1,000	1980
	M-20 SLBM	80	1	1,000	1977
	M-4 SLBM	16	6*	150	1985
	MIRAGE IVA BOMBER	18	1	150	1964
	S-3 IRBM	18	1	1,000	1980
	M-20 SLBM	16	1	1,000	1977
	M-4 SLBM	80	6*	150	1985
	M-5 SLBM	16?	≥6*	?	1995
	S-X IRBM?	?	?	?	?

NUCLEAR DELIVERY VEHICLES in the current strategic arsenals of Britain and France are printed in black in this table; those expected to be in the arsenals a decade from now are printed in color. All the SLBM's in the British arsenal were originally developed and manufactured in the U.S., although the warhead-dispensing mechanism on the Chevaline version of the Polaris missile and on the Trident II is of British design. An asterisk indicates that a missile's warheads are contained in multiple independently targetable reentry vehicles, so that the warheads of a single missile can strike widely separated targets.

cities" nuclear posture. As the strategic nuclear force expands over the next decade, however, it is conceivable that the government might reconsider its nuclear doctrine.

Although each country's capacity to inflict damage on the Soviet Union will increase over the next decade, a combined European nuclear deterrent based on Anglo-French cooperation could be even more militarily effective. Because the damage that can be inflicted on urban and industrial areas quickly reaches a point of saturation, the main advantages of a joint force would have little to do with overall damage levels. Rather, a combined Anglo-French force would allow greater targeting flexibility because there would be a larger pool of warheads to choose from; the combined force would be able to keep greater numbers of warheads in reserve after initial strikes and, because of its greater dispersal, it would increase Western European confidence in its ability to retaliate.

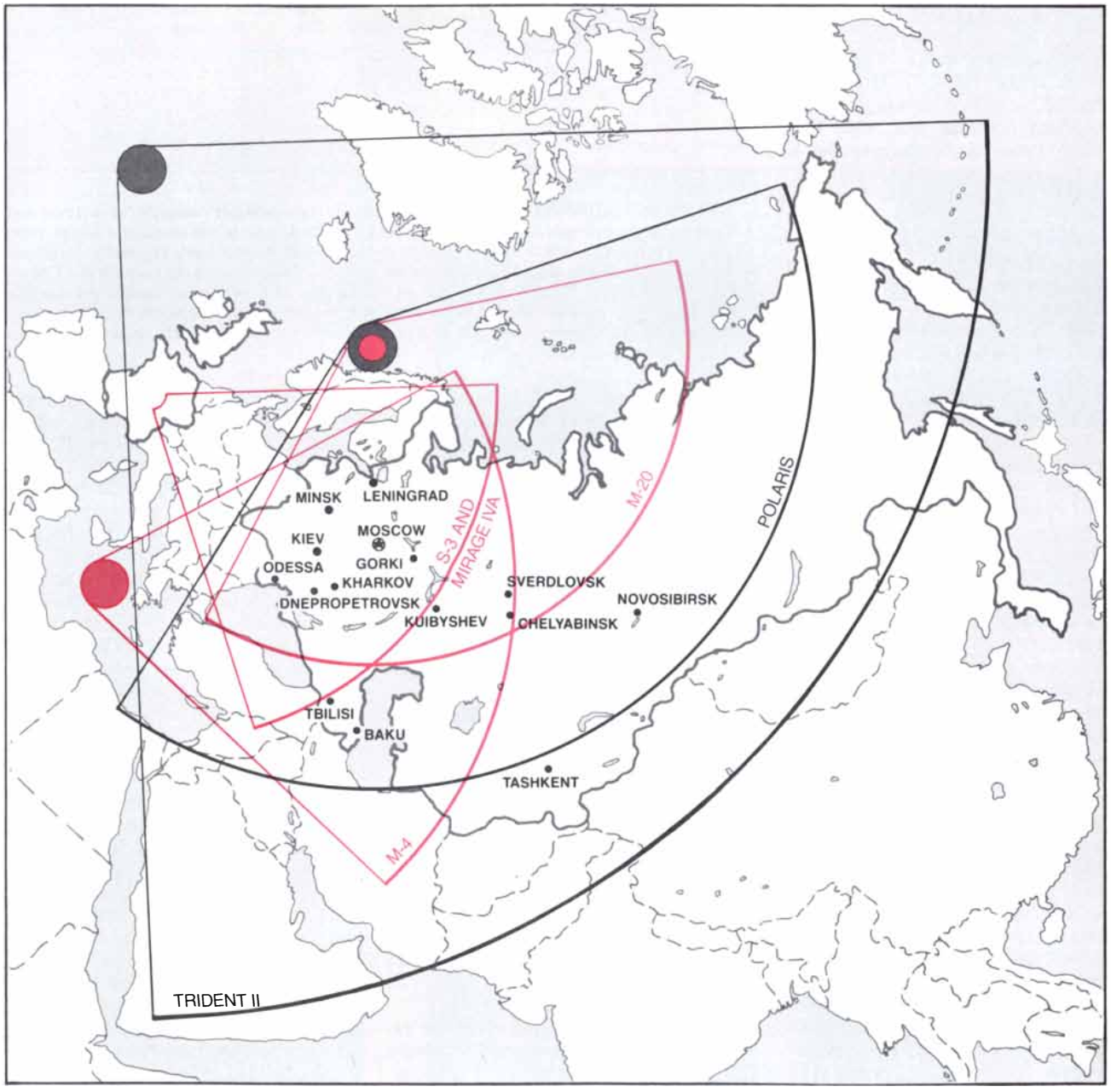
From the Soviet perspective the expansion of European strategic forces has broad and serious implications. Certainly the adoption of limited-target options by the British or French would severely complicate Soviet military planning, particularly that related to land battles in Europe. Yet it is the sheer destructive potential wielded by the European nuclear powers that presents the worst possible scenario to the U.S.S.R.: a nuclear attack by a third nuclear power could seriously cripple the Soviet Union, leaving it extremely vulnerable to a subsequent U.S. attack. To forestall such a scenario the U.S.S.R. might consider deploying a nationwide ballistic-missile defense. Up to this point we have been assuming that the U.S.S.R. could count on a maximum of 100 ABM-interceptor launchers around Moscow (as permitted under the ABM Treaty of 1972) to defend itself against a strategic nuclear attack. Could an expansion of the Soviet ABM system to cover other urban and industrial areas of the Soviet Union severely hinder the effectiveness of the European nuclear forces?

If we assume that the Soviets abrogate the ABM Treaty in the late 1980's or early 1990's and deploy up to 1,000 ABM launchers by 1995, the damage European forces could inflict on the Soviet Union might be significantly reduced. This would be the case only if the Soviet defense system were highly reliable: if each ABM interceptor were to have at least a 60 percent probability of destroying an incoming RV, and if there were to be a large number of ABM-interceptor launchers. A less capable defense would minimize only slightly the destruction the European nuclear forces could wreak.

For a combined Anglo-French nuclear-attack scenario, damage potential begins to decline only as the size of the Soviet missile defense approaches 1,500 deployed interceptors. Such a defense would require a large-scale and very costly Soviet ABM program. Hence, according to our calculations,

as long as no more than a moderately large number of interceptors (fewer than 800) are deployed in an upgraded Soviet ABM system, and as long as the interceptors are no more than moderately capable (less than 60 percent effective), the British and French nuclear forces can continue to be fairly reliable in their respective deterrent roles into the 1990's.

Soviet deployment of even a moderately large missile-defense system of limited effectiveness would nonetheless be likely to affect European target planning. British and French defense planners might feel compelled to expend large numbers of warheads on a smaller number of targets to ensure that at least some warheads would penetrate the defenses protecting key targets. That might preclude the adoption of strategies based on limited-target options, since the entire inventory of nuclear weapons might be required to execute an urban and industrial attack. A large-scale Soviet missile defense could also have a broad impact on Britain and France by introducing greater uncertainty as to the amount of damage that their relatively modest



APPROXIMATE RANGES of British (*black*) and French (*color*) missiles and aircraft determine which cities in the U.S.S.R. could be targeted by which delivery vehicles. The presumed launch areas for SLBM's are shown as disks. This information was used for calculating the loss in lives and in production capacity that would be inflicted on the U.S.S.R. by a hypothetical European nuclear attack (*see illustrations on opposite page and page 41*). Each target city within range of a delivery vehicle was allocated one warhead of the type carried by the vehicle; any warheads remaining after this initial distribution were apportioned to the cities according to each city's size. It was assumed that warheads would be detonated at altitudes maximizing damage from the blast of the nuclear explosion. The area subject to blast overpressures equal to or greater than five pounds per square inch was taken as the unit of measurement: it was assumed that all people within the area would be killed and all buildings would be destroyed. A target city was deemed to be "saturated" when all its urban area was subjected to the lethal overpressure. It was assumed that an anti-ballistic-missile (ABM) system protecting a target city could be countered by allocating more warheads to compensate for the ABM system's effectiveness.

strategic forces could inflict in a massive attack.

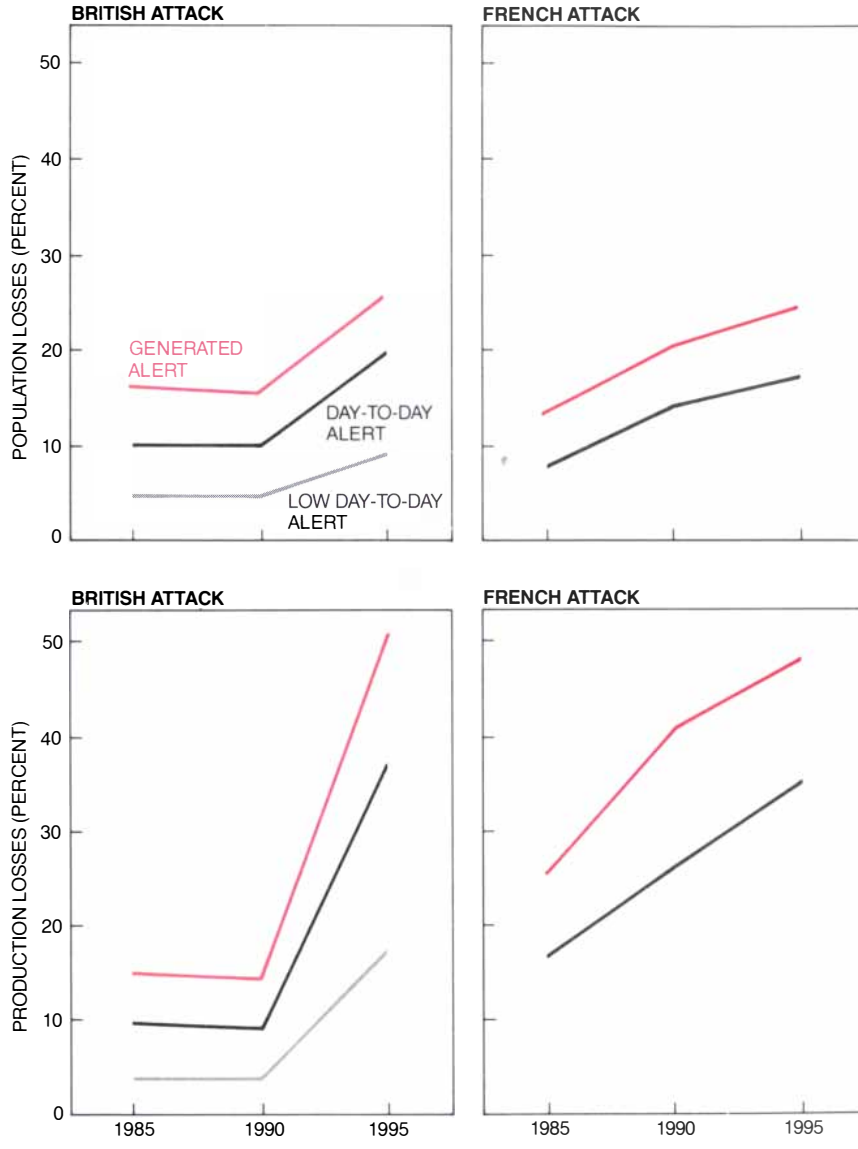
The effectiveness attributed to a Soviet ABM interceptor is a critical factor in our computer model, and we have probably overestimated it in our calculations. More realistic assumptions about the performance of a Soviet ABM system, of course, would mean that the Soviets would need to deploy even more ABM-interceptor launchers to destroy the same number of incoming RV's. Moreover, in the likelier crisis scenarios, where forces have been placed on generated alert (in which almost all nuclear delivery systems are on station, standing by to launch an attack), very large numbers of ABM launchers are needed before a Soviet ABM system could have a significant impact. Hence, even after a substantial investment, an augmented ABM system might still fall short of acquiring practical military utility against the European forces.

What is perhaps the most critical assumption in our computer model has nothing to do with the ABM hardware; it has to do with the Soviet strategy for engaging incoming RV's. In our analysis we assumed that multiple interception attempts could be made against each incoming RV. In other words, it is assumed that if an ABM interceptor misses its target RV, other interceptors would be fired at the RV until it is destroyed or until all interceptors are expended.

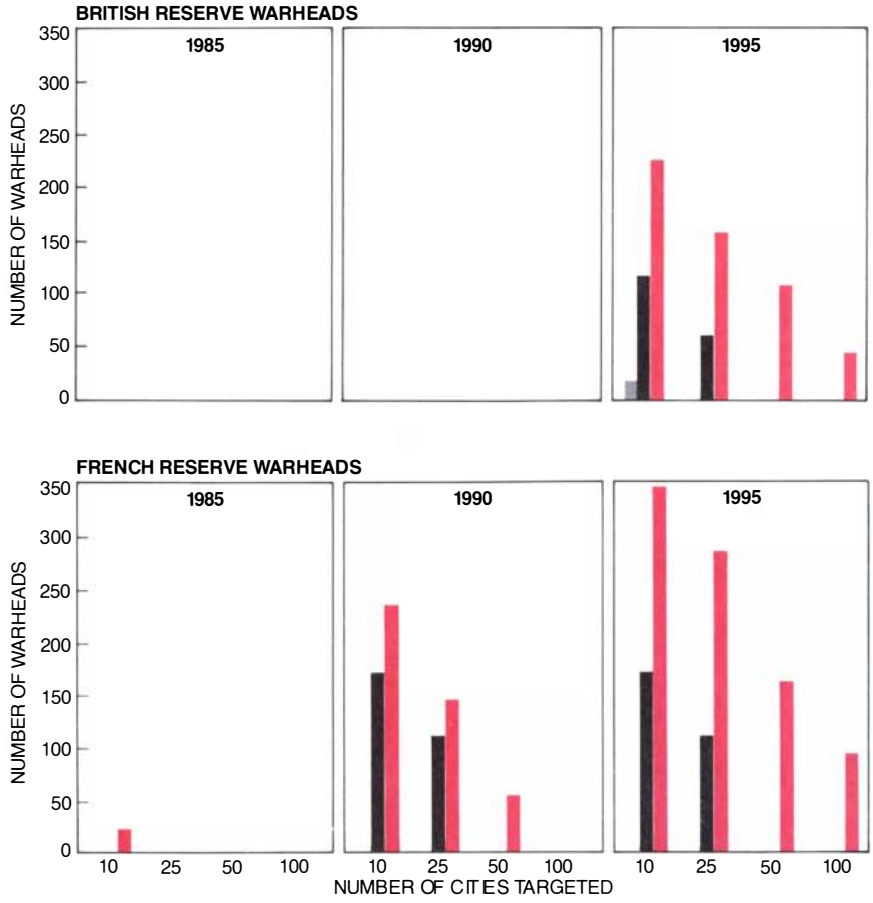
Yet in the event of a nuclear war Soviet strategists would probably consider any European attack as a precursor to an American one, and they would be reluctant to exhaust their defensive potential against the first and less powerful attacker. If the assumed rules of RV engagement are altered to allow only one interception attempt against each European RV, the performance of the Soviet missile defense declines precipitously. For example, against a 1995 British attack in a state of low day-to-day alert (when only one submarine is on station and ready to launch its missiles), a 500-launcher ABM system that allows multiple interception attempts reduces the postulated immediate Soviet fatalities to only 309,000. The identical attack and defensive deployment, allowing this time for only a single chance at interception, would lead to a projection of some 13 million fatalities.

In sum, although a Soviet ABM system could pose a challenge to European nuclear forces, the mere existence of third-country nuclear forces complicates the operation of any Soviet ballistic-missile defense. In order to maintain an effective defense the Soviet ABM system must be sufficiently large and robust to absorb all third-country attacks while retaining the capability to meet any subsequent American attack. A defense of this magnitude would be an extremely difficult and costly undertaking. Clearly a less costly option for the U.S.S.R. might be found at the negotiating table.

Since the late 1960's, however, Britain, France and the U.S. have been united in excluding European nuclear forces from negotiated constraints. Every Soviet attempt to place British and French forces with American ones on one side of the scale when balancing Western and Soviet strategic forces has been rejected by the U.S. as a thinly veiled attempt to legitimize Soviet strategic superiority over the U.S. Even in discussions more pertinent to the Europeans themselves, such as the recent negotiations on intermediate-range nuclear forces (INF), Britain and France, supported by the U.S., have in-



THREAT OF DESTRUCTION to Soviet society, with which Britain and France aim to deter aggression against them by the U.S.S.R., will increase markedly as force-modernization programs are implemented. Potential fatalities resulting from a European nuclear attack on major urban and industrial centers are plotted as a percent of the Soviet population; ruined production facilities are plotted as a percent of total Soviet manufacturing capacity. Various levels of readiness can be attributed to the attacking forces: generated alert (*color*), in which almost all forces have been ordered to stand by for an attack; day-to-day alert (*black*), in which aircraft, submarines or missiles in routine overhaul are excluded, and, in the case of the British forces, low day-to-day alert (*gray*), in which only one submarine is assumed to be ready to fire its missiles. The calculations do not take any Soviet civil-defense measures into account. Only Moscow is credited with an ABM system, each interceptor of which was given a 30 percent chance of stopping an incoming reentry vehicle.



RESERVE WARHEADS are those left over after enough warheads have been allocated to ensure the saturation of all target cities. The number of reserve warheads depends on whether the largest 10, 25, 50 or 100 Soviet cities have been designated as targets and whether the forces attacking the cities are in a state of generated alert (*color*), day-to-day alert (*black*) or low day-to-day alert (*gray*). Already by 1985 French forces in generated alert could have had a few warheads left over after effectively destroying the 10 largest Soviet cities. Britain would have to wait until 1995, when the Trident II SLBM is deployed, before counting on any extra warheads after a massive nuclear attack on the U.S.S.R.

sisted on remaining exempt from any limits that might be placed on American nuclear weapons based in Europe.

The U.S., Britain and France have advanced a number of arguments for this exclusion of British and French forces from any limitations, some of which are not entirely convincing. For instance, the American assertion that these forces are not meant to deter attacks against other alliance members may be true for France, but it is not true for Britain. British nuclear weapons are normally considered part of NATO's forces; only in the most extreme circumstances would Britain assume unilateral command of its nuclear forces.

The Western assertion that the European forces are not "theater" forces but rather are "strategic" and ought not to be considered in the INF forum contradicts an earlier stance in the Strategic Arms Limitation Talks (SALT), in which the American delegation maintained the reverse. Finally, the Western allies have argued that the European forces are too small to be considered a threat. This may be true if they are compared in terms of sheer numbers with the entire Soviet strategic force. It certainly is not true if they are compared with Soviet theater forces, however. Moreover, numerical comparisons evade the real point, namely that European forces as they now stand could inflict unprecedented damage on the Soviet Union.

It is not clear whether Western arguments will enable the two European nuclear powers to avoid inclusion in arms-control negotiations much longer. As we have stated above, given their planned force modernizations, the combined British and French forces would in fact account for a significant percentage of the Soviet strategic force. Any new bilateral strategic arms-control agreement would only worsen the Soviet Union's position vis-à-vis Europe. For example, a strategic-arms agreement in accordance with the current U.S. proposal would limit Soviet strategic forces to about 7,000 deliverable warheads and bombs. In such a case the forces available to an unconstrained Britain and France in 1995 might equal about 16 percent of the Soviet total.

The increasing ability of the European forces to inflict severe damage on the Soviet Union, not to mention the problems they present for Soviet defense planners, is almost certain to result in continued Soviet pressure to include the British and French forces in any new arms-control agreement. To the extent that such treaties are considered beneficial to international security in general, a solution to the third-nuclear-power issue may have to be found sooner rather than later.

Ironically, it might well be in the interest of the U.S. and its allies to submit to the long-standing Soviet wish that European forces be included in the arms-control process. Although the modernization of European forces will no doubt enhance their contribution to the deterrence of a Soviet attack, uncontrolled expansion of European nuclear arsenals may lower the threshold of confrontation at which Britain and France would fire their nuclear weapons. In such a case the U.S. could conceivably be dragged into a full-scale nuclear war as the result of a situation not of its own making. For example, the Soviet Union may not be able to tell a European SLBM from an American one. Consequently, if one of the European nuclear powers were to launch an SLBM toward the Soviet Union early in a crisis, the Soviet Union might respond with a retaliatory attack indiscriminately aimed at Western forces, including those of the U.S.

Furthermore, a move toward greater autonomy in nuclear matters on the part of Britain and France has a multitude of political overtones that reverberate to the very core of U.S.-European relations. An increasingly independent European defense based on the threatened use of nuclear weapons against the Soviet homeland probably would be detrimental to the goal of a politically and militarily unified Western alliance. In this context, refusal by Britain and France to take part in the arms-limitation process could be viewed as a further sign of diminishing European confidence in the American nuclear guarantee and, by extension, in the NATO alliance as a whole. If the alliance is to remain as close as it has been in the past, it may be wise to encourage the integration of European forces into the arms-control process.

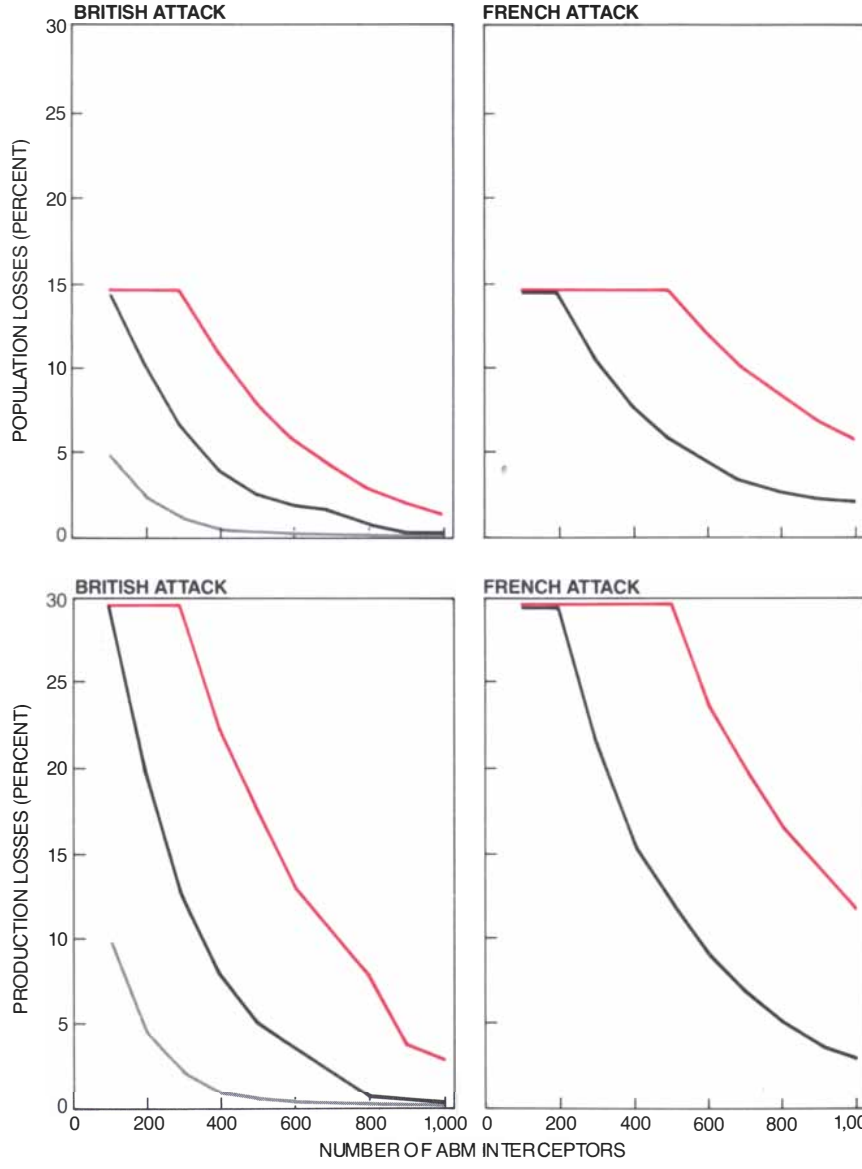
Inclusion of the European strategic

forces in agreements could take any of several forms. One simple solution would be to place equal ceilings on the total number of nuclear delivery systems and warheads of the Soviet Union and of NATO. Such an approach would probably be politically unacceptable in Europe and the U.S. Alternatively, the British and French could each negotiate bilaterally with the U.S.S.R. for protocols or separate treaties linked to a comprehensive Soviet-U.S. agreement. This would be a straightforward approach to Anglo-French participation in arms-control negotiations. Such separate limits on the British and French could help to lay the groundwork for future multilateral negotiations that might limit nuclear forces according to certain assigned ratios.

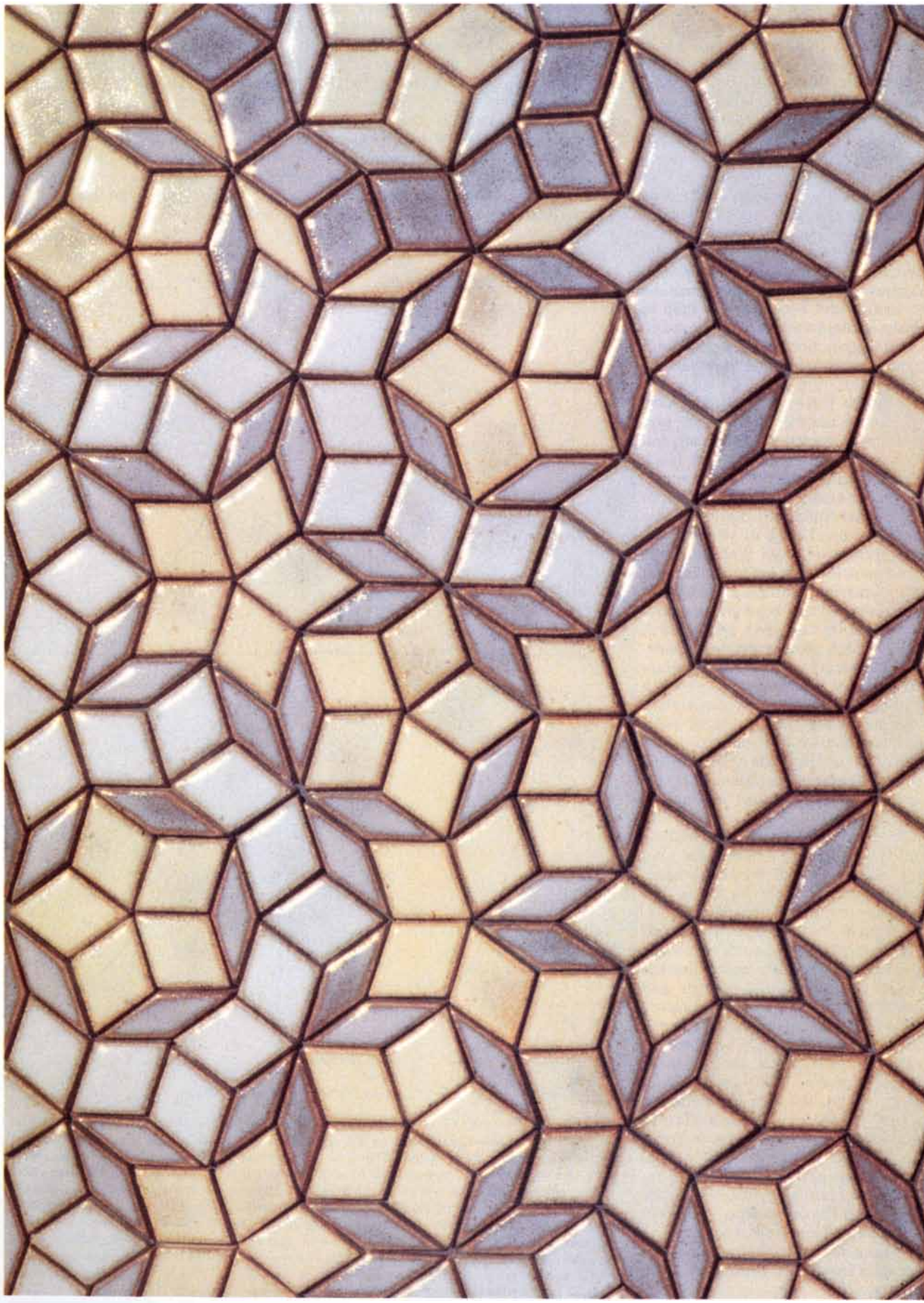
Although Britain and France have expressed a number of stringent preconditions for participation in any nuclear-arms-control discussions (such as a significant reduction of the superpowers' arsenals), securing European cooperation could ultimately hinge on how the critical issue of strategic defense is resolved. Both Western European countries have made it clear that before entering strategic-arms negotiations they want reassurances that constraints on defensive systems will be given high priority on the agenda. They are afraid that the SDI program and its Soviet counterpart, as well as a budding antisatellite arms race, threaten to undermine the ABM limits existing under the SALT regime. The ultimate result could be abrogation of the ABM treaty and the deployment of extensive strategic defenses in the U.S. and the U.S.S.R.

Clearly Britain and France would be relatively better off if the status quo, which remains almost free of ballistic-missile defenses, were preserved. Although expansion of the current Galosh system could limit British and French targeting plans, it probably would not significantly undermine the deterrent effect of the British and French nuclear forces. A 21st-century Soviet ABM system could undermine it, however. Furthermore, a 21st-century American ABM system such as those being proposed under SDI could decouple European defenses from that of the U.S. In spite of American offers to extend any future defensive umbrella over Europe and in spite of official British participation in SDI-related research and French enthusiasm for the multilateral Eureka defense-technology research effort, the European allies have been lobbying the superpowers to refrain from a long-term defensive buildup.

It is not clear how the U.S. will treat this issue in future arms-control negotiations. Unless the U.S. concedes some limitations on defensive weaponry, it is unlikely that the U.S.S.R. will agree to limit its offensive nuclear weapons; the issue of British and French participation in arms control will remain moot. On the other hand, an American offer that reinvigorates limitations on defensive systems and accommodates some limitations on antisatellite weapons might induce the U.S.S.R. to cut its offensive strategic forces. Under such circumstances it would be propitious for Britain and France to take part in the arms-control process, particularly if U.S. concessions on defensive limits were portrayed, in part, as an effort to take allied concerns into account. The participation of Britain and France could help to seal a formal agreement ensuring the continued viability of their independent deterrent.



NATIONWIDE SOVIET ABM DEFENSE based on 20th-century technology could reduce the level of damage inflicted by European forces, but it would not significantly hinder the effective destruction of the 25 largest Soviet cities if the attacking European forces were in a condition of generated (*color*) or day-to-day (*black*) alert unless more than about 800 ABM interceptors were deployed. A British nuclear attack initiated under conditions of low day-to-day alert (*gray*) could be blunted by about 300 interceptors. For these graphs it was assumed that each ABM interceptor would be 60 percent effective, an assumption that probably exaggerates its likely effectiveness. A 21st-century ballistic-missile defense could be substantially more effective than a 20th-century one, and it could pose a more serious threat to the independent nuclear deterrent of Britain and France. Negotiating an arms-control treaty with the U.S.S.R. might be the best way to prevent deployment of a Soviet ABM system that undermines the European deterrent in this century as well as in the next.



Quasicrystals

These newly discovered materials embody a novel kind of order, intermediate between crystalline and amorphous. Their structure can be understood through the mathematical theory of tiling

by David R. Nelson

In 1984 investigators working at the National Bureau of Standards found a material that seemed to violate one of the oldest and most fundamental theorems of crystallography. The material appeared to have the same kind of order that is inherent in a crystal, but it also appeared to be symmetrical in ways that are physically impossible for any crystalline substance. Further investigations into the microstructure of this material have shown that it embodies a new kind of order, neither crystalline nor completely amorphous. Materials structured around this new kind of order seem to forge a link between conventional crystals and the materials called metallic glasses, which are solids formed when molten metals are frozen so rapidly that their constituent atoms have no time to form a crystalline lattice. The new materials have therefore been called quasicrystals.

A conventional crystal is a particularly well-ordered arrangement of atoms or molecules. It is a latticework, in which identical "unit cells"—the building blocks of the crystal, each of which contains precisely the same distribution of atoms—fit together regularly and periodically to fill space. Inherent in every crystal structure are certain symmetries. For example, a crystal is said to have threefold rotational symmetry if the latticework of the crystal would look precisely the same after the crystal was rotated through one-third of a full circle, or 120 degrees. (An elementary example of a shape that has threefold rotational symmetry is the equilateral triangle.) Crystals might also have fourfold or sixfold rotational symmetry (as do the square and the hexagon respectively). A crystal can never have fivefold rotational symmetry, however, for much the same reason that it is impossible to tile a plane—to cover it completely with shapes that do not overlap—using only shapes that have fivefold symmetry, such as pentagons.

It was therefore a tremendous surprise to most crystallographers and condensed-matter physicists when Dany Shechtman, Ilan Blech, Denis Gratias and John W. Cahn reported the crystallographic properties of a particular rapidly cooled sample of an aluminum-manganese alloy. It had scattered a beam of electrons onto a photographic emulsion in a way that formed a sharply defined pattern displaying fivefold symmetry. The sharp definition of the pattern indicated that the material had long-range order, because it meant that atoms in many parts of the sample had reflected the electron beam in the same way. Yet the symmetry of the pattern, which was apparent no matter where in the sample the beam was aimed, indicated that the material's underlying structure had some kind of fivefold symmetry. Other investigators have found that these same properties can be induced in many metals besides aluminum and manganese.

Further analysis of the scattering pattern has revealed the origin of these remarkable properties. The unit-cell structure of most crystals is based on such Platonic solids as the cube, the tetrahedron and the octahedron. These new materials (which as a class I shall call shechtmanite, following a suggestion by Cahn) have a structure based instead on another Platonic solid, the icosahedron: a figure that has 20 faces, each of them an equilateral triangle. Until this discovery there had been almost a conspiracy of silence about the icosahedron in the physics literature. Many respected physics texts state flatly that the icosahedron is "of no physical interest." The reason for such skepticism is that the icosahedron has a fivefold symmetry—five faces meet at each of its vertexes—and so it cannot serve as the unit cell of any conventional crystal.

How has shechtmanite managed to find a way around a universally recognized theorem of crystallography? The answer to this question has come from a study of the microscopic structure of rapidly cooled metal alloys and from the mathematical theory of tiles and tiling.

In order to understand the underlying structure of shechtmanite quasicrystals it is first necessary to have some understanding of the structure of a conventional crystal. Conventional crystals possess two kinds of long-range order, known respectively as orientational order and translational order. These can be seen in one of the simplest crystalline arrangements of

PENROSE TILING is a two-dimensional analogue of newly discovered materials called quasicrystals or shechtmanite. The pattern is not periodic (it cannot be broken down into a single unit cell that repeats indefinitely), but it has certain kinds of order that are usually associated with periodic tilings. For example, the decagons (10-sided, nearly circular figures) found throughout the tiling all have the same orientation (the sides of one decagon are parallel to the sides of the others) just as the unit cells of a periodic tiling all have the same orientation. Unlike any possible periodic tiling, however, a Penrose tiling has a kind of fivefold rotational symmetry: there is a sense in which the pattern is unchanged when it is rotated through one-fifth of a full circle, or 72 degrees. Like Penrose tilings, the microscopic structure of quasicrystals also has a fivefold symmetry. Fivefold symmetry is impossible in any conventional crystalline material (any material built up of a single unit cell that repeats periodically), and so shechtmanite, although it is highly ordered, cannot be a conventional crystal. Penrose tilings were invented by Roger Penrose of the University of Oxford. Tiles here were manufactured by Saxe Patterson (Taos Clay Products) of Taos, N.M.

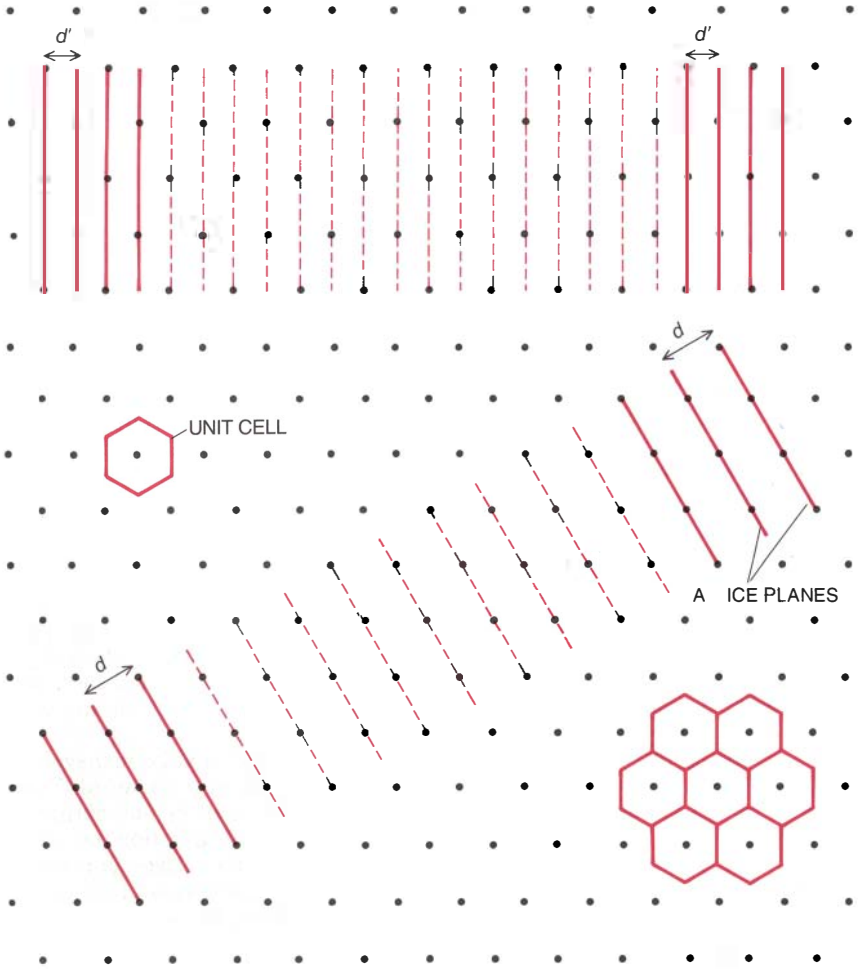
atoms: the triangular lattice, which is the configuration assumed by billiard balls when they are racked up at the start of a game [*see illustration below*].

In this two-dimensional lattice every atom sits in a hexagonal cage formed by its six near neighbors. This hexagon, with an atom at its center, is a unit cell of the crystal: the crystal can be broken down into a repeating pattern of hexagons. Because all the hexagons have the same orientation—that is, because the sides of each hexagon are parallel to the sides of all the others—the crystal is said to exhibit long-range orientational order.

The other kind of long-range order present in a crystal can be demonstrated by drawing a family of parallel lines on the lattice. When the lines are drawn so that every atom lies on one line or another, the lines will be spaced exactly evenly across the crystal. It is therefore possible to translate, or move, a few lines from one part of a crystal so that they match up precisely with the lines in another part of the crystal. Thus by examining any small part of the crystal one can determine the exact position and spacing of the lines in any other part: the crystal has long-range translational order.

In a conventional crystal there are many families of parallel lines, each family facing in a different direction; the spacing between the lines may vary from family to family. In a three-dimensional crystal the lines are actually planes and are called lattice planes. When beams of X rays or electrons are directed at a crystal, they are reflected and scattered by the lattice planes. By studying the directions in which the beams are scattered and the intensity of each scattered beam, investigators can determine which families of lattice planes must exist in the crystal; often they can deduce the precise locations of atoms. It was by this method that shechtmanite was found. The sharp pattern Shechtman and his colleagues noted indicated their sample had lattice planes, but the pattern's fivefold symmetry indicated the material could not be a conventional crystal.

To understand why a conventional crystal cannot have fivefold symmetry, one need only consider an attempt to fill the plane with pentagonal, rather than hexagonal, cells. Regular pentagons, the simplest shapes that have fivefold symmetry, cannot serve as the unit cells of a crystal. The reason is that pentagonal cells, in contrast to hexagons, cannot fit tightly together without leaving a space: if two pentagons are joined along matching parallel faces, a third cannot join them and fit tightly with both its neighbors. This misfit is an example of what condensed-matter physicists call frustration. In three dimensions as well as in two, it is impossible to pack shapes that have fivefold symmetry without some degree of frustration.



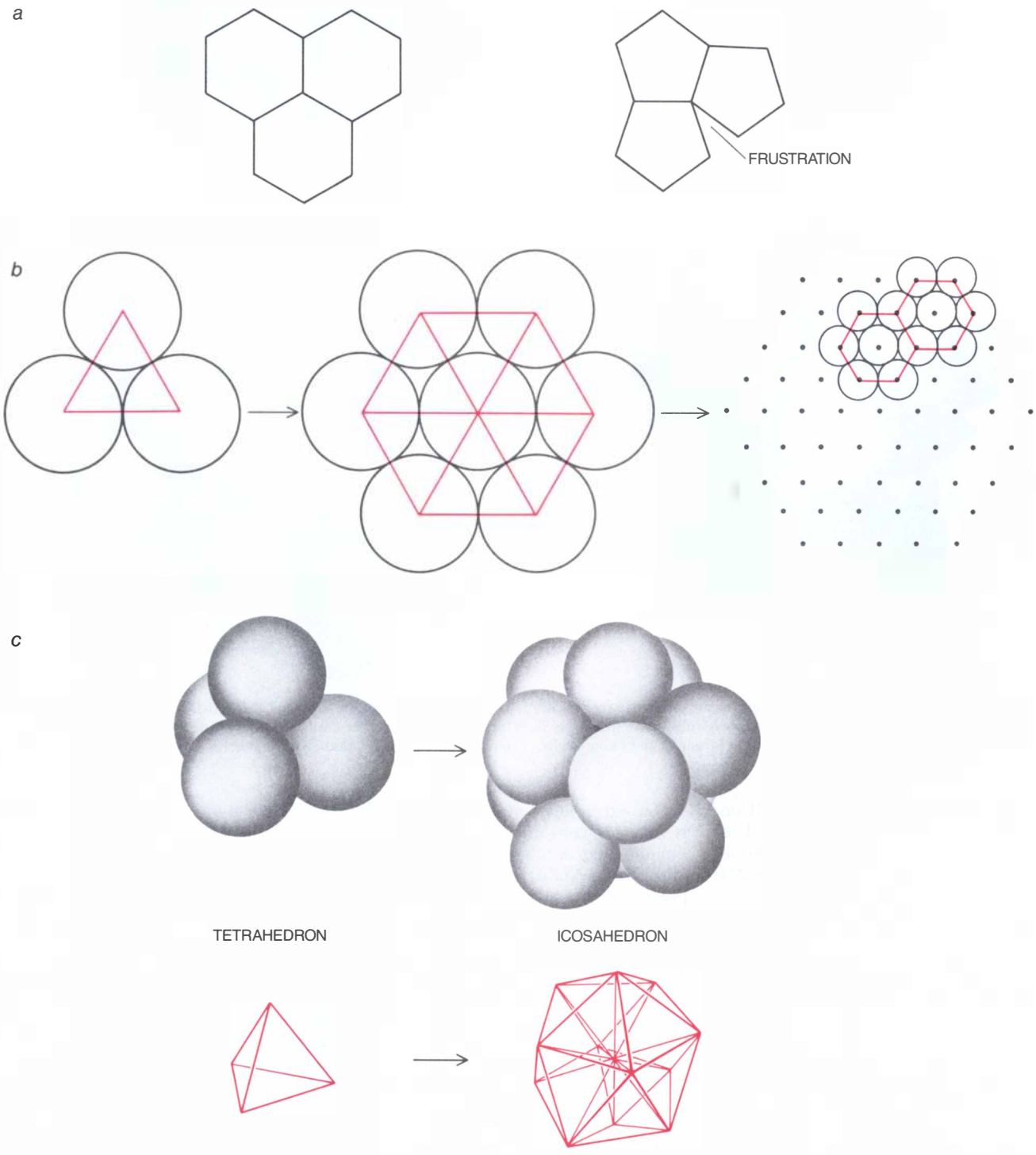
PERIODIC LATTICE illustrates two kinds of long-range order that are inherent in conventional crystals. The pattern can be broken down into a number of hexagons, each with an "atom" at its center. Because the hexagons in one part of the crystal (*lower right*) have the same orientation as those in any other part (*left center*), the lattice is said to have "long-range orientational order." The crystal's "long-range translational order" is demonstrated by the family of parallel lines that runs from the lower left to the right center of the lattice. When the lines, which are called lattice planes in a three-dimensional crystal, are drawn so that every atom is on one line or another, the distance between two adjacent lines is exactly the same throughout the crystal. The position and direction of the lines in any part of the crystal could therefore be determined by translating, or moving, lines from any other part of the crystal. There are many families of parallel lines; another is drawn across the top of the lattice. The distance between adjacent lines may vary from family to family.

In two dimensions the hexagon is the most efficient packing unit: three disks fit tightly together to form a triangle and six such triangles combine to form a hexagon consisting of seven disks. Remarkably, a similar line of reasoning indicates that the icosahedron (which, because of its fivefold symmetry, cannot be the fundamental unit of a regular packing) should actually be the natural packing element in three dimensions: four hard spheres, which might represent atoms, fit tightly together to form a tetrahedron (the "pyramid" shape in which cannonballs are usually stacked at battlefield monuments) and, with small distortions, 20 such tetrahedrons fit together to form an icosahedron consisting of 13 atoms.

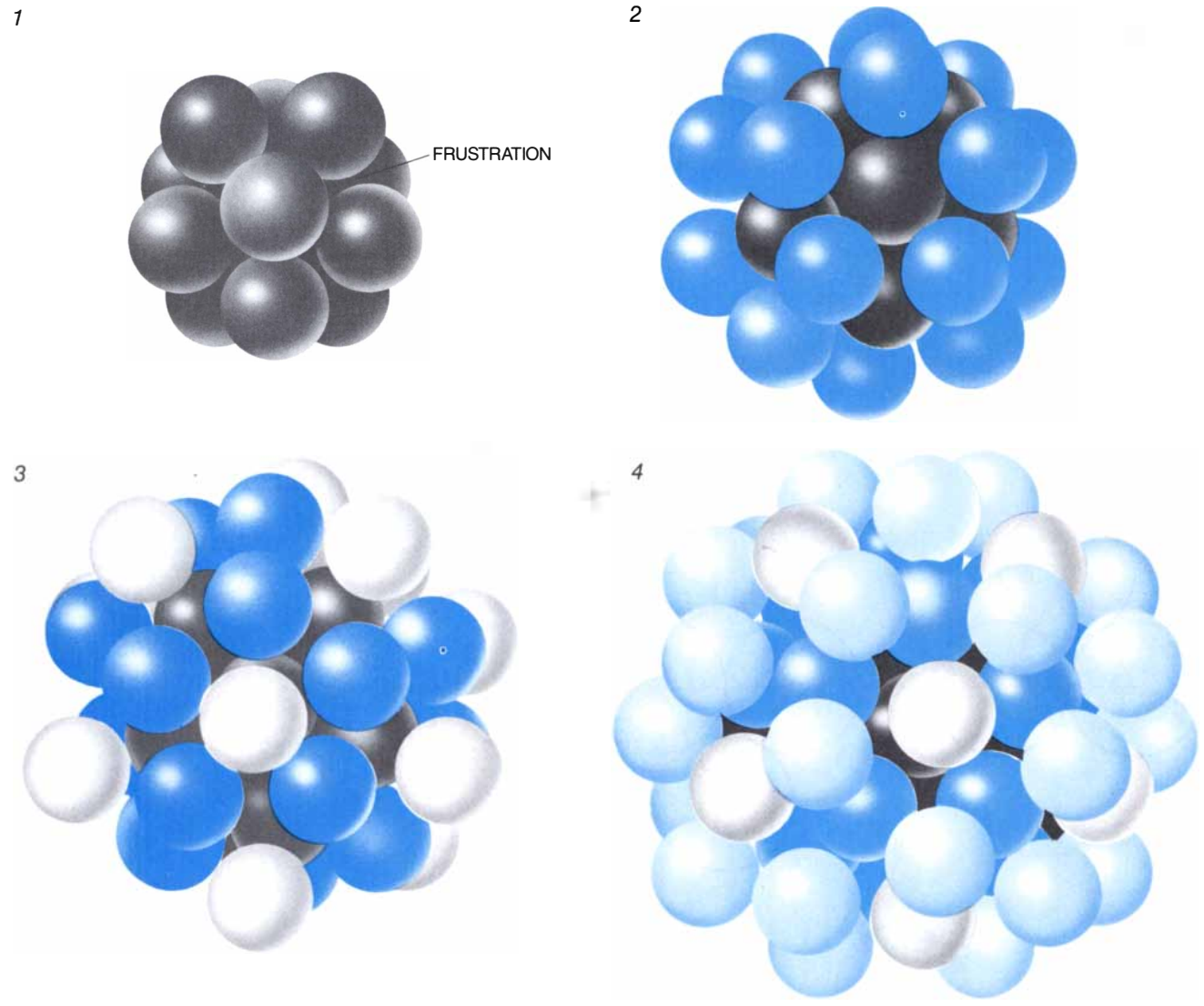
What happens, then, when one tries to pack spheres in a formation modeled on an icosahedron? To begin with, the 13-atom icosahedron is not perfect. It is slightly frustrated, because there are small cracks between the 12 atoms arranged on its surface; each atom is about 5 percent farther from its neighbors on the surface than it is from the central atom. Like pentagons arranged on a plane, the atoms cannot simultaneously sit at identical distances from all their near neighbors. Adding shells of particles only aggravates the frustration; the cracks

between atoms in the outer shells become progressively larger [*see illustration on next page*].

Because of this increasing degree of frustration, icosahedral order cannot propagate throughout a crystal. Over short distances, however, the icosahedral packing is very efficient. Following this line of reasoning, Sir Charles Frank of the University of Bristol suggested in 1952 that small regions of icosahedral symmetry should be quite prevalent in dense undercooled liquids (liquids that have been cooled below their freezing point). In such liquids, he conjectured, there might be many regions in which small clusters of atoms have nucleated to form icosahedrons. It was later realized that if such an undercooled liquid were to form a glass before crystallizing, the glass



TIGHT PACKING is not possible using only shapes that have fivefold symmetry. Three hexagons can pack tightly together without leaving a crack, but three pentagons cannot (*a*); the misfit is called frustration. In two dimensions (*b*) the hexagon is the most efficient packing unit: three disks fit tightly to form a triangle, six triangles fit tightly to form a hexagon and the hexagons fit tightly to cover a plane. In three dimensions (*c*) four spheres pack tightly to form a tetrahedron and, with small distortions, 20 tetrahedrons fit tightly into an icosahedron. Icosahedrons, however, have fivefold symmetry (five triangular faces meet at each vertex), and they cannot fit together tightly. Because icosahedrons cannot pack in a way that fills space completely, they cannot serve as a crystal's unit cells.



ICOSAHEDRAL CLUSTER of atoms is an efficient way to fill space over short distances, but it is progressively less efficient as it grows: cracks between adjacent atoms are larger in the outer shells. In some alloys, such as $Mg_{32}(Al,Zn)_{49}$ (which can form shechtmanite), this frustration is relieved (not shown) by forcing additional atoms into the cracks that first appear in part *3* of the illustration.

might consist of a number of small icosahedral clusters somehow linked together. The atoms of the glass would then exhibit short-range icosahedral order in many very small regions.

Frank's prediction seems to have been correct. In the early 1960's investigators first became able to cool metals rapidly enough to solidify them before they could crystallize. In the late 1960's and early 1970's it was found that the atomic positions in these glasses are modeled very well by certain disorderly groupings of hard spheres. The groupings contain many fragments of icosahedrons, and so short-range icosahedral order does indeed appear to be an important structural element in rapidly cooled liquids and metallic glasses [see "Metallic Glasses," by Praveen Chaudhari, Bill C. Giessen and David Turnbull; SCIENTIFIC AMERICAN, April, 1980].

One extremely simple metallic glass is formed from an alloy of magnesium and zinc. A related compound, the alloy $Mg_{32}(Al,Zn)_{49}$, is especially important in the study of quasicrystals. This is one of the alloys that, when they are cooled rapidly enough, form shechtmanite. When they are cooled even more rapidly, alloys of this kind often form glasses containing many small regions of icosahedral symmetry. When $Mg_{32}(Al,Zn)_{49}$ is cooled more slowly, it forms a crystal in which small regions still have icosahedral symmetry, although the crystal itself is not a pure packing of icosahedrons. The icosahedrons are packed, with slight distortions, into a pattern that has an overall cubic symmetry; the forbidden long-range icosahedral symmetry is therefore absent. It is likely that the short-range icosahedral symmetries found in such glassy and crystalline states are somehow related to the long-range icosahedral symmetry of shechtmanite.

In crystalline $Mg_{32}(Al,Zn)_{49}$ the natural frustration of short-range icosahedral packings is relieved by forcing additional atoms into the cracks between atoms in the icosahedral shells, thereby disturbing the symmetry of the clusters. The resulting dense configuration of particles can be seen as a number of slightly distorted tetrahedrons with atoms at their vertexes; the bonds between atoms are the edges of the tetrahedrons. Tetrahedrons meet at common faces, and each edge, or bond, is shared by several tetrahedrons [*see illustration on opposite page*]. Both metallic glasses and crystalline alloys can be analyzed in this way.

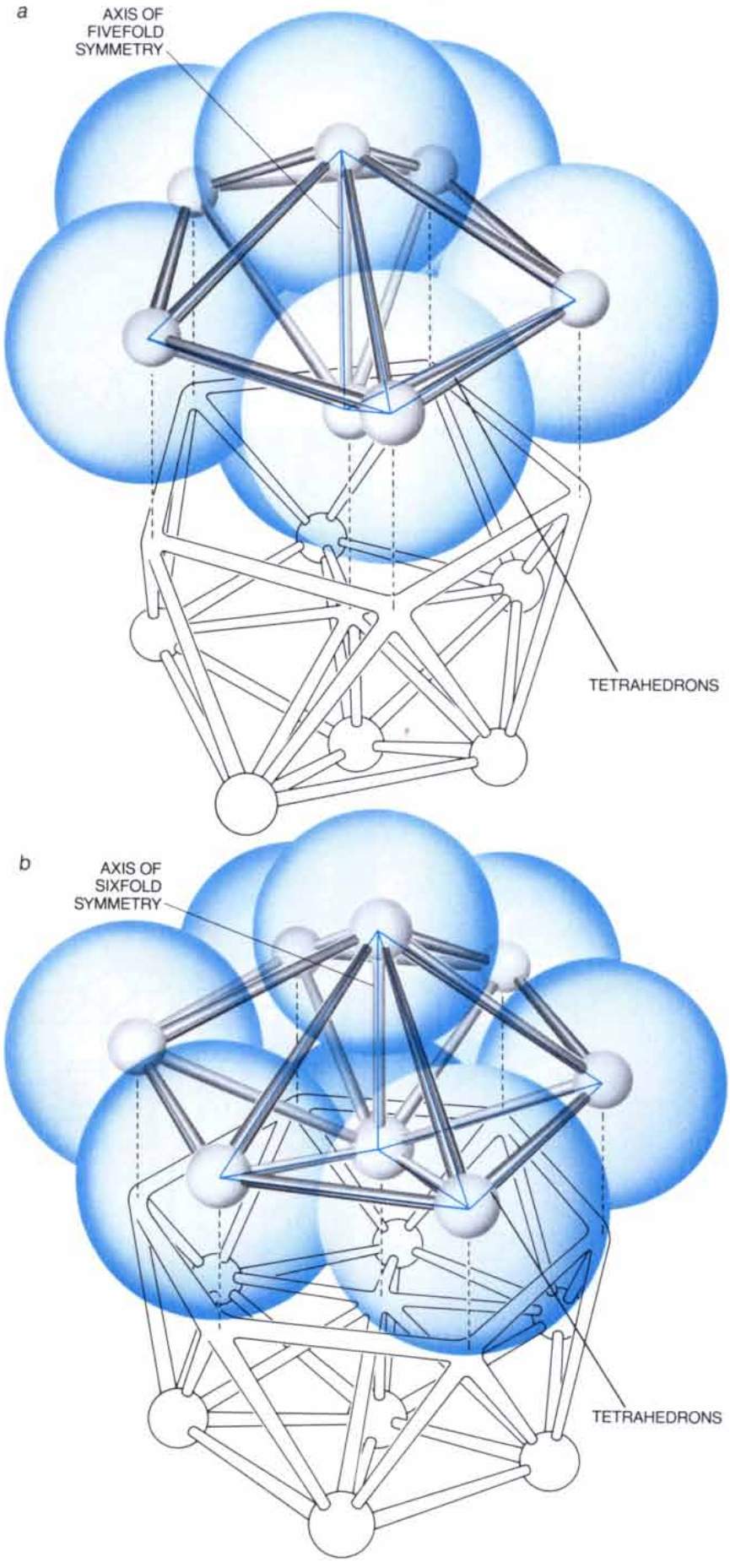
Each bond is shared by four, five or six distorted tetrahedrons; most are

shared by five. Such a grouping of five tetrahedrons around a common bond can be seen as a fragment of an icosahedron. Wherever these groupings occur, then, the alloy has short-range icosahedral order. To accommodate the frustration inherent in a packing of icosahedrons, 10 percent or more of the bonds must be shared by six, rather than five, tetrahedrons.

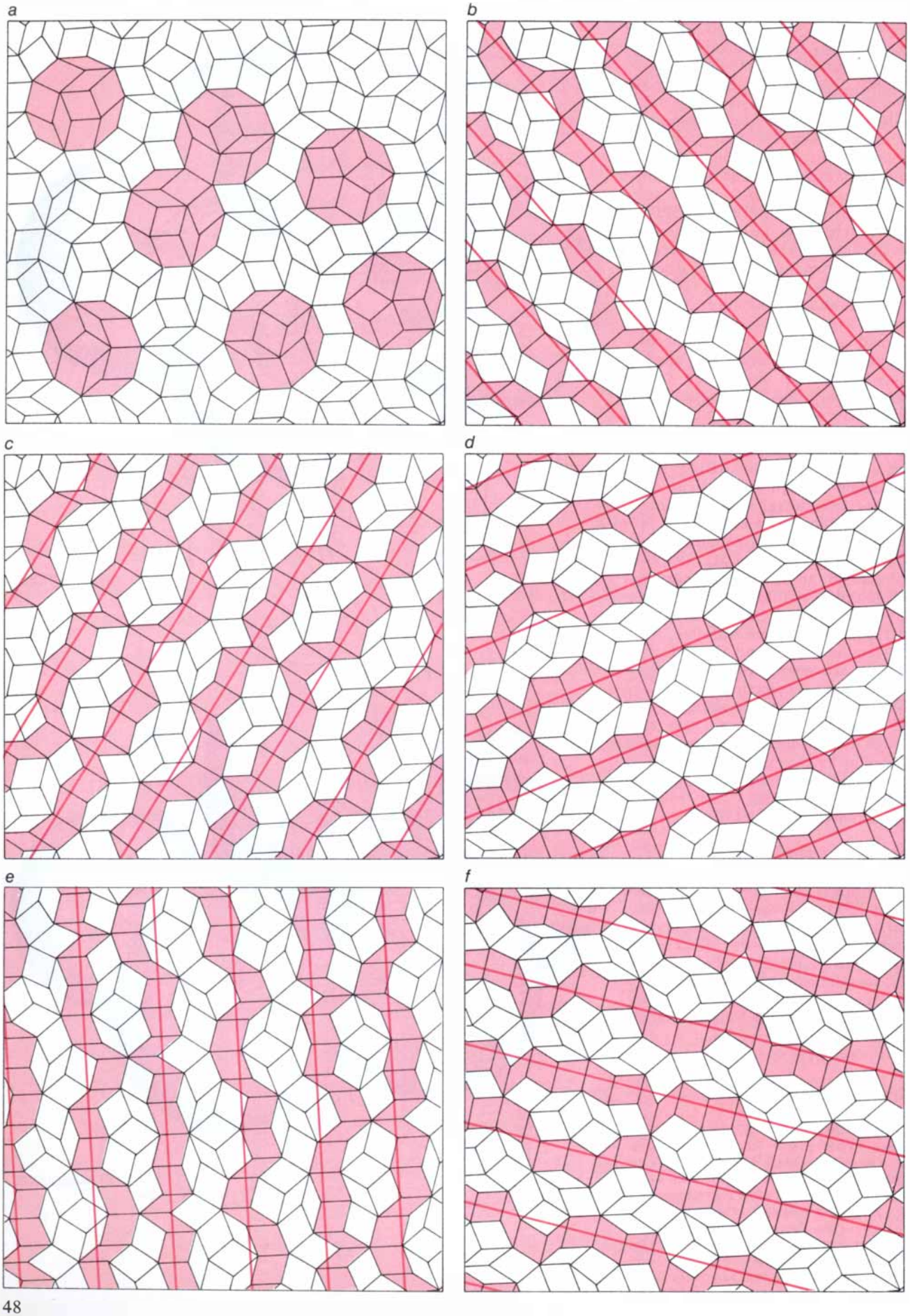
It turns out that all the sixfold and fourfold bonds in the material are connected into long lines, which are called wedge disclinations. Wedge disclinations cannot begin or end inside the material, and at low temperatures they rarely cross. When molten metals are cooled rapidly, the lines can become tangled, thereby forming a metallic glass. When the metals are cooled more slowly, the wedge disclination lines have enough time to cross, and they link up into periodic networks. In many alloys the material then consists of a regular net of sixfold bonds threading through an otherwise completely icosahedral medium. This is the crystalline phase of the alloy; it is known as a Frank-Kasper phase. In the Frank-Kasper phase of $Mg_{32}(Al,Zn)_{49}$ magnesium atoms lie at the sites that are joined by sixfold bonds, and smaller aluminum and zinc atoms occupy the remaining, icosahedral sites.

Metallic glasses and Frank-Kasper phases are good examples of how short-range icosahedral order can form and how it can be accommodated in a crystal that does not have long-range icosahedral order. To understand the long-range icosahedral order in shechtmanite requires a few additional tools, which come from the mathematical theory of tilings.

A tiling can be a good analogue of a crystal. Just as a crystal fills three-dimensional space with unit cells, so a tiling fills two-dimensional space with tiles. Many properties of three-dimensional crystals are also observed in two-dimensional tilings. For example, the two-dimensional hexagonal packing has the lattice planes and long-range orientational symmetry that are found in three-dimensional crystals. It is reasonable to suppose, then, that



CONFIGURATIONS of atoms found in a dense agglomeration of particles can be seen as close packings in which slightly distorted tetrahedrons share common bonds. Most bonds are shared by five tetrahedrons (*a*), forming a fragment of an icosahedron. To relieve the frustration inherent in icosahedral packings, some of the bonds must be shared by six tetrahedrons (*b*), forming a fragment of a solid that resembles an icosahedron but has a sixfold symmetry axis.



some kind of tiling might also capture some of the unusual properties of quasicrystals.

The tiling that is the best model for quasicrystals was actually developed a decade before shechtmanite was discovered [see "Mathematical Games," by Martin Gardner; SCIENTIFIC AMERICAN, January, 1977]. In 1974 Roger Penrose, a mathematical physicist at the University of Oxford, was experimenting with ways of tiling a plane aperiodically (that is, in such a way that the pattern of tiles cannot be broken down into a single unit cell such as a hexagon). Penrose found a tiling that requires tiles of only two shapes, both of which are rhombuses (parallelograms all of whose sides are the same length). One rhombus has internal angles of 36 and 144 degrees, the other has internal angles of 72 and 108 degrees; the rhombuses are fitted together according to certain "matching rules." In an infinite Penrose tiling the ratio between the number of "fat" rhombuses and the number of "thin" rhombuses is exactly equal to the golden mean (about 1.618). Because this ratio is an irrational number, it is impossible to break down the tiling into a single unit cell that contains an integral number of each kind of rhombus.

Like shechtmanite, Penrose tilings are not crystalline in a conventional sense, but they have many crystalline properties. For example, in a Penrose tiling it is possible to pick out many decagons, or regular 10-sided polygons [*see illustration on opposite page*]. Like the hexagons that are the unit cell of the two-dimensional "billiard ball" crystal, all the decagons have precisely the same orientation. Like shechtmanite, the Penrose tiling has the long-range orientational order that is usually associated with conventional crystal lattices.

In a subtler way Penrose tilings have a kind of translational order as well. One way to see this is to shade all the rhombuses that have sides parallel to a given direction. The shaded rhombuses form a series of jagged irregular lines each of which, on the average, approximates a straight line. All the lines are parallel and approximately evenly spaced. In an average, statistical sense, therefore, a Penrose tiling has long-range translational order as well as orientational order.

Like quasicrystals, Penrose tilings also have a kind of fivefold symmetry. In a Penrose tiling the shaded rhombuses fall into five families of parallel lines. The lines run in directions that might be parallel to the edges of a regular pentagon. They intersect at angles that are multiples of 72 degrees, or one-fifth of a full circle. It can be shown that the lines, like the lattice planes of an ordinary crystal, will scatter beams of electrons or X-ray radiation. Beams reflected from a Penrose pattern would have fivefold rotational symmetry no matter where in the pattern they were aimed. The disorderly appearance of the lattice planes is similar to that found in a conventional crystal at temperatures above absolute zero, when the atoms are disordered because of thermal vibrations. In Penrose tilings, of course, the disorder would be present even at a temperature of absolute zero.

The first work that suggested Penrose tilings might be applicable to real materials was done by Alan L. MacKay of Birbeck College of the University of London in 1981. In 1984 Peter Kramer and Reinhardt Neri of the University of Tübingen and Dov I. Levine and Paul J. Steinhardt of the University of Pennsylvania independently developed a three-dimensional generalization of Penrose tilings that has turned out to be closely related to the structure of shechtmanite.

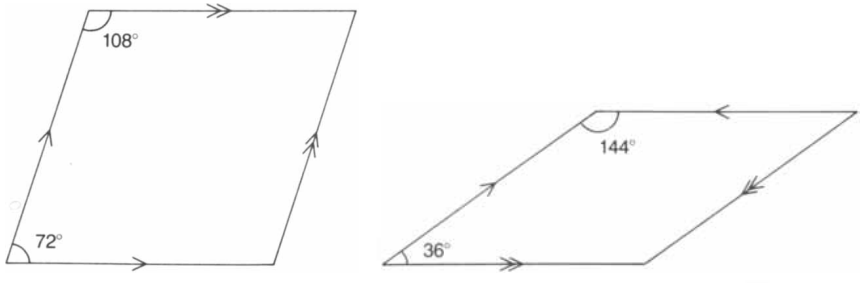
Like their two-dimensional counterparts, three-dimensional Penrose tilings have long-range orientational and translational symmetry. In addition they have long-range icosahedral symmetry, in much the same way as two-dimensional Penrose patterns have long-range fivefold symmetry.

The basic units of the three-dimensional tilings are two rhombohedrons (six-faced figures, each of whose faces is a rhombus; a rhombohedron looks somewhat like a squashed cube). The interior solid angles of the rhombohedron are precisely equal to the angles formed by certain bonds in an icosahedral cluster of atoms [*see illustration on next page*]. In an infinite three-dimensional Penrose tiling the ratio of the number of one kind of rhombohedron to the number of the other kind of rhombohedron is the golden mean, and so a three-dimensional Penrose tiling, like shechtmanite, cannot be described in terms of a single unit cell. Calculations of how three-dimensional Penrose tilings would scatter radiation have produced results remarkably similar to the experimental results achieved with actual samples of shechtmanite.

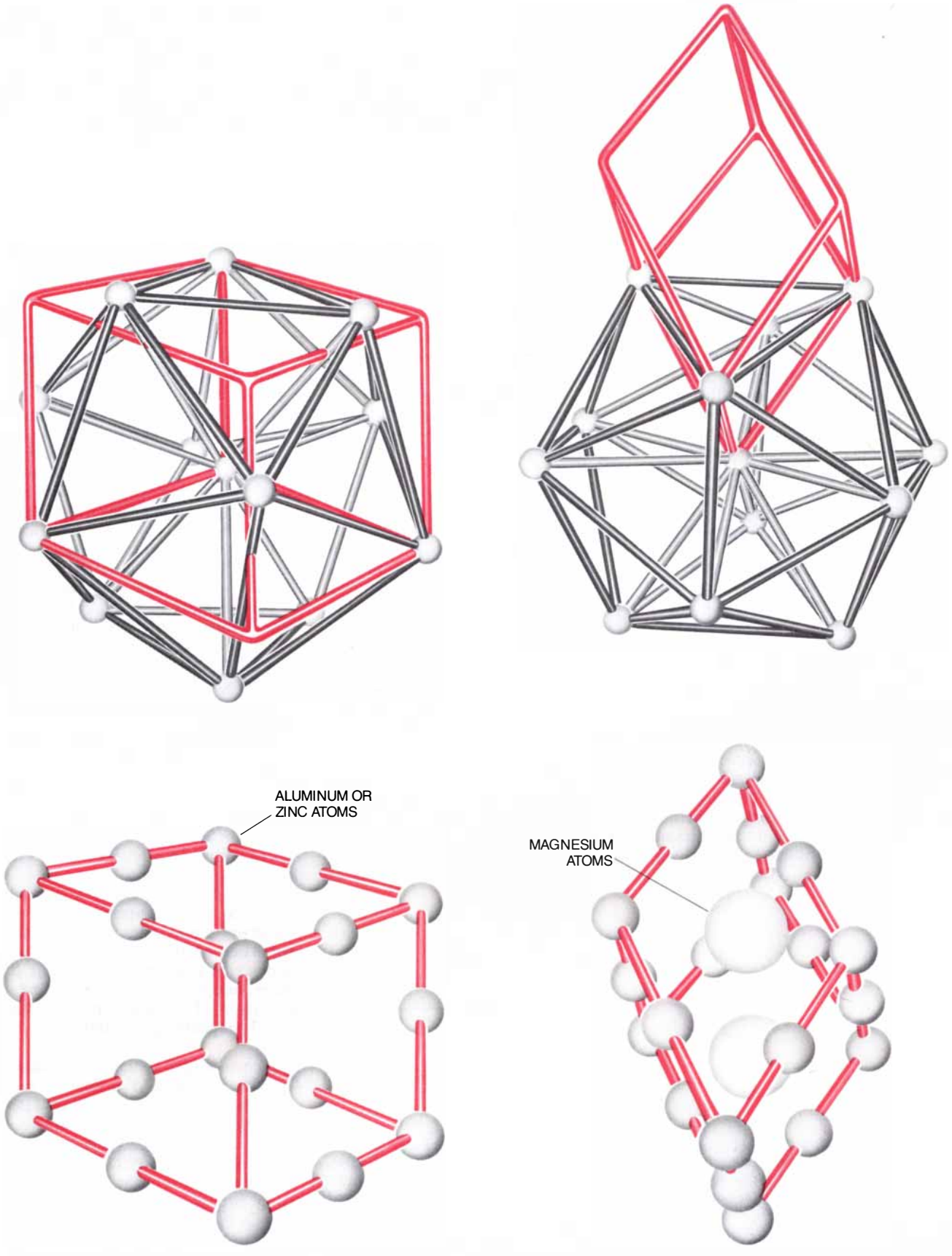
Three-dimensional Penrose tilings provide an excellent starting point for understanding the atomic arrangements in quasicrystals. To understand the structure of any given shechtmanite alloy in terms of three-dimensional Penrose tilings, one must devise a way of "decorating," or filling, the two rhombohedrons with atoms in such a way that an agglomeration of rhombohedrons would contain the correct proportion of atoms of each element. To describe a conventional crystal, of course, one would need only to find a decoration of a single unit cell that could be repeated indefinitely.

Penrose tilings, as they are realized in shechtmanite, represent a new notion of the idea of crystallinity. Because the two "unit cells" are arranged without strict periodicity, every unit cell sits in a slightly different environment. Consequently the forces due to distant atoms will vary from cell to cell, causing variations in the positions of the atoms in each cell. If the Penrose rhombohedrons are to be a good description of a particular shechtmanite alloy, the variations must be small.

In real materials the matching rules



LONG-RANGE ORDER in a Penrose tiling is similar to that in a conventional crystal, but there are subtle differences that allow the Penrose tiling to have a kind of fivefold symmetry. The decagons found throughout the pattern (a, *color*) all have the same orientation, demonstrating the pattern's long-range orientational symmetry. The pattern also has a kind of long-range translational symmetry. When all the rhombuses that have sides parallel to a given direction are colored (*b*), they form a series of jagged lines that are roughly parallel and approximately evenly spaced, like the lattice planes of a conventional crystal. There are five families of lattice planes (*b–f*), which intersect at angles that are multiples of 72 degrees (one-fifth of a full circle). Any pattern formed by X rays or electrons reflected off these planes would have fivefold symmetry. A Penrose pattern is formed when the rhombuses pictured above are fitted together so that the arrows superposed on their edges match.



PENROSE RHOMBOHEDRONS (*color*) **can fit together to form a three-dimensional structure (not shown) that resembles two-dimensional Penrose tilings. The interior angles of the rhombohedrons are derived from angles between bonds of an icosahedral structure. A three-dimensional Penrose tiling has icosahedral symmetry in the same rough sense that a two-dimensional Penrose tiling has fivefold symmetry. Particular alloys are described in terms of Penrose tilings by "decorating" each rhombohedron in such a way that an assembled tiling includes the correct ratio of atoms of various types. A decoration for $Mg_{32}(Al,Zn)_{49}$ is shown at bottom.**

that determine the placement of Penrose rhombohedrons are probably not strictly obeyed. It can be shown that small groups of Penrose cells can be rearranged locally in a way that violates the strict matching rules but still preserves the overall order of the material. In two-dimensional Penrose tilings, for example, the tiles inside any of the decagons can be taken out, scrambled and put back in the decagon in a way that violates matching rules in some places. Nevertheless, the orientation of the decagon remains the same, and so the quasicrystal still keeps its long-range orientational order. The scrambling operation does not interrupt the lattice planes (they merely zig where they once zagged), and so it changes only small details in the way X rays or electrons would scatter from the tilings.

Penrose tilings also provide a link between the crystalline order inherent in the Frank-Kasper phase of many alloys and the long-range icosahedral order of shechtmanite. For example, Christopher Henley of Cornell University and Veit Elser of AT&T Bell Laboratories have shown how the unit cell of the Frank-Kasper phase of $Mg_{32}(Al,Zn)_{49}$ can be seen as a slightly distorted fragment of a Penrose tiling.

In the Frank-Kasper phase the periodic repetition of this distorted fragment throughout space destroys any inherent icosahedral symmetry. In shechtmanite, however, this fragmentary region of icosahedral order propagates throughout the material to form a macroscopic quasicrystal. Henley and Elser have also shown how to decorate Penrose rhombohedrons in a way that explains the structure of the shechtmanite phase of $Mg_{32}(Al,Zn)_{49}$. The icosahedral order in this phase of the alloy is long-range, in spite of the presence of a network of wedge disclination lines similar to those found in Frank-Kasper phases.

Similar studies and proposals have been made to describe the aluminum-manganese alloys originally discovered by Shechtman and his colleagues. The proposed models include many tetrahedrons and fragments of icosahedrons, although they also include some octahedrons. The models can account for the positions of about 80 percent of the atoms in the material.

Penrose tilings also suggest exotic crystallographies based on symmetries that are not icosahedral. For example, Leonid A. Bendersky of the National Bureau of Standards has shown that another phase of rapidly cooled aluminum-manganese behaves as if it were composed of layers of two-dimensional Penrose tilings that are stacked periodically in the vertical direction.

SHECHTMANITE "SNOWFLAKES" form when a molten alloy of aluminum and manganese is cooled rapidly. The "crystals" have the branched pattern of dendrites, or lobes, normally found in ice crystals, but because of the underlying icosahedral symmetry of shechtmanite, these samples have fivefold symmetry rather than the sixfold symmetry of ordinary snowflakes. The samples measure about 10,000 atomic spacings across and are magnified about 10,000 times in this electron micrograph, which was made in the laboratory of Leonid A. Bendersky and Robert J. Schaefer at the National Bureau of Standards.

There are still many unanswered questions. What, for example, is the connection between short-range icosahedral order in liquids or metallic glasses and the long-range icosahedral order in shechtmanite? It seems to be much easier to form microscopic fragments of shechtmanite crystals in an undercooled liquid than it is to form the first microscopic fragments of a conventional crystal. Once these small fragments form, they must grow to macroscopic dimensions. How shechtmanite grows is still not known, although it is clear that its growth processes must be quite different from those of conventional crystals.

Another puzzle concerns the range over which the material's orientational order and translational order extend. Although the orientations of the icosahedrons in shechtmanite are correlated over distances as large as thousands of atomic spacings, the corresponding lattice planes are in phase only over a few hundred spacings. We do not know whether the short correlation lengths of translational order are an intrinsic feature of the phase or an artifact of the cooling process. It may soon be possible to grow large shechtmanite crystals under more controlled conditions; such crystals should help to resolve the puzzle.

Shechtmanite quasicrystals are no mere curiosity. The study of quasicrystals has tied together two existing branches of theory: the theory of metallic glasses and the mathematical theory of aperiodic tilings. In doing so it has brought new and powerful tools to bear on the study of metallic alloys. Questions about long- and short-range icosahedral order should occupy solid-state physicists and materials scientists for some time to come.

Transplantation in the Central Nervous System

Transplanted embryonic neurons can establish functional connections in the adult brain and spinal cord, long believed to be immutable in mammals. Such grafts might reverse damage from disease or injury

by Alan Fine

Can damage done to the brain or the spinal cord by disease or injury be repaired? Neurons, or nerve cells, cannot regenerate in adult mammals. The great majority of them are in place by the end of infancy, and in primates, including human beings, the development of the nervous system is complete by puberty. Axons, the threadlike extensions of nerve cells along which messages travel, can regrow after damage—a capacity underlying the slow return of feeling and movement after certain injuries—but they generally do so only in peripheral nerves. In the brain or spinal cord a damaged pathway rarely re-forms.

Even though the mammalian central nervous system (the brain and spinal cord) shows little capacity on its own for regeneration, over the past decade it has been found capable of sustaining new growth of another kind. In a series of experiments done mainly in rats, other investigators and I have shown that grafts of embryonic brain tissue can be anatomically and functionally incorporated into the adult central nervous system. The interactions of grafts and their host nervous system have revealed much about the factors governing development and regeneration in the central nervous system. The success of certain transplantation experiments has also suggested ways to treat currently incurable disorders, such as Parkinson's disease and Alzheimer's disease, in which parts of the central nervous system degenerate.

The work of the past decade was foreshadowed by a number of earlier efforts. In 1890 W. Gilman Thompson of New York University Medical College tried to transplant pieces of cerebral cortex (the outer layer of the brain) from adult cats to dogs. No neurons survived these procedures or similar manipulations by other workers over the next 15 years. Elizabeth Hopkins Dunn, a physician working as a research assistant at the University of Chicago, guessed that although adult brain tissue did not stand up to the rigors of transplantation, immature tissue might. In work done in 1903 (but not reported until 1917) she exchanged fragments of cortex between 10-day-old rats. In 10 percent of the cases she found surviving neurons in the grafts three months after surgery.

In addition to establishing the fact that immature neural tissue has a better chance of surviving in another brain than adult tissue has, Dunn noticed that the surviving grafts in most cases were those that were richly supplied with new blood vessels. Such transplants had been placed in the ventricles, or fluid-filled spaces, of the host brain, in contact with choroid plexus, the highly vascular membrane that lines the ventricles. In 1940 W. E. Le Gros Clark of the University of Oxford confirmed the importance of immaturity and a rich blood supply for the survival of transplanted neurons. He successfully transferred pieces of cortex from rabbit embryos to the lateral ventricles in six-week-old rabbits. Although the neurons were immature when they were grafted, four weeks later he found that many of them had completed their development.

Many investigators were skeptical about these early results, and the work was not pursued. Current interest in transplantation in the mammalian central nervous system dates from 1971, when Gopal D. Das, Joseph Altman and their students at Purdue University demonstrated beyond doubt that transplanted immature neurons can survive and mature. These workers injected radioactively labeled thymidine, one of the building blocks of DNA, into seven-day-old rats, where the compound was incorporated into the genetic material of newborn cells, including neurons. Das and Altman then transplanted fragments of the cerebellum (a convoluted structure at the back of the brain) from the injected animals to the matching site in untreated animals. Two weeks later they took sections of the host cerebellum and coated them with photographic emulsion. Radioactivity in the sections exposed the emulsion, indicating that transplanted neurons had survived.

In later experiments Das and Altman established the general principle that embryonic brain tissue transplanted during the period when neurons multiply and migrate, before they extend their filamentous axons and dendrites, has the best chance of survival. Because different parts of the brain develop at different rates, the optimum age of the donor animal varies according to the kind of graft that is intended. The age of the host animal has far less bearing on the survival of transplanted tissue than do factors such as the physical stability of the implant and the blood supply available to the graft.

A condition that is now known to have less effect on transplant survival than had been expected is the degree of kinship between the donor and the host animal. In other kinds of graft immunologic rejection swiftly ensues if the donor and the host are not closely related. Accordingly most of the early transplantation experiments were done in closely related or genetically identical animals. It was later learned that immunologic rejection of a graft in the central nervous system may not occur even when the donor and the host animal are genetically different. Indeed, transplants to the central nervous system can succeed

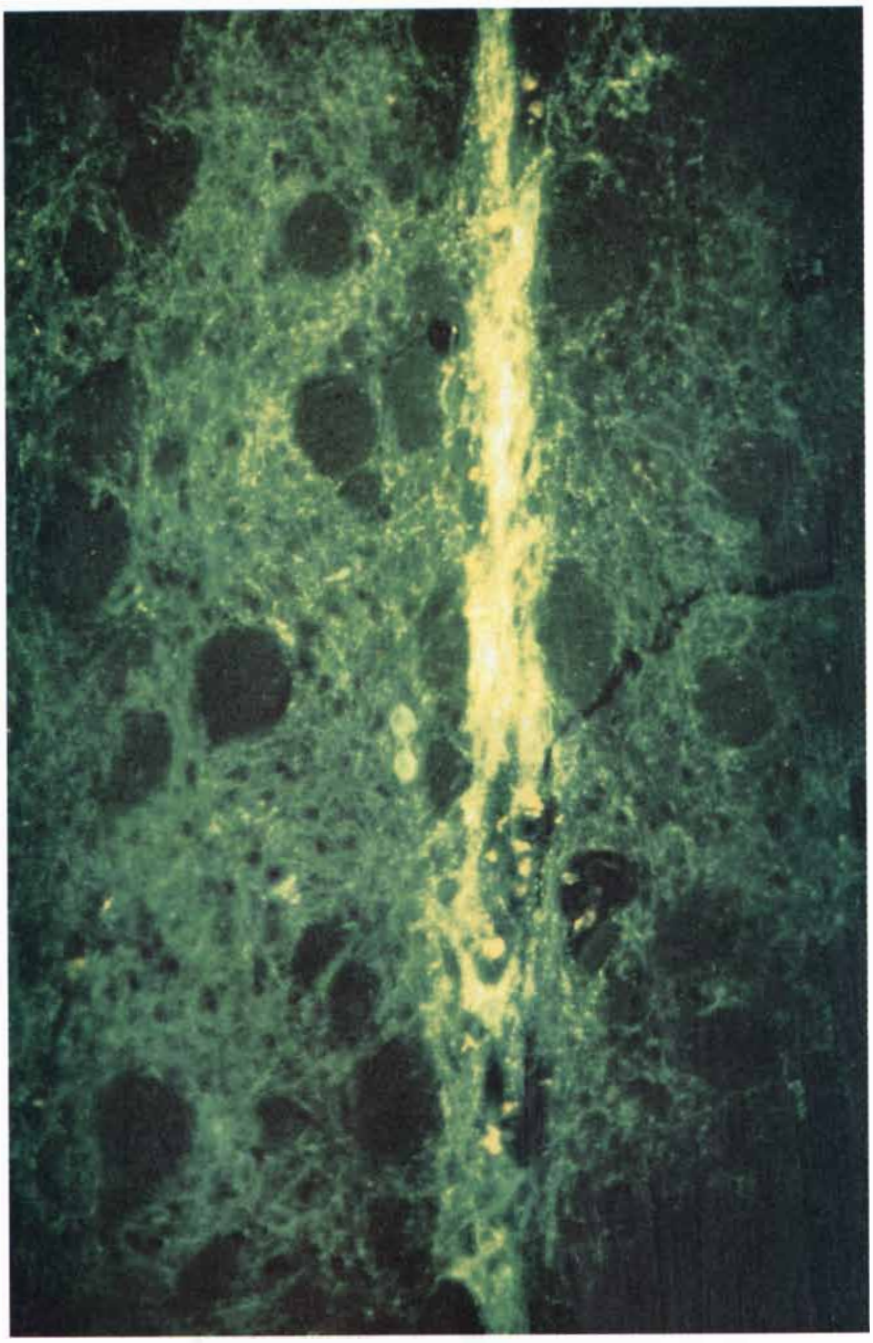
between animals of different species, particularly when the host animal has been treated with an immunosuppressive drug such as cyclosporin.

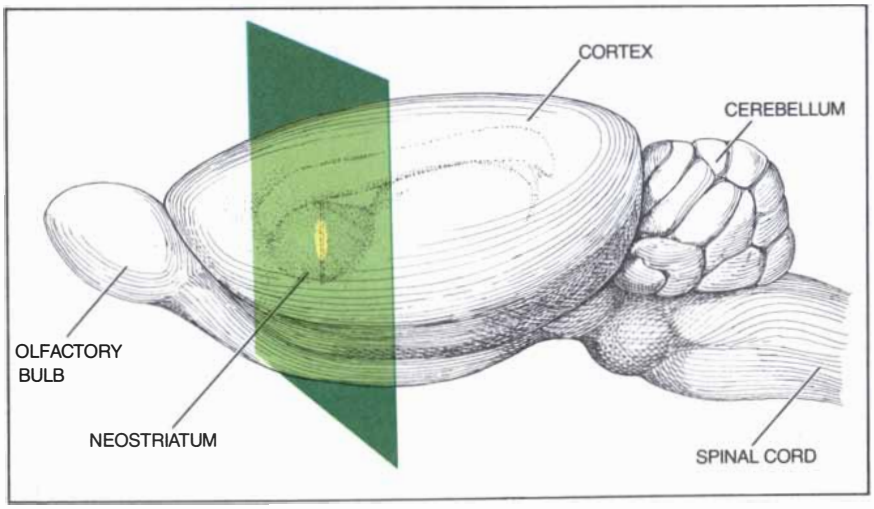
The observation that the brain is an "immunologically privileged" site for transplantation did not originate with grafts of neurons. Earlier work had shown that grafted skin and tumor tissue could survive in the brain even in an animal that promptly rejected similar grafts to its skin. It is thought that because the brain lacks lymphatic vessels and lymph nodes, from which many of the cells of the immune system are deployed, and because the walls of blood vessels in the central nervous system are specialized to create a "blood-brain barrier," the access of the immune system to foreign tissue in the brain is limited.

In the case of neuronal transplants the lack of rejection may also reflect the characteristics of nerve cells proper. On their surface most cells bear large molecules known as class I major histocompatibility antigens. The antigens are distinctive in each animal; they are the molecules the immune system recognizes as foreign when it rejects grafted tissue. It is now known that those antigens are normally rare or absent on most neurons.

The routine survival of transplanted neural tissue opens the way to study of its interaction with the host. One line of inquiry to which the technique has been applied is the identification of factors governing the response of the central nervous system to injury. Transplants can serve to alter conditions in a traumatized brain or spinal cord, making it possible to gauge the relation of a given factor to the limited regenerative capacity of central neurons.

It was known, for example, that the





YELLOW FLUORESCENCE marks nerve cells transplanted into a rat brain. Nerve fibers containing the neurotransmitter dopamine had previously been eliminated from the neostriatum, the brain structure that is pictured, in order to induce an experimental analogue of Parkinson's disease in the animal. Neurons that synthesize dopamine were then taken from a rat embryo and injected as disaggregated cells into the neostriatum. Six months later a cross section of the host brain was made along the plane indicated in the diagram at the right. The section was treated with formaldehyde vapor, which combines with dopamine to form a fluorescent compound. The fluorescence at and around the implant shows the transplanted neurons had survived, had sent fibers into the neostriatum and had continued to make dopamine. The image is from Anders Björklund of the University of Lund.

difference between the regenerative capacity of axons in peripheral nerves and that of axons in the central nervous system is correlated with a difference in their glial cells. Those cells surround the axons and can myelinate them, that is, encase the fibers in a sheath of fatty, electrically insulating material. Peripheral nerves are myelinated by Schwann cells; the myelinating cells in the brain and spinal cord are oligodendrocytes.

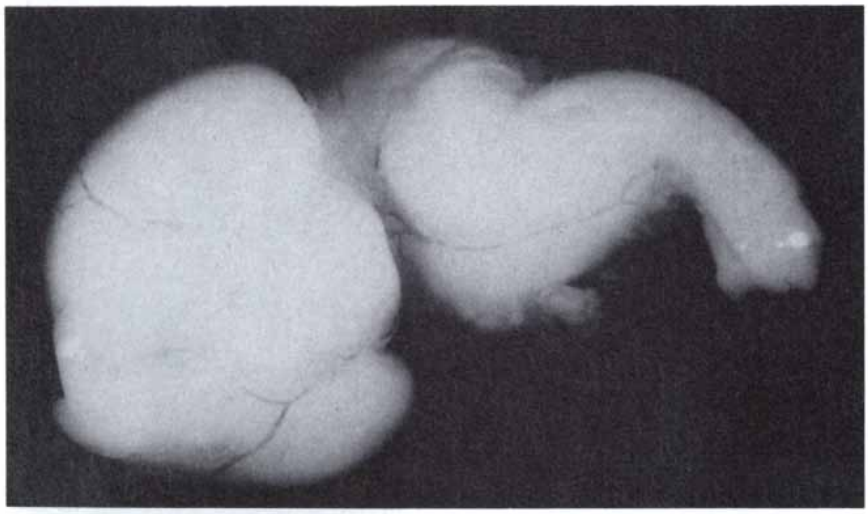
Studies of sensory neurons, which have both a central and a peripheral fiber, suggest glial cells influence regeneration. From a site just outside the spinal cord a sensory nerve cell extends one fiber to a sensory structure in the periphery of the body and another into the spinal cord. The peripheral fiber, which is myelinated only by Schwann cells, ordinarily regenerates after damage. The great Spanish neuroanatomist Santiago Ramón y Cajal noticed 60 years ago that if the central fiber is crushed, it will regrow only to the point where it enters the spinal cord and oligodendrocytes replace Schwann cells.

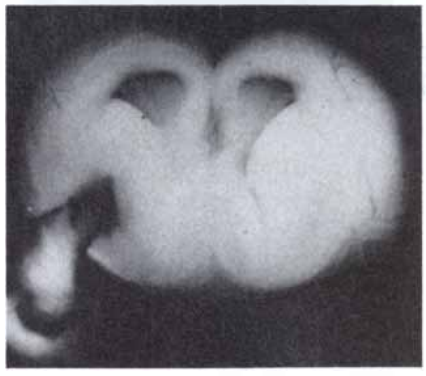
A number of workers have suggested that either oligodendrocytes secrete a substance inhibiting axon growth or Schwann cells somehow stimulate axons to grow. Several ways in which Schwann cells might promote fiber growth have already been identified. They secrete proteins such as laminin that have been found in culture to promote the growth of axons and their adhesion to a surface. In response to peripheral-nerve injury Schwann cells may also produce nerve growth factor, a well-studied substance that fosters survival and fiber growth in certain peripheral nerve cells.

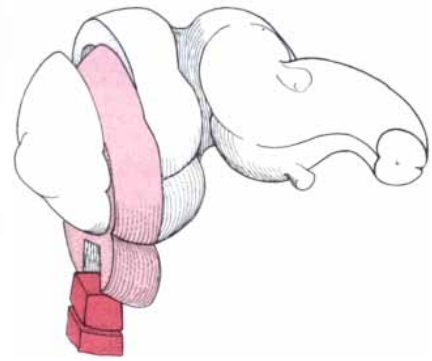
In 1977 Carl C. Kao and his collaborators at Georgetown University tested the possibility that peripheral glia might support the regeneration of damaged fibers in the central nervous system. They used delicate surgical techniques to splice lengths of sciatic nerve (a major peripheral nerve) into gaps made by removing a section of the spinal cord in rats. Within weeks, the workers found, the graft and the cord healed to form a smooth, scar-free union, and some fibers seemed to have grown from the spinal cord into the peripheral-nerve segment.

Albert J. Aguayo, Peter M. Richardson and their colleagues at the Montreal General Hospital and McGill University have confirmed and extended those findings. A month after replacing lengths of spinal cord with segments of sciatic nerve from the same rats, they examined the cords under both the light and the electron microscope and noted many fibers growing into the grafts. After three months, by injecting minute quantities of the enzyme horseradish peroxidase into the spinal cord adjacent to the graft, the workers were able to demonstrate that some of the regenerating fibers had succeeded in growing all the way through the graft, bridging the severed ends of the cord.

Horseradish peroxidase is taken up by nerve terminals and carried back along the axons to the cell bodies; incubation in a solution that is chemically altered by the enzyme stains cell bodies whose axons have transported it. In the grafted rats staining appeared in spinal neurons on the other side of the graft from the injection site; their fibers must have grown through the segment of peripheral nerve. In later work Aguayo and his colleagues found that although the regenerating fibers can span grafts several centimeters long, they continue into the spinal cord beyond the graft for only a few millimeters at most.







EMBRYONIC RAT BRAIN (*top*), seen from below, is the usual source of neurons for transplantation to adult rats. The four-millimeter-wide brain is from a 17-day embryo (rat pups are born at about 22 days). At that stage of development the neurons are still multiplying and migrating through the brain. In many areas they have not yet grown axons and dendrites, the processes through which they will communicate with other cells and tissues. When they are transplanted to an adult brain, such neurons can mature and send axons into host tissue. The inner piece of tissue dissected from a slice of embryonic brain (*left*) contains cells that are the precursors of a structure known as the nucleus basalis. Such cells, implanted into an adult rat brain, can reverse a defect that simulates a feature of Alzheimer's disease. The diagram (*right*) shows the orientation of the slice in the intact brain.

Transplantation experiments thus can shed light on the differing regenerative capacities of the peripheral and central nervous systems. A similar contrast in the ability to regenerate and reestablish connections is evident between the embryonic brain, which undergoes extensive changes as it develops and often can recover from injury, and the adult brain. By transplanting pieces of embryonic rat brain to injured adult brains Anders Björklund and Ulf Stenevi of the University of Lund and Lawrence F. Kromer, then at the University of California at San Diego Medical Center, identified one possible basis of the difference.

Before doing the transplants the investigators severed a band of nerve fibers known as the fimbria at its entrance to the hippocampus, a folded and curved sheet of evolutionarily ancient cortex that in human beings plays

a role in memory and emotion. The fimbria includes fibers arising in two structures of the forebrain, the medial septal nucleus and the diagonal band nucleus. The fibers contain the neurotransmitter acetylcholine, one of the brain's chemical messengers. Cutting the fimbria eliminates acetylcholine and its synthesizing enzyme, choline acetyltransferase, permanently from most of the upper hippocampus.

When the workers transplanted a piece of embryonic rat hippocampus into the gap in the fimbria, the cholinergic (acetylcholine-releasing) fibers grew through the graft and into the host hippocampus. Six months later choline acetyltransferase in the hippocampus adjacent to the graft was restored to as much as 50 percent of its normal level. Similar experiments using other parts of the embryonic brain have indicated that embryonic tissue can act as a bridge for the growth of axons in the central nervous system, provided the grafted tissue is the normal target of the axons. The results suggest embryonic brain structures may produce specific neurotrophic factors (substances that stimulate the growth of nerve cells and fibers) that guide developing fibers to their targets. Alternatively, grafted embryonic tissue may somehow induce the host brain to produce such factors.

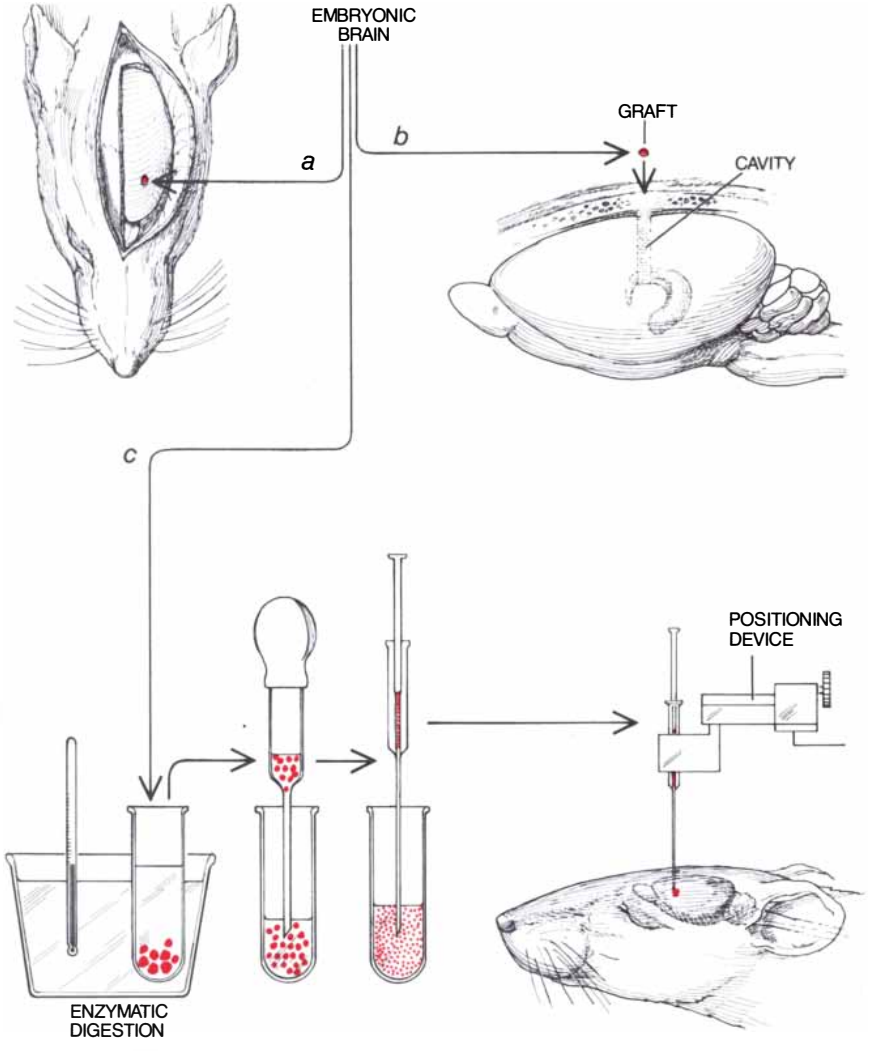
That the adult brain can secrete neurotrophic factors under certain conditions is borne out by observations made by Ellen R. Lewis and Carl W. Cotman at the University of California at Irvine in 1982. They noted that embryonic brain tissue transplanted to cavities made in the brain of adult rats survived better in cavities formed between three and six days before the implant than in cavities made earlier or more recently. During that interval, the investigators surmised, some neurotrophic factor made or accumulated by the central nervous system following an injury must reach its greatest concentration.

Together with Manuel Nieto-Sampedro and other colleagues, the Irvine workers demonstrated the presence of such a factor by collecting fluid from the cavities and adding it to neurons in culture. The fluid extended the survival of the neurons, and the effect was greatest when the cavity was between three and six days old. Such a factor might act as an analogue of nerve growth factor and stimulate the limited fiber growth seen after certain injuries to the central nervous system.

In addition to pointing to influences on the overall extent of fiber growth, the fate of neural transplants hints at factors that control the establishment of specific patterns of connections in the brain. In contrast to the normal inability of adult neurons to grow new fibers or regenerate injured ones, grafted embryonic neurons are able to extend fibers into the host brain routinely. This growth is not indiscriminate. The normal anatomical relation of the grafted and the host tissue and the existing innervation of the host structures both influence the extent and pattern of fiber growth.

Fibers extending from grafted tissue are most likely to grow into the structures normally innervated by that tissue. Björklund, Stenevi and their coworkers transplanted embryonic neurons to the cortex of adult rats and also to the neostriatum, a region that lies under the cortex in the forebrain and is related to movement. The neurons were taken from the substantia nigra, a midbrain structure that sends many fibers containing the neurotransmitter dopamine to the neostriatum but very few to the cortex. The cells survived in both places but sent out appreciable numbers of fibers only in the neostriatum, normally their principal target.

Raymond D. Lund, then at the Medical University of South Carolina, and his students C. B. Jaeger and Steven C. and Linda Kirschen McLoon made detailed observations about the specificity with which grafts interact with host brains. They transplanted embryonic



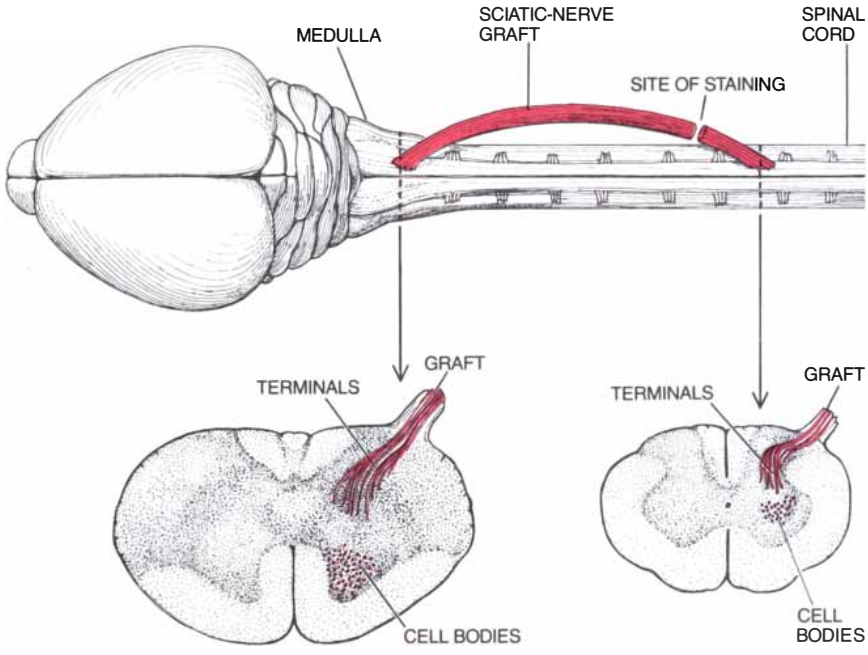
MODE OF TRANSPLANTATION varies according to the aim of the research. Solid tissue dissected from the appropriate part of embryonic rat brain can be implanted directly into the host brain (*a*). Alternatively, a cavity can be prepared in advance (*b*); blood vessels grow into the walls of the cavity and ensure a rich blood supply to the graft. Those approaches, in which embryonic tissue is implanted with its internal organization intact, make it possible to study the growth of fibers from the host brain into the graft. A third approach (*c*) is less disruptive of the host brain. Embryonic tissue is digested by an enzyme and then passed through a series of progressively finer pipettes to disrupt it. A suspension of single cells results; the cells can be labeled with a radioactive substance or a dye, stored or manipulated in other ways before being injected accurately into any site in the host rat brain.

retinas to the brains of newborn and adult rats. In some cases the graft site was the superior colliculus, one of the regions in the brain to which retinal neurons normally send fibers; in others it was the cortex or the cerebellum, areas that receive no retinal inputs. Although the retinas survived in every site, fiber outgrowth ensued only in the superior colliculus. The outgrowing fibers attained the normal targets of retinal fibers in the colliculus and sometimes also followed the normal retinal pathways to more distant targets.

In addition to sending out fibers, tissue transplanted to the brains of newborn rats in some cases also received inputs from the host brain; again the interaction was highly specific. Grafts of retina, which ordinarily is not innervated by other brain structures, received no fiber growth from the host brain. Pieces of embryonic colliculus placed in the superior colliculus of newborn rats, however, not only grew fibers themselves but also became threaded with appropriate inputs from the host retina and cortex. Such results hint at the existence of specific local cues governing fiber growth in the central nervous system.

The finding that grafts send the largest number of fibers into areas of host brain that have been deprived of their normal innervation gives further evidence for specific local influences on fiber growth. In the work of Lund and his colleagues, for example, the retinal grafts generated the most profuse fiber growth when the eye normally sending fibers to the site of the implant had been surgically removed beforehand. Björklund, Stenevi and their co-workers made a similar observation when they transplanted embryonic tissue containing the locus ceruleus, a collection of cells in the brainstem that send fibers containing the neurotransmitter noradrenaline to targets throughout the brain, including the hippocampus. When the tissue was implanted in the hippocampus of adult brains, it extended fibers that recapitulated the pattern of innervation normally established there by the host brain's own locus ceruleus—but only if the pathways from the locus ceruleus had been destroyed before the transplant.

Work not involving transplants has shown that damage to a set of fibers in the central nervous system can result in "collateral axon sprouting": other, healthy fibers extend new branches into the denervated area. The much more luxuriant fiber outgrowth from grafts into denervated brain structures seems to present an analogous case. Taken together, such observations suggest that the growth of nerve fibers from structures in the developing brain as well as from grafts may be regulated by competition either for some scarce diffusible material released by target neurons or for occupancy of particular sites on the target.



BRIDGE OF PERIPHERAL NERVE was grafted into rat central nervous system by Albert J. Aguayo and Samuel David of McGill University to study the nerve's effect on axon growth. The workers inserted one end of a segment of sciatic nerve (a major nerve in the leg) into the medulla (part of the brainstem) and the other end into the spinal cord. After the graft had been in place for six months they severed it, applied a stain to the cut surfaces and then examined cross sections of the medulla and spinal cord made at each end of the graft. Sections from both sites contained stained cell bodies; the sections also revealed stained fibers coursing from the nerve segment into the adjacent host tissue. Evidently neurons in both the medulla and the spinal cord had sent fibers all the way through the sciatic-nerve graft; the fibers carried the stain in both directions to their terminals and cell bodies. It is thought the glia, or supporting cells, of peripheral nerves foster axon growth.

Studies of fiber growth confirm that grafted embryonic tissue can be incorporated anatomically into the host central nervous system. Is the integration functional as well? If it is, could grafts restore normal function to a brain impaired by injury or disease?

Several features of the central nervous system might work in favor of recovery mediated by grafts. For one thing, recovery of function in the central nervous system may not require the precise reconstruction of damaged elements. Many important and complex functions in the brain and spinal cord are performed semiautonomously by discrete structures or assemblies of neurons. To restore a measure of function after an injury that isolates such a structure, it may be sufficient to replace only a few of its inputs.

Over the past 15 years, for example, workers have established that the basic elements of locomotion, such as the rhythmic spinal undulations in fishes and the sequential contractions of the limb muscles that underlie stepping motion in mammals, are generated by groups of neurons in the spinal cord. These "locomotion generators" are subject to several stages of control: they are inhibited by other spinal neurons, which in turn are inhibited by fibers that descend from the brainstem and are thought to release catecholamine neurotransmitters such as noradrenaline and dopamine. Cutting the connection between the brainstem and the spinal cord in a cat gives free rein to the inhibitory spinal neurons and hence paralyzes the animal, but if the cat is injected with L-dopa, a precursor of dopamine, and is supported on a treadmill, it will walk. These observations have suggested that paralysis caused by spinal-cord injury might respond to grafts of embryonic catecholamine neurons into the spinal cord near the locomotion generators, even if the spinal pathways are not repaired.

The spinal cord includes not only long pathways consisting of a single continuous fiber but also pathways made up of multiple short links, in which messages are relayed by a series of neurons. It is therefore conceivable that in a severed spinal cord these multiple-relay pathways, at least, could be mended with transplants of embryonic neurons, which would transmit infor-

mation across the gap in the pathways.

Neurochemical factors as well as neuroanatomical ones brighten the prospect for graft-mediated recovery of function. The chemistry of the brain changes after it is damaged. These effects have been studied in detail in an experimental analogue of Parkinson's disease. The disease is the result of the degeneration of the dopamine-releasing fibers that extend from the substantia nigra to the neostriatum. The degeneration can be induced in the rat brain by injecting either of the two structures or the bundle of fibers with the substance 6-hydroxydopamine, which selectively destroys neurons, fibers and terminals containing catecholamines such as dopamine.

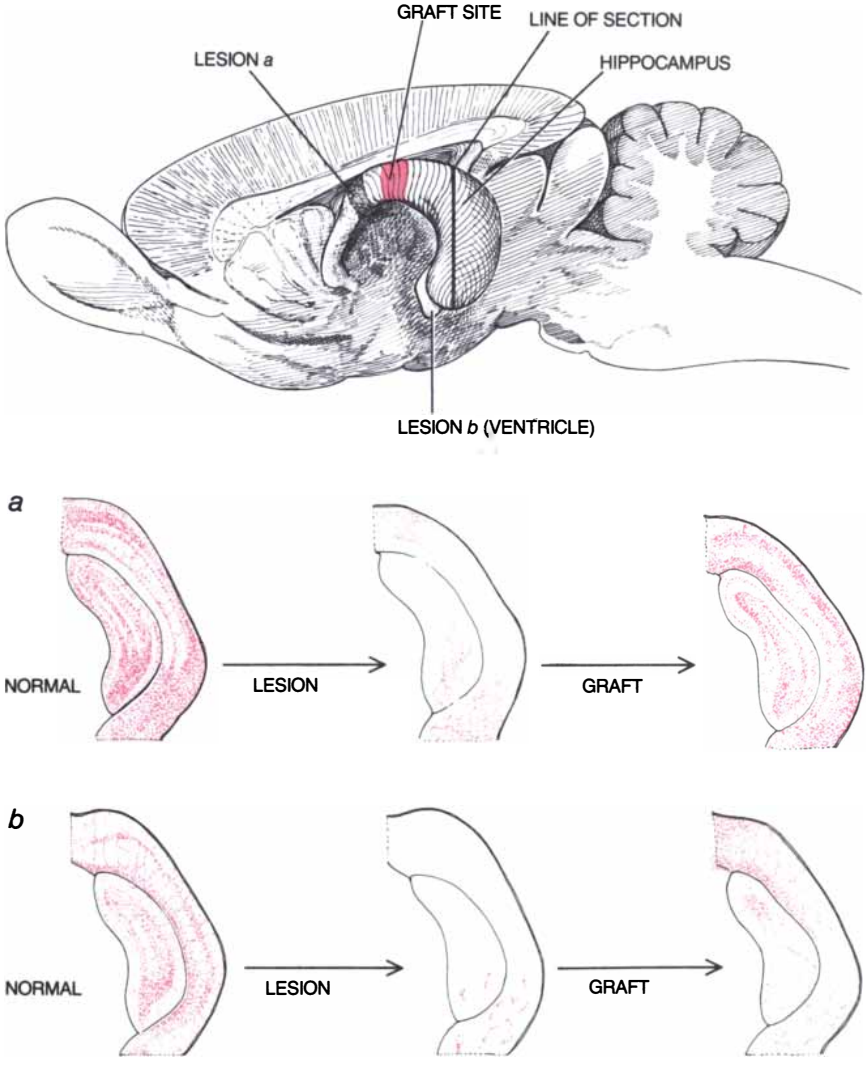
When the dopamine pathway is destroyed on both sides of the brain, the animal becomes immobile and may die unless it is nursed intensively. A rat injected on only one side no longer responds to sensations on the opposite side of the body, which is controlled by the damaged half of the brain. It develops an asymmetric posture, twisting away from the unresponsive side. When the animal moves, it turns toward its "good" side—the one that received the injection.

If the lesions (the damaged areas) spare more than 5 percent of the dopamine-containing cells or fibers, the animal can recover. Much of the recovery takes place within one or two weeks, far too soon to be the result of regeneration of the damaged fibers or collateral sprouting of other dopaminergic neurons. Experiments done in a number of laboratories, notably that of John F. Marshall of the University of California at Irvine, have identified two kinds of neurochemical changes that take place in the nervous system following injury or damage and may account for the rapid recovery.

One change is an increase in the rate at which surviving fiber terminals synthesize and release neurotransmitter. The other is an increase in the sensitivity of the target neurons to the small amounts of transmitter still being secreted. Such "denervation supersensitivity," which usually results from an increase in the number of receptors for the neurotransmitter on the surface of the target cell, would increase the efficacy of the surviving dopaminergic fibers in the lesioned rats.

Denervation supersensitivity probably accounts for some of the remarkable ability of embryonic neuronal transplants to reverse behavioral abnormalities in animal models of degenerative neurological diseases. The 6-hydroxydopamine model of Parkinson's disease provided the earliest example. Mark J. Perlow, William J. Freed and Richard Jed Wyatt of the National Institute of Mental Health, working with Lars Olson and Åke Seiger of the Karolinska Institute and Barry Hoffer of the University of Colorado Health Sciences Center at Denver, destroyed the substantia nigra on one side of the brain in rats. They then transplanted fragments of embryonic brain containing the precursor cells of the substantia nigra to the lateral ventricle adjacent to the denervated neostriatum.

The movement asymmetries that had resulted from the lesions were reduced after the grafts. Injections of the drug apomorphine confirmed that the transplanted neurons were releasing dopamine. Apomorphine binds to dopamine receptors and activates them. Its effect yields a measure of denervation supersensitivity: neurons that lack their usual input of dopamine will be unusually sensitive to apomorphine. Thus injecting it into the neostriatum of a lesioned rat causes the rat to reverse its turning and twist away from



REESTABLISHMENT OF NORMAL INNERVATION after transplantation of embryonic tissue into denervated rat hippocampus is evident in cross sections treated to distinguish fibers containing particular kinds of neurotransmitters. The diagram (*top*) indicates the orientation of the sections and the sites of the experimental manipulations. The upper series of sections were exposed to a substance that reveals the distribution of acetylcholine before (*left*) and after (*middle*) the bundle of cholinergic (acetylcholine-containing) fibers entering the hippocampus was cut (*lesion* a). Four months after cholinergic neurons from an embryo were implanted in a denervated hippocampus, a third section (*right*) reveals that the grafted cells had extended fibers recapitulating the normal pattern of cholinergic innervation. The lower series of sections show the normal pattern of noradrenaline-containing fibers (*left*) and their elimination (*middle*) after the injection into the adjacent ventricle of a toxin that specifically destroys such fibers (*lesion* b). Six months after the implantation of embryonic tissue containing noradrenaline the original pattern of innervation was largely reestablished (*right*). Björklund and Ulf Stenevi of the University of Lund did the work.

the lesion: the neostriatum on the damaged side in effect overreacts to the drug. Because some dopamine input to the neostriatum had been restored by the grafts, the grafted rats' response to apomorphine was only half as great as the response of lesioned rats that had not received the grafts, and the reduction lasted for the length of the experiment: more than six months.

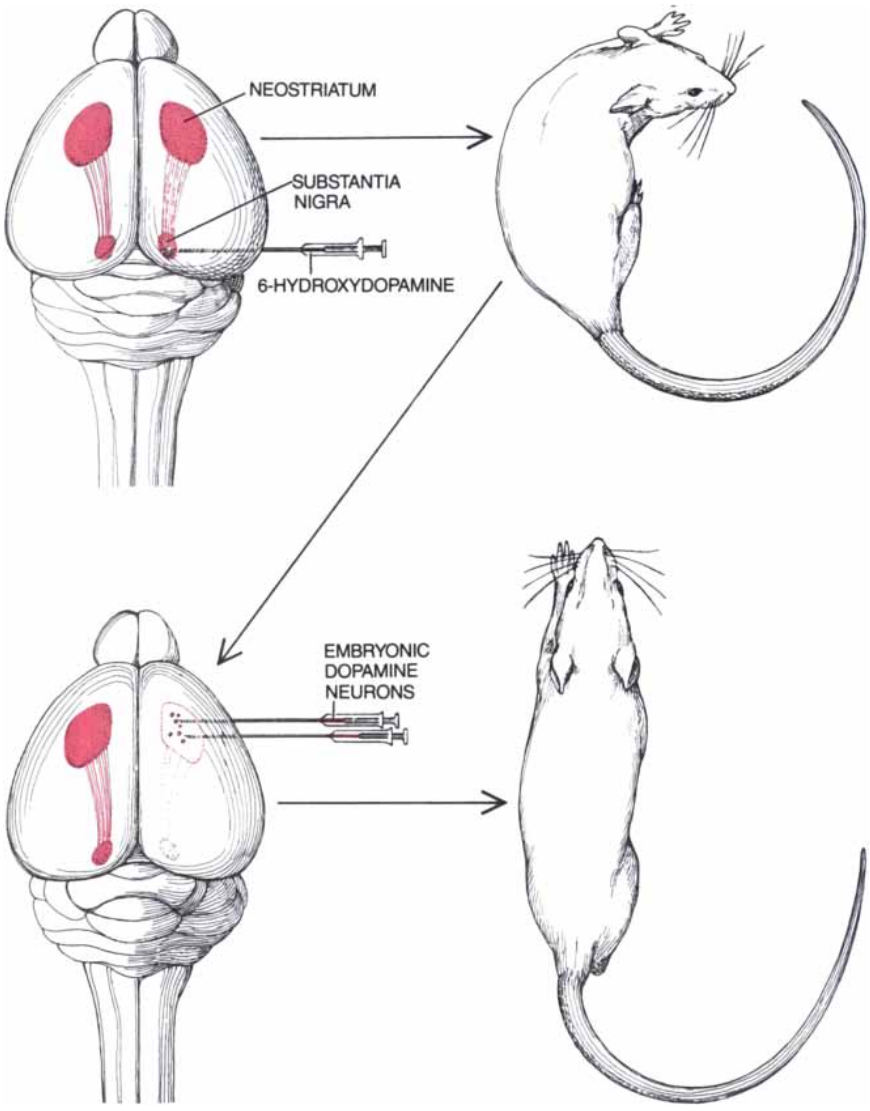
Freed and Wyatt have found that implants of cells from the adrenal gland, a hormone-secreting structure above the kidney, can also reverse some of the effects of 6-hydroxydopamine injection. The cells, taken from the medulla, or central part, of the gland, normally produce adrenaline, a hormone derived from dopamine. When medullary cells are removed from the gland, they may secrete dopamine instead. Like the neuronal transplants, the cells were placed in the lateral ventricle, and they reduced the rats' movement asymmetry but not their sensory abnormalities. Grafts of tissue from embryonic substantia nigra into sites within or very near the neostriatum, done by Björklund and Stenevi, in collaboration with Susan D. Iversen and Stephen B. Dunnett of the University of Cambridge, reduced or abolished the rats' movement disorder and also eliminated the animals' sensory asymmetry altogether.

Experimental models of Alzheimer's disease offer another striking example of transplant-mediated recovery. The progressive loss of memory and other higher mental functions in Alzheimer's disease is associated with widespread degeneration of neurons and neurochemical abnormalities, notably a depletion of acetylcholine in the hippocampus and much of the cortex. Depriving the rat hippocampus of its main acetylcholine input by cutting the fimbria has been found in many laboratories to impair certain kinds of learning, in particular the spatial kind that is needed to learn a maze.

Björklund and his co-workers, including Fred H. Gage, transplanted solid pieces of tissue from the forebrain structures that normally supply the hippocampus with acetylcholine into the cut fimbria of rats; in other rats the workers introduced the tissue directly into the denervated hippocampus in the form of disaggregated cells. The grafts improved the animals' ability to learn a maze—an effect that was most pronounced when the animals were also injected with physostigmine, a drug that blocks the degradation of acetylcholine and thereby enhances its effects.

The cerebral cortex, which is also affected in Alzheimer's disease, gets its input of acetylcholine from the nucleus basalis of Meynert, a cluster of cells at the base of the forebrain. Those cells degenerate in Alzheimer's disease. To simulate this feature of the disease my colleagues, including Dunnett and Guy Toniolo of the University of Strasbourg, and I injected the region of the nucleus basalis on one side of the rat brain with the toxin ibotenic acid. The toxin, which destroys neurons but spares nerve fibers passing through the injected area, permanently eliminates almost all the acetylcholine input from the nucleus basalis to the cortex on the injected side. Certain sensory and movement asymmetries result from the injection, along with serious impairment of the rat's ability to learn and remember simple tasks, such as avoiding the dark chamber in a two-chamber maze or finding a submerged platform in a pool of water.

Because the neurons of the nucleus basalis are interspersed among various other brain structures, the ibotenic acid inevitably destroys a range of cells that send fibers to parts of the brain other than the cortex. Consequently it is unclear whether the behavioral abnormalities of injected rats stem only from the destruction of the cholinergic pathways from the nucleus basalis to the cortex or also reflect damage to other pathways. The ability

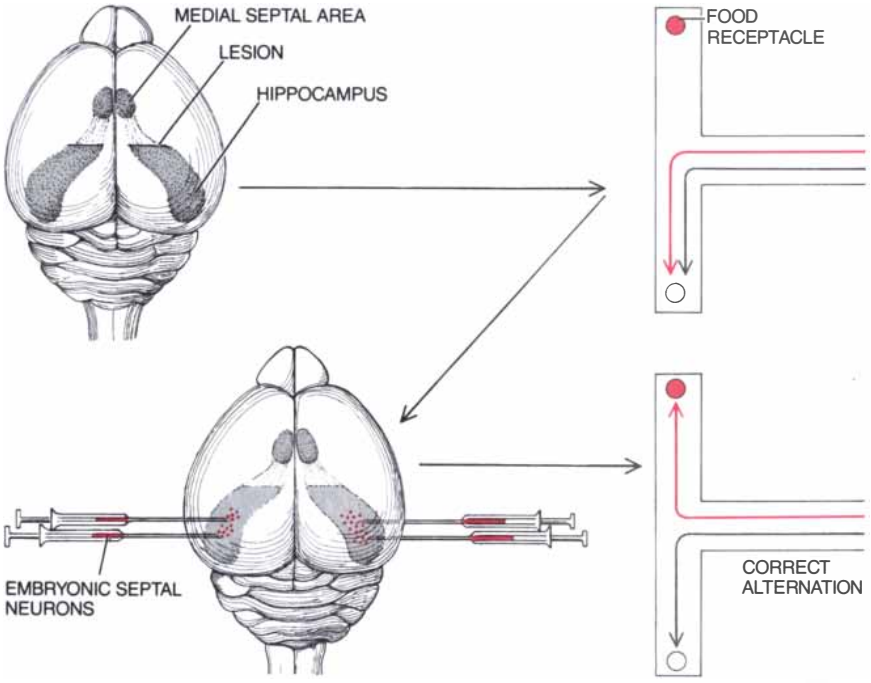


EXPERIMENTAL PARKINSON'S DISEASE induced in a rat can be reversed by transplants of dopamine-producing neurons. An injection of the toxin 6-hydroxydopamine destroys the substantia nigra, which sends dopamine-containing fibers to the neostriatum. Bilateral injections severely debilitate the animal, but if it is injected on one side only, the rat develops an asymmetry of posture and movement: it turns spontaneously toward the injected side. William J. Freed, Richard Jed Wyatt and their collaborators at the National Institute of Mental Health, and independently Björklund, Stenevi and their co-workers, have found that transplantation of dopamine-producing neurons taken from rat embryos to the denervated neostriatum eliminates the movement asymmetry for at least six months.

of transplants of embryonic nucleus basalis to reverse the abnormalities, we thought, would resolve that question and thereby clarify the role of the nucleus basalis in normal functioning.

We found that grafts of cells taken from the precursor region of the nucleus basalis in rat embryos and implanted directly into the denervated cortex of the toxin-injected rats restored their spatial memory to normal, although the grafted rats still learned their tasks more slowly than normal rats did. The grafts also corrected some of the movement asymmetries but did not reverse the sensory deficits. Control transplants of cells from embryonic hippocampus, which lack acetylcholine, were entirely without effect.

As well as suggesting that transplants may be able to reverse Alzheimer's-like deficits, the outcome also confirms that the cholinergic pathways from the nucleus basalis to the cortex play a role in the spatial memory that was impaired in lesioned rats. Transplantation could be applied in much the same way to the analysis of functional neuroanatomy throughout the central nervous system. Where it is difficult to remove only part of a widely projecting pathway or a complex structure in order to investigate function in detail, it may be possible to achieve the same end by destroying the entire structure or pathway and then reconstructing with grafted tissue the part that is of interest.



FEATURE OF ALZHEIMER'S DISEASE, a loss of acetylcholine from the hippocampus, can be simulated in the rat by cutting the cholinergic fibers leading from the medial septal area to the hippocampus. Whereas normal rats can be trained with rewards of food to enter opposite arms of a T-maze in strict alternation, the rats with cut fibers do not learn to alternate. Instead they reenter the same side of the maze several times in a row. Björklund and his colleagues found that when embryonic neurons taken from the medial septal area are injected into the hippocampus of such rats, they can again learn to alternate sides.

Denervation supersensitivity by itself is not enough to account for the substantial behavioral effects of neuronal grafts. In the normal brain many nerve cells relay precisely patterned messages in response to appropriate inputs from other neurons. Grafted neurons often are transplanted not to their normal site but to a distant target, and it is not known whether they establish a normal pattern of synapses: the zones of communication with target neurons at which most neurons release neurotransmitter. If transplanted neurons do not communicate specifically with host neurons, how might transplants restore any impaired behavior at all?

Their efficacy suggests that dopamine, acetylcholine and presumably other neurotransmitters often act not to relay critically patterned messages but to modulate the general level or nature of target cells' activity or their response to other inputs. Given those conditions a neurotransmitter might be effective even if its release is largely unregulated and it reaches its target by gradual diffusion.

In contrast to neurotransmitters, the hormones that are secreted by certain structures in the brain normally travel considerable distances to their targets, and their release may be regulated not by signals from adjacent cells but by diffusible factors carried in the blood or released by other brain structures. Neuroendocrine diseases, the result of deficiencies in particular brain hormones, are therefore promising candidates for treatment with neural grafts. The first such disease in which the technique has been tried in the laboratory is diabetes insipidus. It is characterized by excessive thirst and urination. The symptoms reflect a lack of vasopressin, a hormone that regulates fluid balance and blood pressure by directing the kidney to concentrate urine and causing the peripheral blood vessels to constrict.

The hormone, which also seems to exert behavioral effects in the brain, is made by neurons in the front of the hypothalamus, a brain structure that regulates many physiological functions. The neurons send axons into the posterior part of the pituitary gland, where the vasopressin is released. Diabetes insipidus in humans almost always results from tumors or trauma affecting the pituitary, but a strain of laboratory rats displays an inherited form of the disease. They are born without the vasopressin-producing neurons.

Don M. Gash and John R. Sladek, Jr., of the University of Rochester School of Medicine and Dentistry transplanted fragments of embryonic hypothalamus to a brain ventricle in affected rats. In about 25 percent of the cases the workers observed a sustained improvement in fluid balance: water intake declined and urine became concentrated. Six months after the transplantation the workers found many vasopressin-containing neurons surviving within the grafts.

The neuroendocrine disease known as hypogonadotrophic hypogonadism or Kallmann's syndrome also has a laboratory model. The disease, an inherited condition associated with the *X* sex chromosome, prevents sexual maturation in boys. The missing hormone is gonadotropin-releasing hormone (GnRH). Made in the hypothalamus, GnRH causes the pituitary to secrete other hormones, which in turn bring about the maturation of the testes and the production of testosterone, the male sex hormone.

The same condition occurs in certain mutant mice. The late Dorothy T. Krieger, working with Perlow and Marie J. Gibson of the Mount Sinai School of Medicine in New York, Harry M. Charlton of the University of Oxford and other investigators, treated hypogonadal mice by transplanting GnRH-containing neurons from embryos to a cerebral ventricle, close to the cells' normal position in

the hypothalamus. Within two months the recipients' testes had partially matured and were making normal sperm. A later examination of the mouse brains showed a luxuriant growth of fibers extending from the transplanted neurons toward their normal targets: specialized capillaries that transport GnRH to the pituitary gland.

The root of a hormonal deficiency need not lie within the brain for an intracerebral transplant to have potential value. Efforts to treat endocrine disorders such as diabetes mellitus by transplanting hormone-secreting tissue to other sites in experimental animals or human patients fail unless the subjects are given strong immunosuppressive drug treatment to prevent graft rejection. Implanted in the immunologically privileged confines of the brain, however, such tissue might survive and function. The circulating chemical stimuli that trigger hormone secretion are present in the cerebrospinal fluid that fills the brain ventricles, and secreted hormones could return to the circulatory system by way of the special cellular membranes that transport minute droplets of cerebrospinal fluid to the blood.

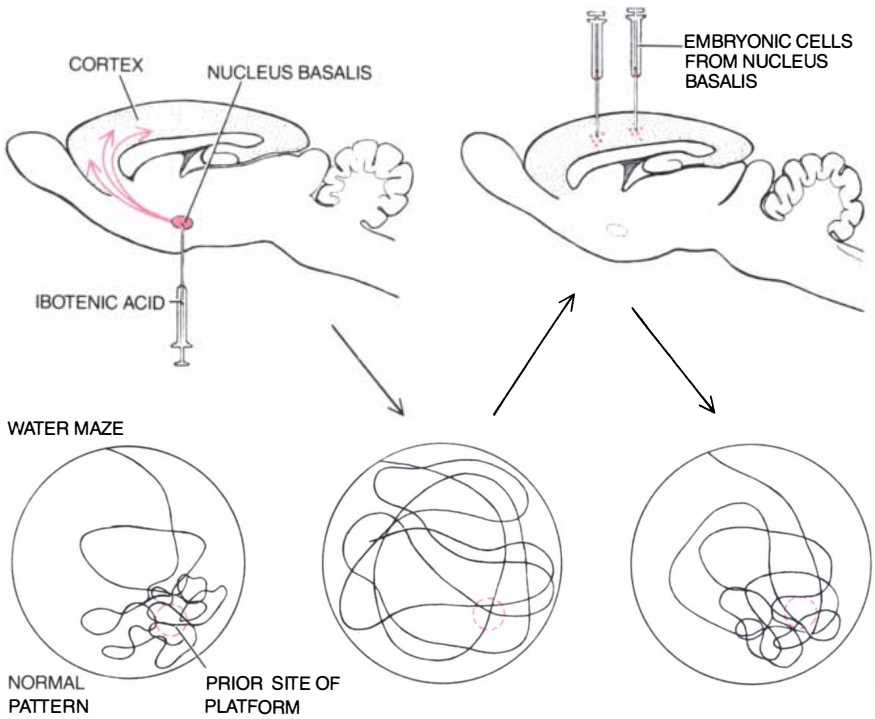
The idea was tested more than 40 years ago by Charles M. Pomerat and his co-workers at the University of Alabama. They transplanted tissue from the adrenal gland of newborn rats to the brain of adult rats whose adrenals had been removed. The grafted tissue survived and matured, and for more than eight months it corrected the hormonal deficiencies brought on by the glands' removal.

Recently I and, independently, Wah Jun Tze and Joseph Tai of the University of British Columbia reversed chemically induced diabetes mellitus in rats by transplanting insulin-secreting cells from the pancreas of other rats to the brain of the diabetic animals. Tze and Tai have shown that such transplants are also effective in the spontaneously occurring diabetes, found in a particular rat strain, that is considered the closest animal analogue of human juvenile-onset, insulin-dependent diabetes.

As understanding grows of the biology of transplants in the mammalian central nervous system, their potential to halt its deterioration or restore lost capacities will be tested in a growing number of disorders. Together with Francesco Scaravilli of the National Hospital in London, for example, I am exploring the possible value of such transplants in the hereditary metabolic defects known as lysosomal storage disorders. Caused by deficiencies in enzymes needed to break down complex molecules within the cell, the disorders often lead to severe neurological deterioration. Appropriate cells transplanted to the brain might replace the missing enzymes and prevent the degeneration. In other work, undertaken with Brian Meldrum and Smita Patel of the Institute of Psychiatry in London and Harry Robertson and Georgia Cottrell of Dalhousie University, I am studying the ability of transplanted inhibitory neurons to suppress epileptic seizures induced experimentally in rats.

The work I have described is only a first step toward the development of reliable therapies for human diseases; the procedures that have succeeded in rats are only now being tried in monkeys, in several laboratories including my own. Yet a sense that it is urgent to attempt neuronal transplants in human beings is widely felt. Many of the human disorders that have been simulated and treated by transplantation in experimental animals are currently incurable. The symptoms of Parkinson's disease can often be alleviated by the drug L-dopa, but the fatal progress of the disease continues. For Alzheimer's disease there is no such palliative. Olson and Seiger, working with Erik-Olof Backlund of the Karolinska Hospital and Institute of Stockholm, have already transplanted cells from the adrenal glands of four patients with severe Parkinson's disease into their brains. None of the patients experienced a significant and lasting recovery, although several of them showed temporary improvements.

For the moment transplantation in the central nervous system of human beings poses an ethical problem. Should experimental procedures that are shown to be successful in imperfect animal models but are unproved in primates and that carry unknown but perhaps serious risks be used to treat patients with progressive and fatal disease? The issue warrants wider consideration than it has received so far. Further ethical questions will arise if experiments in primate models of human disease clearly establish the value of the procedures. Nerve cells derived from certain tumors can be grown in culture and have been transplanted successfully into rodents. The likeliest source of embryonic neurons for transplantation to human beings, however, appears to be tissue from aborted fetuses.



ANOTHER FEATURE OF ALZHEIMER'S DISEASE, the depletion of acetylcholine in the cortex, was simulated by the author and his colleagues. An injection of ibotenic acid destroyed the large cells of the nucleus basalis, which send cholinergic fibers to the cortex. A water maze revealed the resulting memory deficit. A normal rat, trained to find a submerged platform, will swim over the site as though searching for the platform even if it is removed. The path of an injected rat, in contrast, suggests the animal does not remember the platform's original position. After transplantation of embryonic cells from the nucleus basalis to the cortex, such a rat can again be trained to remember the site of the platform.

Pleasure Trip

Don't let radar spoil it: Remember your Passport

Imagine setting out on a long day's drive and not having to worry about radar. Seeing America might be fun again if you didn't have to watch it unfold in the rearview mirror.

PASSPORT means low-profile protection

PASSPORT has exactly what the vacationing driver needs: advanced radar protection, yet none of the nuisances you've come to expect of ordinary radar detectors.

The experts at *Car and Driver* magazine gave us a rave review: "In a word, the Passport is a winner."

Our customers agree, even though we allow them ample opportunity to do otherwise. Every Passport is sold with the promise that it will live up to the customer's highest hopes, not just our claims. If it doesn't satisfy within the first 30 days, it can be sent back for a full refund.

PASSPORT has been on the market just over a year and already our money-back offer has been declined a quarter of a million times. So we'll stand on our record.

PASSPORT means the size of a cassette tape

PASSPORT provides superheterodyne performance in a package the size of a cassette tape. This miniaturization is possible only with SMDs (Surface Mounted Devices), microelectronics common in satellites but unprecedented in radar detectors.

All you do is clip PASSPORT to your dashtop, windshield or visor, plug it in, and adjust the volume. Upon radar contact, the alert lamp glows and the variable-pulse audio begins a slow warning: "beep" for X band radar, "brap" for K band. Simultaneously a bar graph of Hewlett-Packard LEDs shows radar proximity.

PASSPORT comes complete with all accessories.

As you get closer, the pulse quickens and the bar graph lengthens. PASSPORT watches for radar so you can watch the road.

PASSPORT comes with a leather travel case and everything necessary to get you on the road in seconds, no matter what car you drive. And one more convenience – we deliver. We make PASSPORT in our own factory in Cincinnati and we sell direct to you.

PASSPORT means an easy fit in your travel case

Call toll free. We'll answer your questions. If you decide to buy, we'll ship by UPS the next business day at our expense. For $6.00 extra, Federal Express will deliver to you within two business days of shipment.

And once again, we make this promise: if you're not satisfied within 30 days, return PASSPORT. We'll refund your purchase and your shipping costs. There are no hidden charges. Moreover, PASSPORT is backed by a one-year limited warranty.

Next time you set out on a pleasure trip, don't forget your PASSPORT.

$295 (OH res. add $16.23 tax)
Slightly higher in Canada

VISA MasterCard

PASSPORT®
RADAR • RECEIVER

Call Toll Free 800-543-1608

Cincinnati Microwave
Department 9168
One Microwave Plaza
Cincinnati, Ohio 45296-0100

AT&T Bell Laboratories scientists have generated a beam from a diode laser with a frequency spectrum 1,000,000 times narrower than that of today's most advanced commercial semiconductor lasers.

Part of a record-breaking coherent lightwave system, this laser 'peels away' a tremendous number of unwanted frequencies that can clutter up a beam—to create a lightwave so pure, its frequency variance is limited to 1 part in 40 billion. The equivalent of one drop of ink in a million-gallon, Olympic-size swimming pool.

A drop in 40,000,000,000

Purity Has Its Rewards

The new narrow-spectrum laser is the key element in a coherent lightwave communications system that increases the information-carrying capacity of an optical fiber, as well as the distance over which an unboosted laser beam can be received.

Increased capacity comes from dramatically increasing the number of individual laser beams that can be sent through a fiber's best transmission window.

Each laser in a coherent system produces an exceptionally stable, pure wavelength—allowing thousands of non-interfering wavelengths to travel side-by-side on a fiber. (Only a handful of beams can be combined using today's commercial semiconductor lasers.)

Dial 'M' For Movie

In the future, the capacity of coherent transmission could allow us to send 10 million conversations—or 10 thousand digital TV channels—simultaneously, on a single fiber.

Or, using the full capacity of a fiber, a coherent system could dump a movie like 'Gone with the Wind' into a home memory unit in one second flat. Or deliver Beethoven's '5th' in less than a 50th of a second.

Making A Little Go A Longer Way

Increasing capacity is important in an age of rapidly expanding information needs. But so is reducing costs—in this case, by nearly doubling the distance an unboosted signal can be received.

A newly developed AT&T coherent lightwave receiver contains its own narrow-spectrum laser. The beam from this laser reinforces the transmitted signal as it detects it—a technique only possible with two such pure beams.

Using this receiver, AT&T has achieved a laboratory transmission record of nearly 100 miles at a data rate of 1 billion pulses per second.

We Don't Keep The Future Waiting

Coherent lightwave transmission is just one of the ways AT&T is working toward the high-capacity, high-speed integrated networks of the future.

Meanwhile, we're bringing tomorrow closer with leading-edge lightwave systems we're building today.

AT&T this year introduced a commercial lightwave system—the FT Series G—designed to operate at up to 1.7 billion bits per second, a rate that permits the transmission of 24 thousand simultaneous calls on a single pair of fibers.

And by 1988, we'll have installed the first transatlantic and transpacific lightwave systems to Europe and the Far East—systems capable of transmitting 40 thousand simultaneous conversations on two pairs of fibers.

Clearly, whether on land or underseas, AT&T is lighting the way in lightwave. And peeling the laser beam is part of it.

AT&T

The right choice.

SCIENCE AND THE CITIZEN

The Heart of the Matter

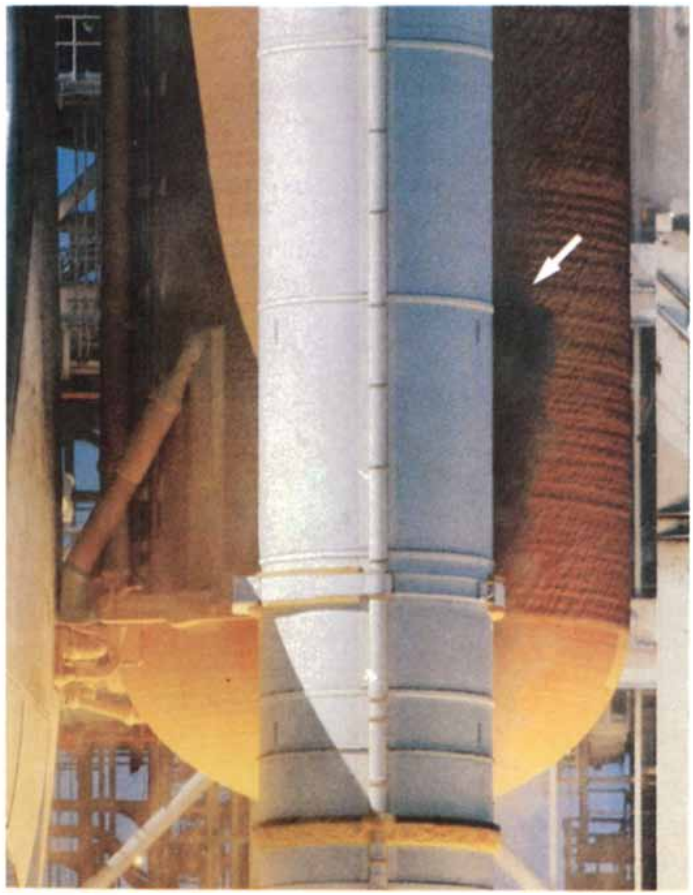
In its final report the presidential commission charged with examining the explosion of the space shuttle *Challenger* identifies the design flaw that caused the accident and describes in detail the events leading up to the tragedy. It does not describe the underlying causes, within the organization of the National Aeronautics and Space Administration, that made it possible for serious dangers to be ignored. One of the commissioners, Richard P. Feynman of the California Institute of Technology, has addressed this question in a separate document, in which he treats the errors in judgment and execution discovered by the commission as symptoms by which to diagnose larger problems within the space agency. He concludes that NASA's effectiveness in selling its projects to Congress has interfered with its effectiveness as a science and engineering agency.

It is well known by now that the immediate cause of the accident was found to be a faulty seal in one of the joints between sections of the shuttle's right-hand solid-rocket booster. Hot gases eroded a rubber O-ring in the seal and "blew by" it, creating a leak that eventually allowed a plume of flame to escape through the joint and pierce the shuttle's external fuel tank.

The finding came as no surprise. Testimony before the Rogers commission (named for its chairman, William P. Rogers, a former secretary of state) revealed that O-rings in the solid-rocket boosters had been a matter of concern for nearly a decade. Seals are an essential part of the boosters because, like all large solid rockets, they are built in sections. There are several reasons for such a design. One is that the fuel is first cast as a liquid, and it might not dry and cure correctly if it were deposited in a single container as large as the shuttle booster. Another reason is that an intact booster rocket would be too large to be transported by rail from the manufacturer to the launch site; because the boosters were made in landlocked Utah, no other means of transportation was available. The particular method of joining the sections that was proposed by the manufacturer, Morton Thiokol, Inc., had been criticized by NASA, however. That was in 1977, when tests first indicated that Thiokol's method of sealing the joints between sections might lead to erosion and leaks.

During the second flight of the space shuttle,in November, 1981, one O-ring in the right-hand solid-rocket booster was eroded, although no gases blew by it. O-rings were eroded during 11 subsequent flights—often in more than one joint—and in nine flights hot gases blew by the "primary" O-ring in at least one joint but did not pass completely through the rest of the seal.

Engineers at Thiokol were alarmed by the unexpected frailty of the seals. In July, 1985, Roger M. Boisjoly, a Thiokol engineer, wrote a memorandum to Robert K. Lund, the vice-president of engineering, "to insure that



FAILURE of a seal doomed the space shuttle while it was still on the launch pad. Smoke leaked from the right-hand booster rocket and spurted upward .678 second after the rocket was ignited. In the first photograph (*left*) the leaking joint is at the bottom of the

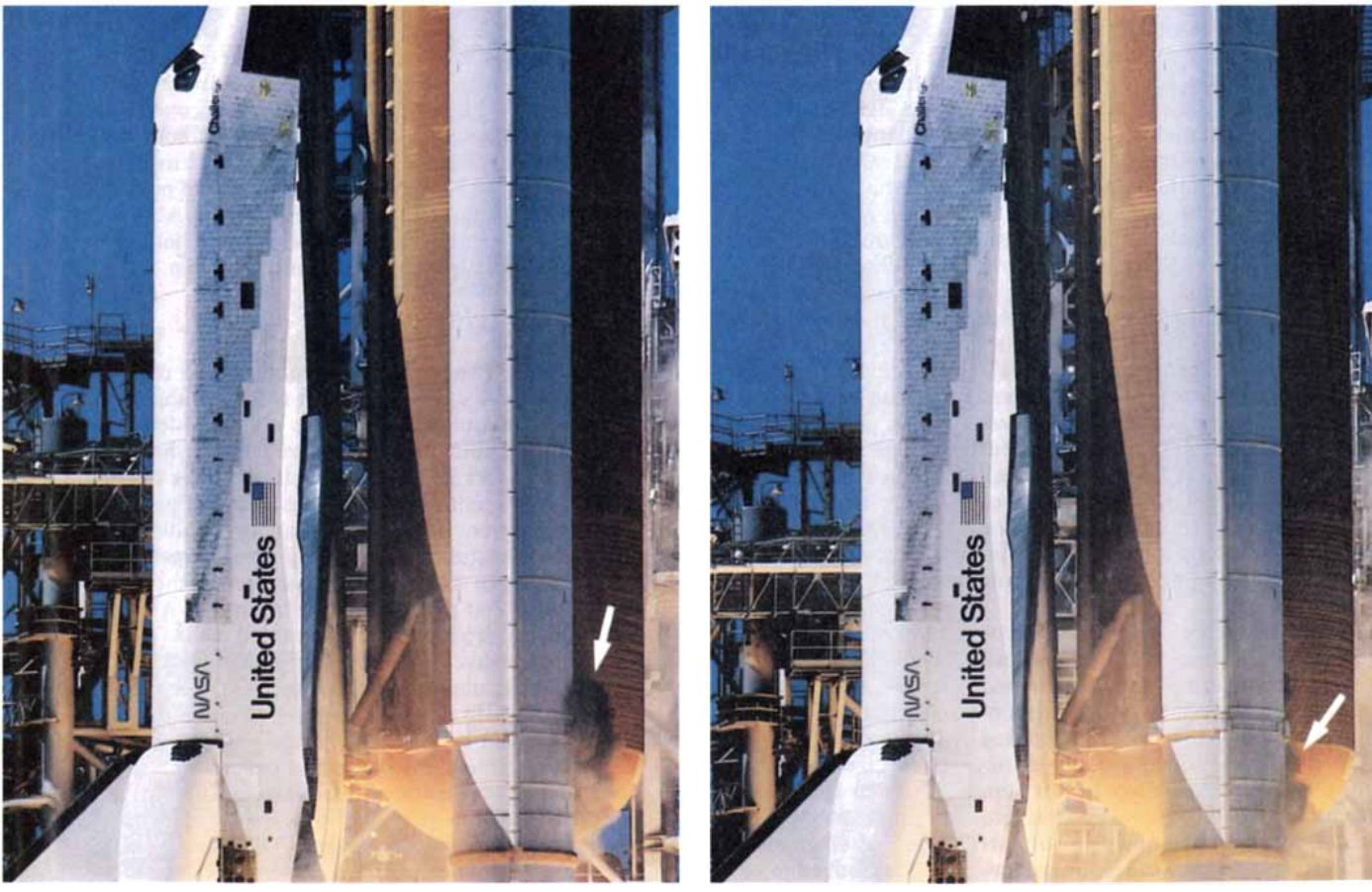
management is fully aware of the seriousness of the current O-ring erosion problem.... It is my honest and very real fear that if we do not take immediate action to dedicate a team to solve the problem...we stand in jeopardy of losing a flight along with all the launch pad facilities." A later memorandum, written in October, 1985, by the head of the task force eventually created to solve the problem, begins with the word "HELP!" and ends with "This is a red flag." The engineers' concern came to a head the night before the *Challenger* launch, when, in a teleconference, they tried to convince both the NASA and Thiokol managements not to launch because of the extremely cold temperatures at the launch pad.

Why were shuttles allowed to fly when critical parts were being damaged in unexpected ways? According to Feynman, managers at NASA and at Thiokol came to regard O-ring erosion as an acceptable risk because O-rings had eroded on previous flights without causing the boosters to fail. Officials noted that in the earlier flights the rings had been eroded by no more than one-third of their radius. Experiments had indicated that an O-ring would have to be eroded by one full radius before it would fail, and so the officials asserted that there was "a safety factor of three." Feynman observes, "This is a strange use of the engineer's term 'safety factor.'... Erosion was a clue that something was wrong. Erosion was not something from which safety can be inferred."

Officials tried to understand the erosion by making a mathematical model, based on data from flights on which the O-rings eroded, to predict the amount of damage to be expected under various conditions. Feynman discusses the way the model was developed and the final form it took and then adds: "There is nothing much so wrong with this as believing the answer! Uncertainties appear everywhere.... The empirical formula was known to be uncertain, for it did not go through the very data points by which it was determined." NASA used this mathematical model to rationalize flying with ever greater risks. Feynman also discusses the design, testing and certification of the shuttle's main liquid-fuel engine and concludes that here too there was a "slow shift toward decreasing safety factor." In these and other cases, "subtly, and often with apparently logical arguments, the criteria are altered so that flights may still be certified in time."

To estimate the chances of a space shuttle's failing, NASA managers substituted what they termed "engineering judgment" for the standard methods of probability. They set the probability of failure at about one chance in 100,000. Working engineers thought the chances were closer to one in 100. "If we are to replace standard numerical probability usage with engineering judgment," Feynman asks, "why do we find such an enormous disparity between the management estimate and the judgment of the engineers?"

Feynman hypothesizes that the fundamental cause of NASA's systemic overconfidence was that a major role of NASA management was to get funding from Congress. To do so, he says, they painted too rosy a picture of what could be accomplished with current technology. At a press conference held when he released his independent remarks, Feynman speculated that "by exaggerating what they said they could do, they got in a position where they didn't want to hear too much about the



smoke cloud, near the top of the hemispherical section of the orange external fuel tank. In successive photographs the puff of smoke hangs in the air as the shuttle, tracked by NASA's camera, ascends. Photographs are from the report of the presidential commission.

truth.... The *Challenger* mission was the final accident of a sequence of things in which there was warning after warning after warning that something was wrong.... For 10 years they discussed this problem and didn't do anything about it...because it was hard for information to come up. But we know the information was there at the lowest levels. Why the engineers are at the lowest levels I have no idea, but the guys who know something about what the world is really like are at the lowest levels of these organizations and the ones who know how to influence other people by telling them how the world would be nice...they're at the top."

Although Feynman judges NASA management more harshly than the official report, the latter does suggest that NASA's original plans for the shuttle were overambitious: the commitment to provide routine and economical access to space locked the agency into a schedule too tight to be met with the available resources. For example, the inventory of spare parts was not large enough to accommodate the launch schedule, and so each orbiter was made ready for launch by cannibalizing parts from other orbiters. The commission suggests that NASA's desire to make the shuttle the only major U.S. launch system put too much pressure on the program to meet tight schedules and to be able to handle any payload. NASA's can-do attitude, its willingness to undertake challenging tasks at the last minute, also strained the resources of the ground crews and forced NASA officials to focus on the near term at the expense of long-term safety and economy.

Yet the report does not recommend any major changes in the overall structure of the space program, nor does it hold the highest levels of management responsible for the accident; it reserves its strongest criticism for management at Thiokol and at NASA's Marshall Space Flight Center, the division of NASA responsible for the boosters. The report concludes by urging the Administration and the country to continue supporting NASA.

Feynman's report goes on to draw the connection between the overoptimistic attitude of top management and the accident. He concludes by admonishing NASA to be realistic in estimating costs and setting schedules. "If in this way the Government would not support them, then so be it. NASA owes it to the citizens from whom it asks support to be frank, honest and communicative, so that these citizens can make the wisest decisions for the use of their limited resources." His final remark is that of a physicist who is galled to see what he calls "fantasy" enter the realm of science and engineering: "For a successful technology, reality must take precedence over public relations, for nature cannot be fooled."

BIOLOGY

Body and Soul

The notion that a good mental attitude can help the body to fight off disease and that depression can help cause the body to succumb suggests that the nervous system may influence cells that mount the immune response. Some tantalizing linkages are known. Certain neurotransmitters (the substances that mediate brain activity) can affect the activity of some lymphocytes, the white blood cells that defend against invasion. Moreover, certain lymphocytes are known to have receptors that are sensitive to neurotransmitters.

The converse possibility has now emerged. The lymphocytes designated as helper *T* cells synthesize a protein found previously in the brain—a protein that incorporates several copies of the neurotransmitter known as enkephalin. It is conceivable that the enkephalin is a blood-borne signal the helper *T* cells dispatch to neural targets. That is, the immune system may influence the nervous system.

Gerard Zurawski, John S. Abrams and their colleagues report the discovery in *Science*. Working at the DNAX Research Institute of Molecular and Cellular Biology, operated in Palo Alto by the Schering-Plough Corporation, the investigators studied mouse helper *T* cells. Activation with the protein called concanavalin *A* induced the cells to secrete a variety of lymphokines, proteins that coordinate the activities of other cellular participants in the immune response. The investigators analyzed the messenger RNA (mRNA) they found in the activated cells but not in resting ones. Each mRNA carries instructions for the synthesis of a protein; the mRNA's present only in activated cells might yield the identity of some of the unknown chemical signals that helper *T* cells manufacture.

One such messenger was intriguing. It specified the structure of a protein in which the amino acids lysine and arginine made frequent appearances. They often mark sites where a protein is cleaved into smaller molecules. The enkephalins (which are peptides, or short chains of amino acids) also appeared frequently. The messenger proved to encode (with 97 percent faithfulness) the structure of rat preproenkephalin, a protein found in the brain and also in the medulla, or core, of the adrenal gland. It incorporates seven enkephalin molecules.

The messenger was notably abundant. In nerve cells that synthesize preproenkephalin (and apparently cleave it into enkephalin) it accounts for less than one out of every 3,000 mRNA molecules; in adrenal cells that synthesize preproenkephalin it accounts for about one out of every 1,000. Yet in the helper *T* cells it appeared to account for one out of every 250; its abundance was comparable with that of the messengers for known lymphokines. Evidently the activated helper *T* cells make significant amounts of preproenkephalin. Perhaps they respond to an invasion of the body by secreting it, or enkephalins themselves, into the bloodstream along with the known lymphokines and other proteins still to be discovered.

Zurawski, Abrams and their colleagues cannot yet specify a function for the secretion. Enkephalin may simply serve the immune system as another, newly identified lymphokine; that is, it may serve locally to coordinate the activity of other lymphocytes. On the other hand, enkephalin could serve as a signal to parts of the peripheral nervous system, such as sensory fibers or autonomic ganglia. Moreover, although the brain and spinal cord are isolated from much of what travels in the blood, it may be unwise to neglect them as candidate targets. Abrams notes occasional findings that helper *T* cells somehow get into the brain.

Telltale Cells

The effects of a mutagen—a chemical or a form of radiation that can damage the genetic material—can in principle be measured in the laboratory by exposing cultured cells to the agent and examining later cell generations for genetic changes. Traditional techniques for gauging mutagenicity are hampered, however, by the fact that genetic change often kills a large fraction of the mutated cells. They leave no altered progeny, and their mutations go undetected.

In *Proceedings of the National Academy of Sciences* Charles Waldren, Carol Jones, Theodore T. Puck and their coworkers at the Eleanor Roosevelt Institute for Cancer Research in Denver and the University of Colorado Health Sciences Center report a new technique that eliminates much of the bias introduced by cell death and is considerably more sensitive than its prede-

cessors. Whereas test cells previously had to be given large doses of radiation or chemical mutagens to allow for cell death and yet produce a countable number of mutations, the new method can yield meaningful results directly at low doses in the range to which human beings are often exposed.

All such techniques make use of one or more genetic markers: heritable traits whose presence or absence can be detected in descendants of the exposed cells. The Denver workers rely on a set of surface proteins borne by all the cells in the initial population. Following exposure the workers test for missing marker proteins in order to determine the frequency of change in the genes encoding them. Antibodies that bind to the proteins act as probes.

The crucial contrast between the new technique and its forerunners is the site of the marker genes. Earlier methods relied on marker genes carried on chromosomes essential for cell reproduction. The cell might survive a mutation affecting only the marker gene but would probably die after larger changes such as the deletion of the chromosome segment that includes the gene or the outright loss of the chromosome. The Denver group has developed hybrid cells containing a full set of Chinese hamster chromosomes and an extra, human chromosome bearing several genetic markers. Because the hamster chromosomes carry all the genetic machinery needed for reproduction, the cell can live to record mutations as large as even the loss of the entire human chromosome.

The new technique makes it possible to distinguish between such large-scale damage and smaller changes. The loss of all the marker proteins encoded by the human chromosome suggests the mutation eliminated the entire chromosome; the absence of two markers specified by closely spaced genes suggests the deletion of a chromosome segment, and the loss of only a single marker suggests a localized mutation.

The results obtained at low doses of X rays testify to the method's sensitivity. In the past investigators estimated the mutagenicity of low doses by extrapolating linearly down from the mutation rates measured at high doses. It now appears that such extrapolations drastically underestimated the effects of the smaller exposures because cell death had skewed the existing data. The authors found that mammalian cells treated with small amounts of X rays mutate at least 200 times more often than was thought.

The workers hope their technique will lead to a clearer appraisal of the health effects of chronic, low-level exposure to radiation and chemical mutagens. They suggest the method could also serve as an assay for potential prophylactic agents that would protect cells against mutagens and conceivably lower the incidence of cancer and certain genetic diseases.

MEDICINE

Sweet News

An experimental drug tested in animals has arrested a biochemical process thought to contribute to several debilitating conditions associated with diabetes. The process may also figure in the genesis of such effects of aging as cardiovascular disease and cataracts. The investigators, Michael Brownlee, Helen Vlassara, Anthony Kooney, Peter Ulrich and Anthony Cerami of Rockefeller University, hope to begin clinical testing of the drug, aminoguanidine, within a year.

Aminoguanidine appears to prevent the formation of potentially destructive cross-links, or tight chemical bonds, between proteins. The steps leading to this cross-linking begin when glucose, the major sugar in the blood, enters the bloodstream and attaches to a protein. If glucose attaches to a long-lived protein, such as collagen or other components of structural tissue, the resulting molecule eventually undergoes a slow rearrangement into an "advanced glycosylation product." Such products tend to form permanent cross-links with other nearby protein-containing molecules.

The investigators think that in the course of several years the accumulation of these cross-linked products can damage blood vessels and other tissues. In the arteries, for example, the advanced glycosylation product formed with collagen can cross-link with other collagen molecules and can also trap proteins from plasma and cross-link with them. Among these plasma proteins are low-density lipoproteins, which contain cholesterol and are a major component of atherosclerotic plaque. Eventually the accumulation of proteins and cholesterol on the vessel wall can stiffen the vessel and clog it.

The cross-linking of proteins may also lead to cataract formation in the lens of the eye, wrinkling of the skin and stiffening of the heart muscle, lungs and tendons around the joints. Atherosclerosis, cataracts and tissue stiffening, all associated with aging in the normal population, tend to occur prematurely in people with diabetes—very possibly, Brownlee says, because diabetes elevates blood-glucose levels, which in turn elevates the rate of protein glycosylation. Brownlee thinks accelerated glycosylation and cross-linking may also result in changes observed only in the capillaries of people with diabetes. In the eye these "diabetic" changes can lead to blindness, in the kidneys to organ failure.

The workers have so far focused their attention on interrupting harmful cross-linking in blood vessels. They chose aminoguanidine for the purpose because laboratory studies had indicated that it binds to the reactive fraction of advanced glycosylation products, thereby blocking their ability to interact with nearby proteins.

In their recent animal studies, which were reported in *Science* and at the annual meeting of the American Diabetes Association, the investigators compared the arteries of untreated diabetic rats with those of diabetic rats given aminoguanidine for 16 weeks. The treated animals' arteries contained significantly lower amounts of cross-linked collagen and collagen bound to lipoproteins. The treated rats also showed less binding of plasma proteins to the walls of kidney capillaries. In the future, Brownlee says, the drug should enable workers to design long-term studies to determine whether advanced glycosylation products do in fact contribute to vascular disease and whether medication can prevent its development in human beings.

Mood and Food

People become obese because they take in more calories than they burn. What makes someone overeat? Evidence cited at a recent conference sponsored by the New York Academy of Sciences suggests that a subgroup of obese people—"carbohydrate cravers"—may overeat in an unwitting attempt to medicate themselves when they feel depressed.

The conference organizers, Richard J. Wurtman and Judith J. Wurtman of the Massachusetts Institute of Technology, recently found that subjects who normally binge on carbohydrates feel more vigorous and less depressed after being given a carbohydrate meal than do people who normally choose other kinds of snacks. The Wurtmans suspect that the well-being experienced by the carbohydrate cravers in the study can be explained by an increased level of the brain chemical serotonin. The production of serotonin, which affects both satiety and mood, increases after a meal or a snack that is primarily carbohydrate. Conceivably, the Wurtmans say, car-

bohydrate cravers—who may account for about half of all obese people—overeat to achieve serotonin-induced mood changes.

Support for a link between mood and excessive carbohydrate consumption has recently come from investigations into depression and its treatment. For instance, Norman E. Rosenthal and his colleagues at the National Institute of Mental Health report that about two-thirds of people who have a recently recognized syndrome, seasonal affective disorder (SAD), crave carbohydrates and gain weight when they are depressed. SAD victims have an exaggerated response to season changes, becoming depressed, lethargic and withdrawn in winter and becoming happier and more energetic in spring and summer. Many SAD victims also lose weight in the spring and, according to preliminary evidence from the Wurtmans, choose fewer carbohydrate snacks in that season than they do in the winter.

Other studies, involving antidepressant drugs that increase the amounts of serotonin in brain synapses, add more evidence to the mood-food connection—and raise the possibility of new treatments. In France, where one such drug, d-fenfluramine, is on the market, Christine Nathan and her colleagues at the International Research Institute Servier, have shown that d-fenfluramine leads to a decrease in stress-induced eating and a decrease in carbohydrate snacking by obese carbohydrate cravers; it also leads to weight loss. Similarly, the Wurtmans and their colleague Dermot O'Rourke have found that d-fenfluramine eased both winter depressive symptoms and the tendency to binge on carbohydrates in eight SAD patients. Other investigators have found that an experimental antidepressant called fluoxetine could produce weight loss in overweight subjects.

Carbohydrate cravers are, of course, just one possible subgroup of people who become obese. Indeed, many conference participants emphasized the importance of recognizing that obesity can no longer be viewed as a single disorder with a single cause. Among the many potential causes might be a defect in the production of serotonin or of a range of other substances that influence satiety or food choice, as well as a decreased rate of calorie burning in certain people.

Clarinetist's Thumb

A few months ago a professional violinist came to Michael E. Charness, a neurologist at the University of California at San Francisco, complaining of a neck pain and numbness in her playing arm. Charness watched her play for an hour and saw that she held her head in exactly the position physicians choose to induce symptoms of a pinched nerve in the neck. He recommended that she change her playing position. Two months later she was able to play for an extended period without pain.

The violinist's affliction and many others, some merely annoying to most people but career-threatening to performing artists, have given rise to a number of clinics devoted to illnesses to which musicians, dancers and other performing artists are susceptible. Massachusetts General Hospital, the Hospital for Joint Diseases in New York, the Cleveland Clinic Foundation and the University of California at San Francisco have all established special teams for the purpose in recent years. There is even a new medical journal, *Medical Problems of Performing Artists.*

Often the problem arises from overuse of a muscle, tendon or joint. A clarinetist supports the instrument by means of a hooklike device that rests on the right thumb; the result can be pain in the middle joint. A pianist is likely to have trouble with an elbow, a wrist or a hand. A singer or an actor can suffer from polyps or strained vocal chords, a dancer from a variety of leg injuries.

Robert D. Leffert, an orthopedist who is part of the special team treating performing artists at Massachusetts General, recently discussed the problems in *Harvard Medical School Health Letter.* "A physician who sees musicians only occasionally," he wrote, "may fail to recognize that there is a problem when he or she first hears the performer's account of it. And it may be phrased in a rather peculiar way, as when a pianist says, 'I notice a bogginess when I attempt to play the Beethoven E-major sonata.' This translates to: 'I notice a new stiffness in my right hand when I try rapidly alternating two keys.' When the performer and physician lack a common language, treatment is difficult enough, and if the doctor has a preconceived notion that artists have eccentric personalities, it can be very difficult for the musician to persuade him or her that the problem is real, and not merely neurotic, or an excuse for an inability to perform."

Once the problem is recognized, treatment can range from surgery to a prescription for anti-inflammatory drugs or physical therapy to rest or changes in playing style. The patient is not the only one helped. According to Frank R. Wilson, a neurologist in the San Francisco clinic, the work with performing artists "can tell us about how the brain and musculoskeletal system work together."

PHYSICS

Sound Reasoning

For the past three decades the speed of sound in dry, open air under so-called standard conditions (at a temperature of zero degrees Celsius and a pressure of one atmosphere) has been listed in handbooks as 331.45 meters per second (741.43 miles per hour). Most scientists and engineers have taken this value for granted. George S. K. Wong of the National Research Council of Canada did not. The value looked suspect to him, and thorough investigation confirmed his suspicions.

Sound is essentially a pressure disturbance propagating through a medium. As such, a formula for its speed of propagation in a gas can be rather simply derived from the thermodynamic properties of the gas (its behavior as a function of pressure, volume and temperature). Indeed, the French mathematician Pierre Simon de Laplace was the first to state such a formula about 200 years ago. If the constants and parameters that describe the thermodynamic properties of air are accurately known, Laplace's formula determines the speed of sound in air with remarkable precision.

Wong recognized that most of the values for those constants and parameters had been revised since the early 1950's, when what have been thought to be the definitive investigations into the matter were made; he was curious to see whether the calculated speed of sound based on the latest values was in agreement with the accepted speed. When he substituted the new values into Laplace's formula, he found the speed of sound in dry air under standard conditions to be 331.29 m.p.s., or 741.07 m.p.h.—.05 percent less than the accepted speed.

Yet the accepted speed had been determined from numerous direct measurements, which were thought to have an average estimated uncertainty of only .02 percent. Could their uncertainty really have been underestimated for 30 or more years? As Wong writes in *Journal of the Acoustical Society of America,* he scrutinized more than 30 reports on the measurement of the speed of sound published in the past 100 years. In virtually all of them he found sources of uncertainty in the specification of such critical factors as air composition and humidity, barometric pressure, temperature and even

unit definitions and conversions; the methods used to reduce the measured results (obtained for the most part at room temperature) to standard conditions were also often questionable.

Wong decided that the uncertainty in every speed-gauging experiment was likely to have been larger than what was cited in the reports. He concluded that his recalculated, theoretical value is actually well within the bounds of experimental error. Until more accurate speed measurements disprove it, his value should be considered as the "true" speed of sound.

Although the more precise value for the speed of sound may not have much impact on the average person, Wong's finding will necessitate some activity among scientists, engineers and their publishers: textbooks will have to be updated, formulas will have to be revised and laboratory acoustical equipment will have to be recalibrated.

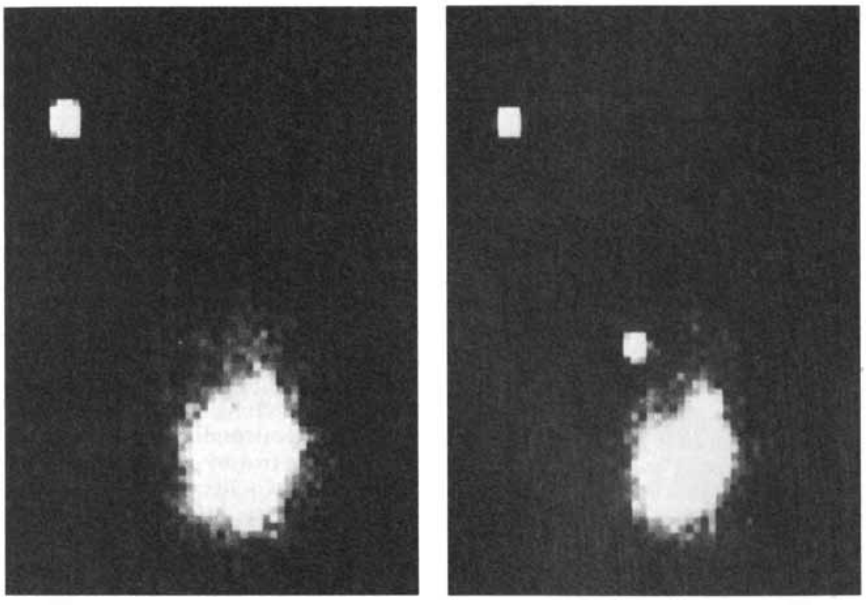
Shadow Play

What drew matter together in the early universe to form galaxies and clusters of galaxies? Prevailing theory, which holds gravity responsible, has been challenged by two investigators at the University of Arizona. Craig J. Hogan and Simon D. M. White write in *Nature* that "mock gravity," an effect of radiation pressure, may have precipitated the coalescence instead.

Radiation pressure is an analogue of the more familiar, readily experienced kinds of pressure. Just as expanding gas in a cylinder head exerts force on the vessel's walls and will cause a piston to move if the pressure is sufficiently high, so light exerts force on particles and will cause them to move if the radiation is sufficiently intense. On the earth radiation pressure is negligible compared with the strength of gravity. In deep space radiation pressure assumes increasing importance.

If two particles lie in a beam of light so that the first particle shadows the second, the first particle "feels" the radiation pressure of the light, whereas the second particle does not. As a result the first particle will move toward the second. In the space between stars, where radiation comes more or less evenly from all directions, two nearby particles can lie in each other's shadow. Mock gravity is the attractive force the particles feel because of such mutual shadowing. Hogan and White estimate that in space the force can be 200 times as strong as the gravitational attraction between the particles.

If the mock-gravity theory is correct, interstellar dust particles should produce a distinct pattern of infrared radiation. The *Cosmic Background Explorer* satellite, which is scheduled for launch at the end of the decade, should detect such a pattern if it exists.



EXPLODING STAR known as a supernova was detected in galaxy M99 in the Virgo cluster by an automated sky-search system. The photograph at the left, made this past spring, shows the core of the galaxy and a foreground star from the Milky Way galaxy (*upper left*). The photograph at the right, made nine days later, shows the new supernova (which is 10 percent as bright as the entire galaxy) between the star and the galactic core.

To Catch a Rising Star

In the course of a year a large number of stars are thought to explode, becoming supernovas. In spite of the relatively high expected frequency of stellar explosions, astronomers spot only about 10 new supernovas per year—usually weeks or months after the most interesting moments of their creation. Now an automated system developed by a team of workers led by Richard Muller and Carl Pennypacker of the Lawrence Berkeley Laboratory has gone into operation. Designed to enhance both the number of supernovas detected and the timeliness of the observations, it has already discovered a supernova in galaxy M99 in the Virgo cluster, about 60 million light-years from the earth.

The automated system, which took five years to develop and test, was mounted last fall on the 30-inch telescope at the Leuschner Observatory in Lafayette, Calif. As the telescope sweeps across the sky a camera makes photographs of all accessible galaxies, approximately 400 per night. A computer then compares the photographs with a reference set of images. When the computer detects a sudden change in the brightness of the stars, it signals that a supernova may have been spotted. Investigators then compare the images to make the final determination. The procedure led to the discovery of the M99 supernova this past spring, within nine days of its creation.

When the system is fully operational, it should be able to detect as many as 100 supernovas per year, identifying them within hours after the light signaling their creation reaches the earth. Rapid detection should significantly advance understanding of how supernovas are formed, as Pennypacker notes. "Our understanding of supernovas is very sketchy," he comments. "We know little about what makes them explode or how bright they are. A lot of what's most interesting in supernovas happens just before the final collapse and in the first few seconds after it. The closer we can get to that moment, the better our understanding of what's going on. We want to look at the fire, not the smoke."

TECHNOLOGY

Tripping the Light Fantastic

It is an irony of solid-state technology that a device as complex as the logic circuitry in the central processing unit of a computer is in essence simply an array of switches. Each switch is a transistor, which is tripped by electric charge and controls the flow of an electric current. If the current flows, the switch is "on"; absence of flow corresponds to "off." A new possibility is that the switches might be photonic.

Each switch would be tripped by photons (particles of light) and would control the flow of a beam of light. Such a switch offers advantages over electronic switches: it could be arranged in solid-state devices in ways that are not possible for electronic switches. Consequently photonic switches should be able to increase greatly the capacity of computers and other devices to handle information.

David Miller and his colleagues at AT&T Bell Laboratories have devised the first working array of solid-state photonic switches. Their invention, a photonic switching chip, is a specially tailored semiconductor wafer. It incorporates a two-by-two arrangement of photonic switches called SEED's (for self electro-optic effect device). In a SEED the arrival of light produces a change in voltage, which in turn affects the material's optical behavior: specifically, the SEED switches from being strongly transmissive to being strongly absorptive.

Arranging the SEED's in a prototype chip presented difficulties. A central problem in semiconductor electronics (and now in photonics) is cascadability, which is to say that the output of one circuit element must create the input for the next. An electronic circuit element has to produce a voltage that induces or halts the flow of electric current in other transistors. The light output of an optical element has to be of a suitable intensity and wavelength to drive other optical elements.

In order to operate, a SEED requires supporting circuit elements. For example, each SEED has an internal voltage; a change in its optical transmissivity is produced by changing the voltage. The voltage that would have been applied to the SEED then has to be "absorbed" by a component such as a resistor.

Miller and his colleagues devised a strategy to incorporate the supporting circuitry in the chip while preserving the chip's minute size. The chip is created by molecular-beam epitaxy, in which a semiconductor crystal lattice is deposited on a wafer atom by atom or molecule by molecule. Each SEED consists of 2,500 layers of gallium arsenide and gallium aluminum arsenide in alternation; the supporting circuit elements are grown directly on top of the SEED's, so that the chip is "vertically integrated." A single battery supplies the voltage for the entire array. The overall thickness of the device is only six micrometers, or six thousandths of a millimeter. The device may nonetheless be the most complex structure the epitaxy technique has ever produced.

Miller predicts that the greatest advantage of the photonic technology is likely to lie in parallel processing, the handling of large quantities of information simultaneously. In electronic circuitry the capacity for parallel processing is limited because data must enter and leave a chip along a single line—the perimeter of the chip, which constitutes a one-dimensional bottleneck. The vertical integration of the photonic technology suggests a better, two-dimensional strategy: data could enter the top face of a chip and emerge at the bottom face, so that a larger quantity of information could be processed simultaneously.

Gene Genie

Biotechnology has acquired a significant new piece of automated equipment: a computerized DNA sequencer that reads off the order of nucleotides in a cloned gene. The device was developed in the laboratory of Leroy E. Hood at the California Institute of Technology and will be manufactured and marketed by Applied Biosystems, Inc.

The genetic information of DNA is encoded in the sequence of its component nucleotides, and so the rapid sequencing of DNA is a critical operation in molecular biology. Within the past decade two powerful sequencing methods have been developed, but they depend on autoradiography and require a fair amount of time and labor. The sequencing machine does the job in real time and without radioactive isotopes.

The Caltech workers exploited the sequencing technique developed by Frederick Sanger of the University of Cambridge. Each nucleotide of DNA is characterized by one of four chemical groups called bases: adenine (*A*), guanine (*G*), thymine (*T*) and cytosine (*C*). In the Sanger method the DNA strand to be sequenced is replicated as four sets of radioactively labeled fragments. Each set contains fragments ending in a particular base (*A, G, T* or *C*) and each fragment is one nucleotide longer than the next. The sets of fragments are injected at the top of four parallel gel columns, one column for each base. An electric field applied to the columns drives the fragments through the gel. Because the smaller fragments travel faster than the larger ones, the fragments are sorted into a series of groups, each of which is seen as a dark band when a photographic emulsion is exposed to the gel. By reading the order of the bands across the four columns the investigator can determine the sequence of the original DNA strand.

The Caltech workers write in *Nature*

that they automated the Sanger method by chemically manipulating the fragments so that each species is marked by one of four different fluorescent dyes: green for *A,* orange for *G,* red for *T* and yellow-green for *C.* The mixture of dye-labeled fragments is then injected into a single long tube containing a gel. An argon laser at the bottom of the tube illuminates the fragments as they migrate through the gel. A photomultiplier tube at the side of the gel column detects the light reflected by each dye molecule as it passes and sends an electrical signal to a computer. The computer translates each signal into a color—and hence a particular base—and thereby determines the DNA sequence.

Buoyant Bugs

Cyanobacteria, or blue-green algae, float near the sunlit surface of lakes and oceans. What if industrially useful bacterial cells and higher cells, which sink and clump in liquid media and are vulnerable to vigorous stirring, could be made as buoyant as the algae? They should mix more easily, grow faster and be skimmed readily off the surface of the medium, increasing yields of bioengineered products ranging from interferon to fuels.

By cloning a so-called flotation gene from cyanobacteria a team of investigators from Pennsylvania State University and the Pasteur Institute in Paris has taken a first step toward making industrially important but normally nonbuoyant cells float. The gene's product is a protein that forms minute hollow cylinders called vacuoles. The vacuoles fill with whatever gas is available in the organism or medium and float the cyanobacterium to the surface.

The team is now cloning the flotation gene in *Escherichia coli,* the favorite bacterium of biotechnologists. If vacuoles are produced in this foreign system, says Donald A. Bryant of Penn State, the gene may well be expressed in other cells too. So far *E. coli* harboring the foreign gene have been observed to float better, but it has not yet been proved that they do so because normal vacuoles are present.

Bryant thinks that rather than being expressed constantly, as it is in cyanobacteria, the flotation gene might be placed under the control of an "inducible promoter" when it is introduced into a foreign cell. (The *lac* promoter, for example, turns its genes on in the presence of the sugar lactose.) Cells containing such a gene might be made to float at will: they would produce vacuoles only after the addition of the appropriate chemical inducer.

Progress in Laser Fusion

Infrared laser beams, converted to shorter wavelengths by optical crystals, may heat heavy hydrogen to 100 million degrees Celsius. Recent work suggests the technique could lead to useful fusion power

by R. Stephen Craxton, Robert L. McCrory and John M. Soures

In the past year four of the world's largest laser systems, the NOVA laser at the Lawrence Livermore National Laboratory in Berkeley, the GEKKO-XII laser at the University of Osaka in Japan, the PHEBUS laser at Limeil, France, and the OMEGA laser in our laboratory at the University of Rochester, have come into operation at visible and ultraviolet wavelengths. At the Los Alamos National Laboratory another kind of laser, operating at a slightly shorter ultraviolet wavelength, is under development. Many large laser systems play an undisputed role in weapons research and development. Nevertheless, the lines of scientific investigation opened up by all these instruments make it possible to be guardedly optimistic about an outstanding peacetime benefit: the scientific feasibility of harnessing fusion power by laser may well be demonstrated in the next 10 years.

Major advances in two disciplines are primarily responsible for this optimism. In optics a phenomenon called harmonic-frequency conversion has been exploited to generate high-powered laser beams that have a much shorter wavelength than was previously possible. The advance is significant because laser energy can be transferred to a fusion fuel pellet much more efficiently at short wavelengths than it can at long wavelengths.

In computer modeling the simulation of fusion reactions initiated by lasers in fuel pellets has become increasingly detailed and realistic. Simulations now make it possible to specify quite accurately the characteristics of fuel pellets and incident laser beams that are needed to generate economically viable fusion power. Moreover, remarkable instrumentation can now trace the evolution of the fusion reactions to within a hundredth of a nanosecond. (A nanosecond is a billionth of a second.)

The controlled release of the energy derived from the fusion of atomic nuclei has been one of the fondest hopes of science since the potential benefits of the process were recognized and serious work on it was begun nearly 40 years ago. It is well known that a mixture of deuterium and tritium, the heavy isotopes of hydrogen, makes the most suitable fuel for fusion. The basic scientific objective is to heat ions of deuterium and tritium to sufficiently high temperatures and maintain the temperatures long enough to realize a net energy gain from the fusion of the ions. In each fusion reaction a deuterium ion and a tritium ion give rise to an alpha particle, or helium nucleus, and a neutron; the total kinetic energy released, which is shared by the two reaction products, is 17.6 million electron volts.

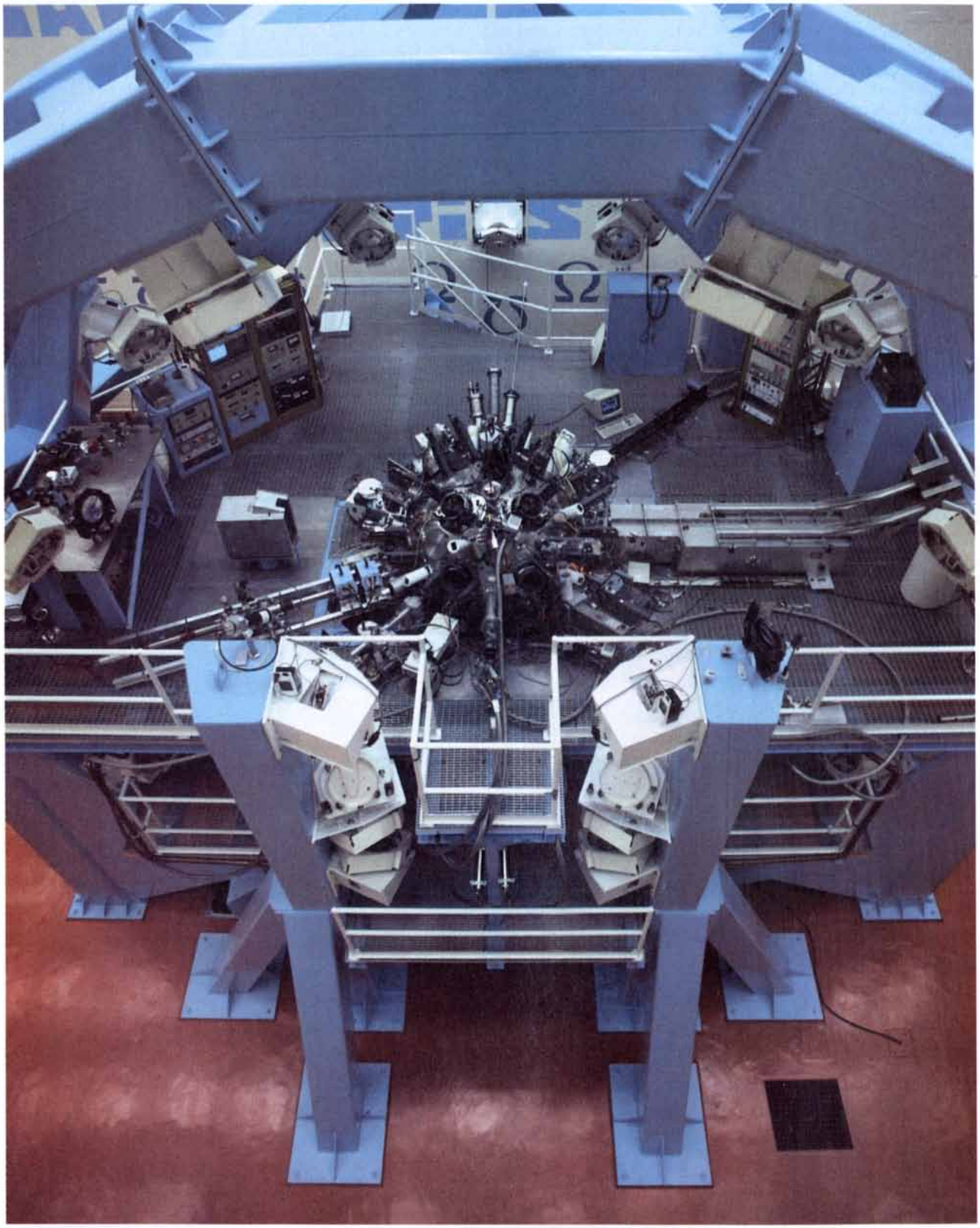
In the long term, fusion between deuterium ions alone could provide a practically limitless source of energy, since vast quantities of deuterium are available from seawater. For example, the energy from the fusion of the deuterium in a pool of water 100 feet on each side and seven feet deep could satisfy the electrical energy needs of the city of Rochester for a year.

Laser Fusion

The best-known scheme for creating the conditions needed for fusion has been to heat a plasma, a hot, electrically neutral gas of ions and free electrons, that is confined by a strong magnetic field [see "The Engineering of Magnetic Fusion Reactors," by Robert W. Conn; SCIENTIFIC AMERICAN, October, 1983]. In the early 1970's, however, workers in the U.S. began experiments on a radically different method of heating and confining a plasma, one that exploits the extraordinary heating potential of the laser [see "Fusion by Laser," by Moshe J. Lubin and Arthur P. Fraas, SCIENTIFIC AMERICAN, June, 1971, and "Fusion Power by Laser Implosion," by John L. Emmett, John Nuckolls and Lowell Wood, SCIENTIFIC AMERICAN, June, 1974]. A beam or pulse of laser light is split into several smaller beams of equal intensity. The split beams are amplified in energy and subsequently brought back together by a system of mirrors and lenses; the beams are thereby focused on a small region from different directions. A charge of deuterium and tritium fuel is encased in a spherical shell a few millimeters in diameter, made of plastic, glass or some other material, and the resulting fuel pellet is placed at the intersection of the beams; the pellet is thus uniformly illuminated.

The laser pulse almost instantly ionizes the atoms in the outermost layer of the pellet, but the material inside a certain critical radius is opaque to the laser energy. Incident energy is consequently absorbed in a dense layer of plasma that surrounds the deuterium-tritium fuel. The heated layer of plasma expands and ablates, or becomes explosively torn free, from the rest of the pellet; the velocity of the ablated plasma is typically 1,000 kilometers per second. An equal and opposite force accelerates the material inside the ablation layer inward, in accordance with Newton's third law, as if it were a rocket propelled by the plasma escaping all around it. The concentric implosive force is sufficient to accelerate the remaining shell to a velocity of several hundred kilometers per second in a billionth of a second. The radius of the fuel is compressed as much as 50 times, and the resulting high temperature and high density of the fuel cause it to fuse.

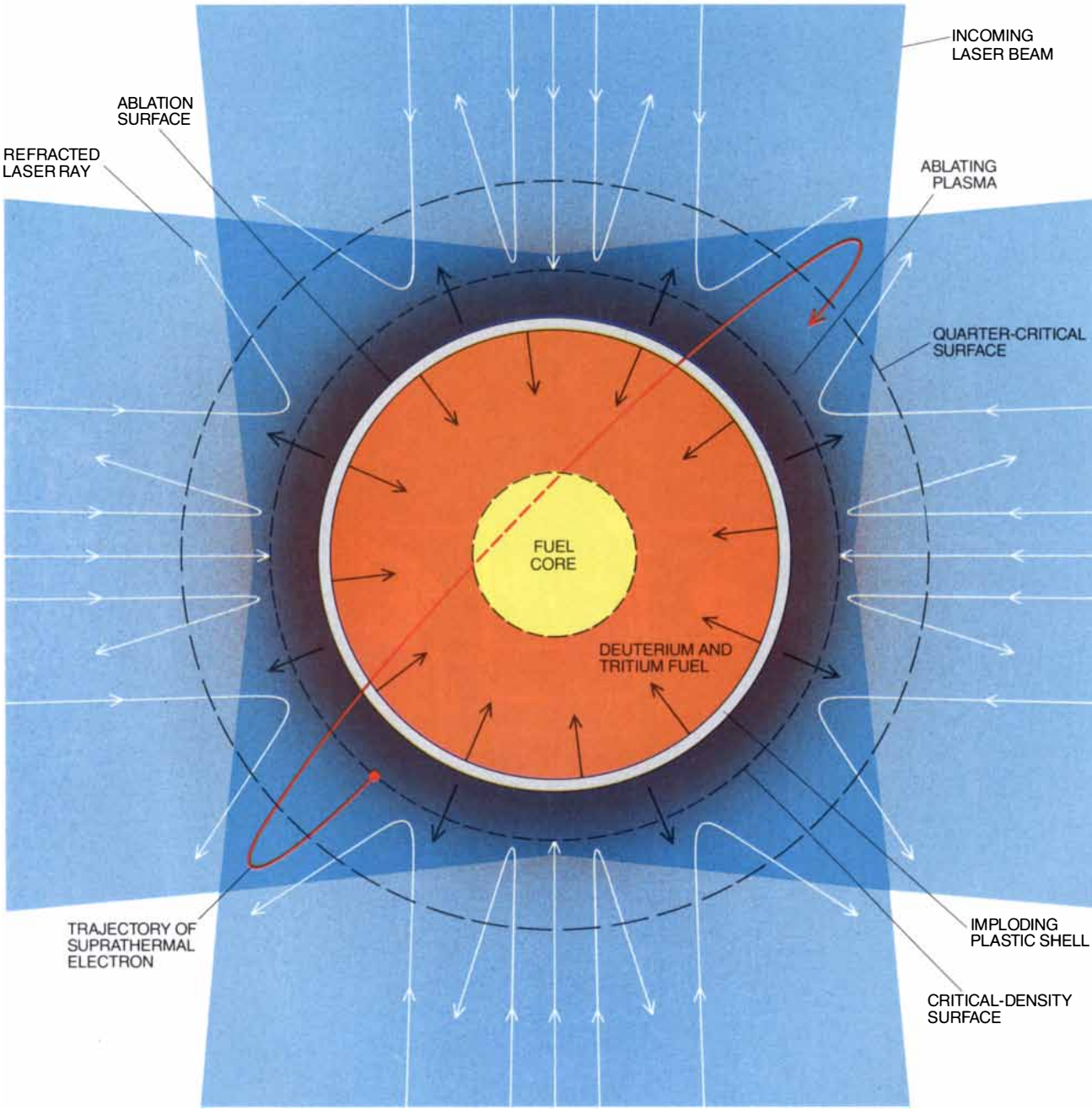
In a working fusion reactor the neutrons to which the reaction gives rise would escape from the fuel at high velocities and deposit their energy in the surrounding reactor chamber. In one design a layer within the wall of the reactor chamber would hold a fluid such



TARGET AREA of the OMEGA laser system, operated in the authors' laboratory at the University of Rochester, is shown in the photograph. The system is a neodymium-glass laser converted to a shorter wavelength in order to study the properties of fuel pellets for laser fusion. The laser first emits a single beam of infrared light at a wavelength of one micron. The beam is then amplified and split into 24 beams. After further amplification each of the 24 beams is passed through a cell holding two crystals of potassium dihydrogen phosphate (KDP), where it is converted into an ultraviolet beam at three times the input frequency. Mirrors and lenses guide and focus the beams into the target chamber, the complex spherical device at the center of the photograph. The target chamber is evacuated, and the fuel pellet is suspended at its center; the 24 focused laser beams converge on the pellet and deposit their energy there. Instruments for monitoring the effects of the laser energy on the pellet can be seen protruding from the upper hemisphere of the target chamber.

as lithium, which is known as a moderator. The neutrons would collide with the fluid and slow down, giving up their kinetic energy to the fluid in the form of heat. The fluid would circulate to a heat exchanger, which would then transfer the heat to pressurized, circulating steam, and electric power would be generated by steam-driven turbines in the usual way. The entire process is known as laser fusion.

Although we shall limit our discussion to laser fusion, we should note that particle-beam accelerators can also drive the compression of a fuel pellet. At the Sandia National Laboratories in Albuquerque construction has recently been completed on an accelerator called PBFA II (Particle Beam Fusion Accelerator), which will energize fuel pellets with beams of high-en-



LASER FUSION takes place in a fuel pellet, typically a small plastic shell about a millimeter in diameter that holds a mixture of deuterium and tritium fuel. The shell is irradiated uniformly from many directions by overlapping laser beams, and its outer portion is vaporized to form a plasma. The laser beams penetrate the plasma only up to the critical-density surface and deposit much of their energy in its vicinity. One mechanism for the energy absorption, called collisional absorption, is particularly effective at short laser wavelengths. Electrons oscillating in the electric field of the laser heat the plasma through collisions with one another and with ions in the plasma. The thermal energy of the heated plasma is then conducted inward to the ablation surface, the boundary between the hot plasma and the dense shell. The plasma outside the ablation surface explodes away from the rest of the shell like the exhaust plume of a rocket. The reaction force causes the shell to be driven inward toward the center of the fuel pellet. The implosion compresses the fuel into a dense core and heats it; thermonuclear reactions can then take place. For long-wavelength lasers a substantial fraction of the laser energy is deposited into unwanted, highly energetic suprathermal electrons. Such electrons can penetrate the core and heat the fuel in the initial stages of the implosion, thereby precluding the high compression needed in the core to achieve fusion ignition.

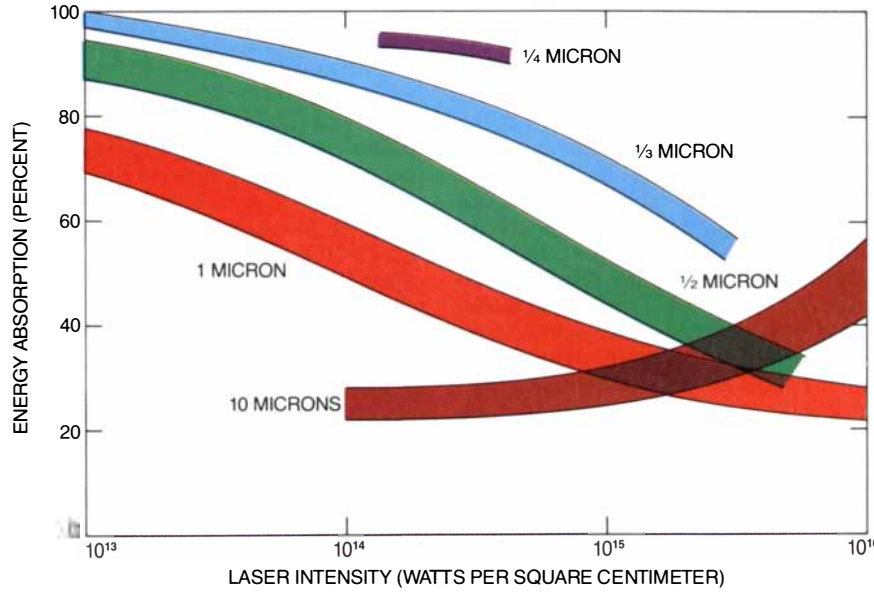
ergy lithium ions. The general term for all such schemes of generating fusion in a pellet by compressing it, whether by laser, particle beam or some other energy source, is "inertial-confinement fusion."

Frequency-converted Lasers

Until quite recently lasers powerful enough to drive fusion reactions had been built only for emission wavelengths in the infrared. Most of the experimental work to develop laser fusion has been conducted with two kinds of infrared laser: the solid-state, neodymium-doped glass laser, whose wavelength is one micron, and the carbon dioxide gas laser, whose wavelength is 10 microns. By the end of the 1970's, however, extensive experiments and elaborate computer simulations had identified a fundamental difficulty. For lasers emitting at the intensities needed for fusion and at wavelengths of one micron or longer, less than half of the laser energy is absorbed by the fuel pellet. Furthermore, of the energy that is absorbed a substantial fraction gets carried away by so-called suprathermal electrons. Such electrons pass freely through the fuel pellet and heat the fuel before it can be compressed. The premature heating precludes achieving the high densities needed for fusion.

Investigators quickly realized that the unwanted preheating of the fuel might be avoided if high-power lasers having wavelengths shorter than the infrared could be built. The apparent impasse posed by the unavailability of such lasers was sidestepped by an interesting optical trick. By carefully growing perfect crystals and then cutting each one at a precisely calculated angle with respect to the orientation of its crystal lattice, devices can be made that generate the higher harmonics of the incident beam. The frequency of the harmonics is two, three or a higher integer times the frequency of the input wave, depending on the number of crystals employed and their orientation within each device; the wavelength of the corresponding higher-harmonic beams is shortened by the same integral factor.

For example, if long-wavelength infrared laser light is passed through one crystal, a large fraction of the light energy can emerge as shorter-wavelength green light at twice the frequency of the input. If the green light and the residual infrared light from the first crystal are passed through a second crystal, they can be combined into a beam of even shorter-wavelength ultraviolet light with three times the frequency (a third of the wavelength) of the original infrared beam.



ABSORPTION EXPERIMENTS at different laser wavelengths show that the energy absorbed by a fuel pellet increases significantly as the wavelength of the laser decreases to less than one micron. The effect is particularly strong in the intensity range of 10^{14} to 10^{15} watts per square centimeter, which is the range required for generating fusion in a reactor with laser irradiation. For such wavelengths the high absorption at low intensity and the decline in absorption at higher intensities are characteristic of collisional absorption, which leads to the desired heating of the plasma surrounding the pellet. The fraction of the incident energy absorbed by unwanted suprathermal electrons increases with laser wavelength.

In one experiment at the École Polytechnique in France a one-micron beam from a neodymium-glass laser was passed through two crystals of potassium dihydrogen phosphate (KDP) in succession; the wavelength of the beam was reduced by a factor of two in each crystal, and the total reduction was therefore a factor of four. A target pellet absorbed almost all the energy of the frequency-converted beam, and suprathermal electrons were not detected. In such early experiments only a small part of the incident light was converted into light of higher frequency, but the results were still very encouraging. Investigators began directing greater attention to the physics of harmonic-frequency conversion and to methods for improving the conversion efficiency of existing neodymium-glass lasers.

Temperature and Density

To appreciate the need for the extraordinary scientific efforts being devoted to frequency conversion and other aspects of laser-fusion technology one must understand the physical environment required if laser fusion is to generate economically useful energy. The temperature of the fuel in each pellet must reach almost 100 million degrees Celsius, several times the temperature at the center of the sun. Simultaneously the fuel must be compressed to a density typical of what is found in the center of the sun: more than 1,000 times its ordinary solid or liquid density of .2 gram per cubic centimeter. The pressures corresponding to such a temperature and density are immense; indeed, pressures on the order of several billion atmospheres have already been generated in laser-driven implosions. Furthermore, macroscopic quantities of fuel must be brought to such conditions. The creation of such an environment not only would have immediate application to fusion power but also would open the possibility of doing experiments of direct relevance to astrophysics in the terrestrial laboratory.

The need to heat fusion fuels derives from the basic energetics of nuclear interactions. At a very short range the protons and neutrons that make up atomic nuclei are attracted to one another by the strong nuclear force. Many systems of nuclei can therefore increase their binding energy if the nuclei fuse. The total mass of the reaction products is slightly less than the sum of the masses of the fusing nuclei. Thus, in accord with Einstein's equivalence between energy and mass, the mass lost in the reaction is converted into energy, which shows up as the kinetic energy of the reaction products.

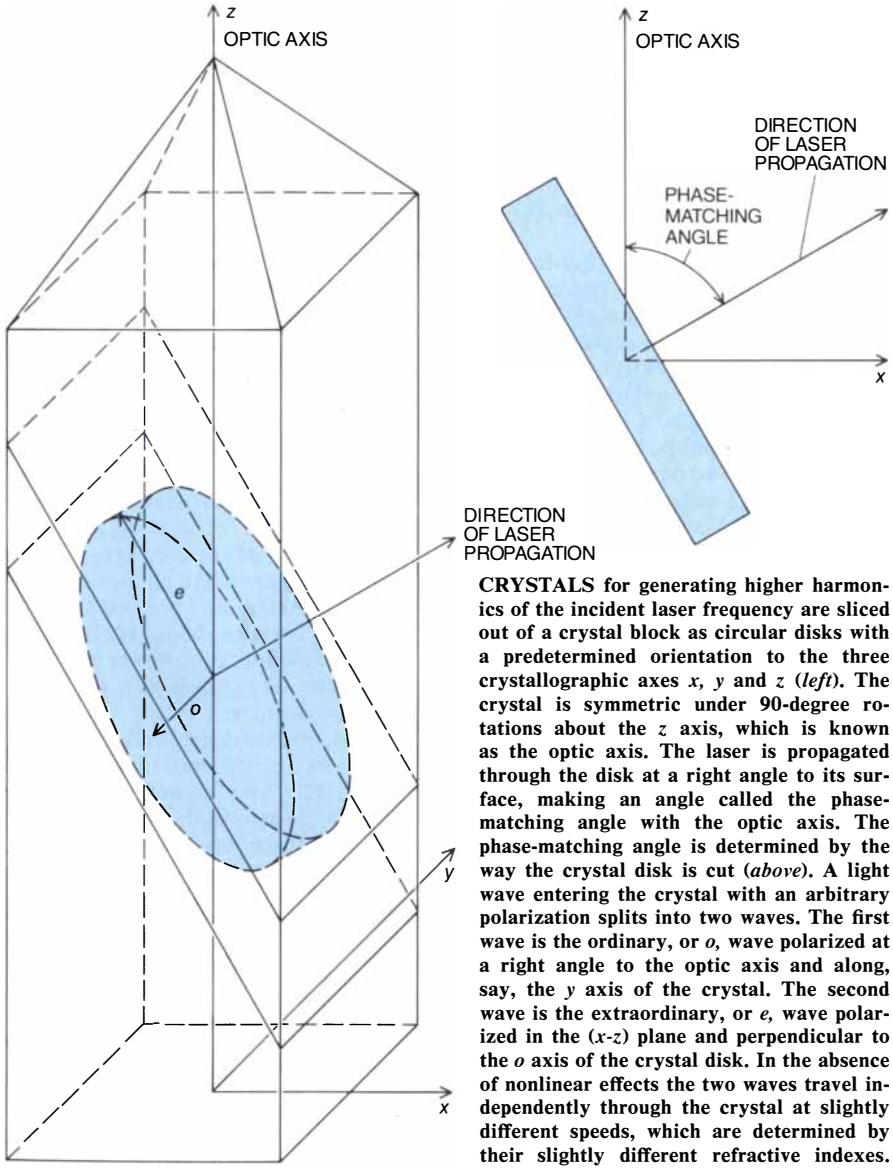
There is an important energy barrier

to this path for reducing potential energy. At temperatures less than about 10 million degrees C. the mutual electrostatic repulsion of the positively charged protons in the nuclei keeps the nuclei from coming close enough to "feel" the attraction of the strong force. Only within a distance of about 2×10^{-13} centimeter—comparable to the radius of the nucleus itself—does the strong force become dominant. To approach within this range two nuclei must be brought together with considerable kinetic energy; in other words, their temperature must be raised substantially. An important reason for the selection of deuterium and tritium ions as fusion fuels is that each ion carries only one positive charge, and that small charge minimizes the mutual repulsion of the fuels.

High compression is not a theoretical prerequisite for fusion, but it enhances the efficiency of the fusion reactions, and it is essential in a practical reactor design. The laser input energy needed to fuse deuterium and tritium efficiently at ordinary densities is well beyond the capacity of present laser technology. When the fuel in the pellet reaches its peak compression, the kinetic energy of the imploding material is converted into heat, and the confinement of that heat to the small compressed region raises the temperature of the fuel. Thermonuclear burning begins in a small "spark plug" region at the center of the fuel. Four-fifths of the energy released by the burning emerges as the kinetic energy of neutrons, which pass through the surrounding plasma and deposit their energy into a fluid circulating in the reactor chamber or into a moderator in the reactor wall.

The other 20 percent of the energy released by the fusion reactions is carried away by alpha particles. Because they are charged particles, they are slowed by the fuel much more effectively than the neutrons are. They collide with the surrounding fuel and give up their kinetic energy over a distance inversely proportional to the fuel density. A sizable fraction of the fusion energy can thus be deposited in the cooler layers of the compressed fuel if the fuel is dense enough and its radius is large enough to stop the alpha particles. Such partial absorption of the fusion energy in the fuel is critical to the efficient propagation of thermonuclear burning; it is called bootstrap heating. When the energy deposited in the fuel by the alpha particles exceeds the energy needed to compress the fuel, the pellet is said to have ignited. If the attainable value of the compressed fuel density is known, the core radius, the fuel mass and the laser energy needed for ignition can all be determined. Achieving ignition would be a milestone in demonstrating the scientific feasibility of laser fusion.

There is a small price to be paid for high compression: soon after fusion begins it must cease because the high internal pressure of the compressed fuel causes it to fly apart. The higher the compression is, the faster the fuel disassembles. Nevertheless, the high reaction rate at high compression more than compensates for the limited burning time. For example, other things being held constant, a tenfold compression of the radius of the fuel increases the reaction rate by a factor of 1,000, whereas the duration of the burning is reduced by a factor of only 10. It should be noted that in magnetic fusion devices the density of the fuel is 10 billion times lower than it is in the burning core of the pellet. The design must compensate for the low density by confining the fuel for 10 billion times as long.



CRYSTALS for generating higher harmonics of the incident laser frequency are sliced out of a crystal block as circular disks with a predetermined orientation to the three crystallographic axes *x*, *y* and *z* (*left*). The crystal is symmetric under 90-degree rotations about the *z* axis, which is known as the optic axis. The laser is propagated through the disk at a right angle to its surface, making an angle called the phase-matching angle with the optic axis. The phase-matching angle is determined by the way the crystal disk is cut (*above*). A light wave entering the crystal with an arbitrary polarization splits into two waves. The first wave is the ordinary, or *o*, wave polarized at a right angle to the optic axis and along, say, the *y* axis of the crystal. The second wave is the extraordinary, or *e*, wave polarized in the (*x*-*z*) plane and perpendicular to the *o* axis of the crystal disk. In the absence of nonlinear effects the two waves travel independently through the crystal at slightly different speeds, which are determined by their slightly different refractive indexes.

Collisional Energy Absorption

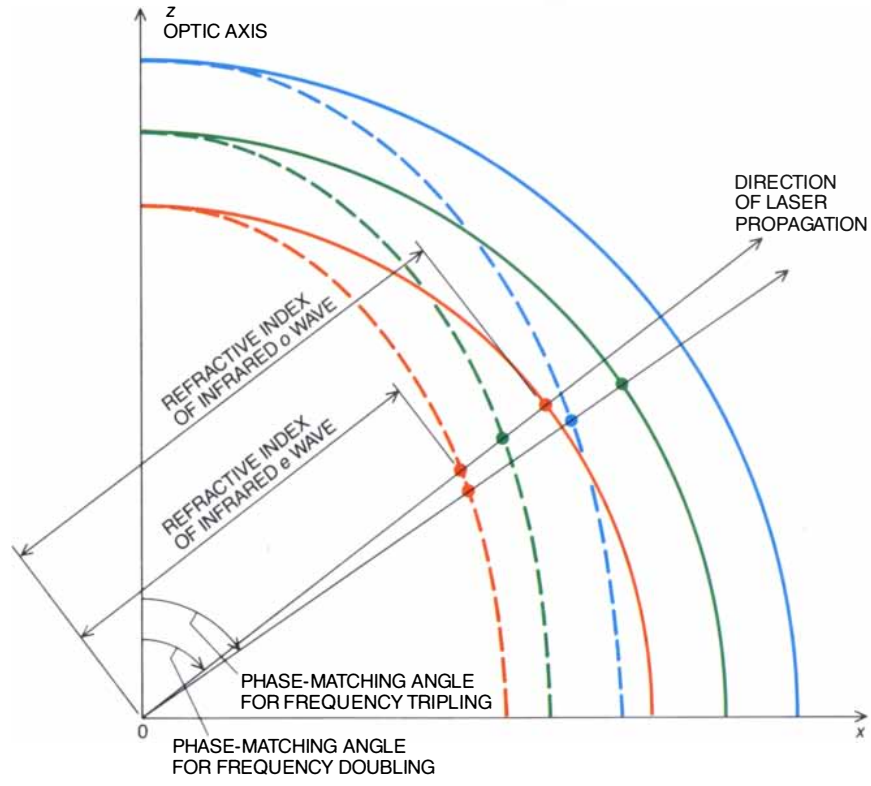
Given the need for high temperature and high density of the fuel, which of the many physical processes relevant to laser fusion are the most important to understand and control? It turns out that the most critical processes depend on details of the interaction of the laser light with the plasma corona, or atmosphere, surrounding the fuel pellet. For example, the transfer of energy from the laser to the ablation layer begins in the corona. Furthermore, the degree to which the fuel is preheated by suprathermal electrons, prior to the implosion, is intimately related to the plasma physics of the corona.

After the first incident laser irradiation ionizes the outer surface of the

pellet, the resulting plasma corona begins to expand outward. The corona is dense near the surface of the pellet and becomes rarefied at larger radii. The rapidly oscillating electric field of the subsequent incoming laser light causes rapid oscillations of the lightest charged particles in the plasma, namely the electrons. A current is therefore set up in the plasma, oscillating at the frequency of the driving electric field and proportional to the density of the electrons in the plasma.

In the rarefied outer regions of the corona the current is small, but in the inner regions, where the electron density is greater, the current becomes larger. At what is called the critical radius the current becomes large enough to shield the plasma at smaller radii from further penetration of the incoming electric field. The oscillating electric current acts as an antenna that broadcasts a second, outgoing electromagnetic wave, equal in frequency to the incoming wave but carrying energy away from the target. The electron density at the depth of maximum penetration of the incoming wave is called the critical density, and it is inversely proportional to the square of the laser wavelength. For example, a threefold reduction of the laser wavelength would give rise to a ninefold increase in the critical density.

Much of the absorption of the laser energy takes place in the vicinity of the critical-density surface. The oscillating electrons collide with ions in the plasma as well as with one another, thereby transferring part of their energy to the plasma as the energy of random motion, or heat. This mechanism for energy transfer is more efficient for a short-wavelength laser than it is for a long-wavelength laser, since the short-wavelength beam can penetrate to higher electron densities where collisions are more frequent. The heat energy is conducted inward to the ablation surface, the boundary between the exploding and imploding regions of the pellet. Collisional absorption accounts for virtually all the absorbed energy that contributes to the implosive compression of the pellet.



ORIENTATION OF DISK SLICE with respect to the crystallographic axes is determined by a simple geometric construction. The red arcs represent the input frequency of the infrared laser, and the green and blue arcs respectively represent the second and third harmonic frequencies. The refractive indexes of the *o* waves vary with wavelength but not with a change in the orientation of the slice with respect to the *z* axis. Hence the refractive index of the infrared *o* wave is shown as a red circular arc, and the indexes of the green and ultraviolet *o* waves are shown as green and blue circular arcs. The refractive indexes of the infrared, green and ultraviolet *e* waves do vary with the angle the slice makes with the *z* axis; the variations are plotted as red, green and blue broken elliptical arcs. The propagation direction for frequency doubling must be chosen in such a way that the refractive index of the green *e* wave is equal to the average of the refractive indexes of the infrared *o* and *e* waves. For frequency tripling the refractive index of the ultraviolet *e* wave must equal one-third of the index of the infrared *e* wave plus two-thirds of the index of the green *o* wave.

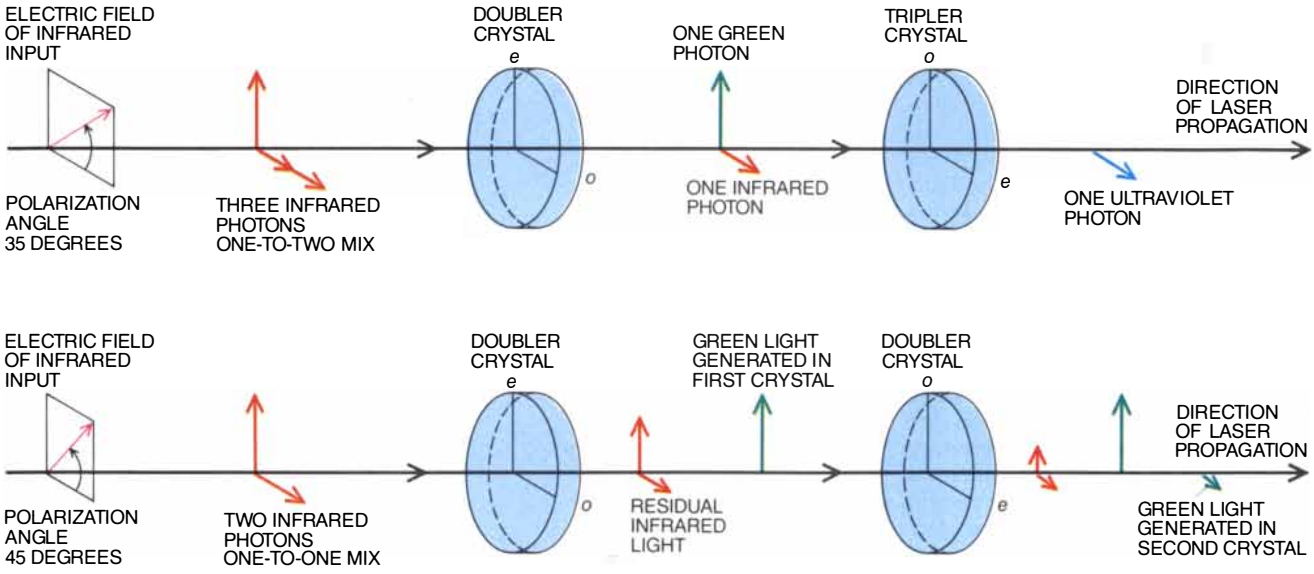
Suprathermal Electrons

There are several other mechanisms for laser-energy absorption that do not contribute to the heating or compression of the fuel but must be understood because they give rise to unwanted suprathermal electrons. As the name implies, the energy of a suprathermal electron is significantly higher than the average energy of the thermal electrons in the plasma. Such excess energy can be acquired only if the electron is accelerated by a large electric field for a sufficient time. The electric field of the laser, although intense, changes direction so often that the velocity of the electrons oscillating in the field is relatively small. Hence in most experimental conditions the electric field of the laser does not lead directly to suprathermal electrons.

Large electric fields can nonetheless arise in waves known as plasmons, which travel through the plasma. Plasmons are somewhat like sound waves in that they propagate as compressions and rarefactions of particles in the direction of the wave motion. Unlike sound waves, however, plasmons do not affect all the particles in the medium: the ions in the plasma are stationary and only the electrons move. Regions of net positive charge are thereby created in the rarefied electronic regions, and regions of net negative charge are created in the compressed regions. A large electric field is formed between a rarefied region and a compressed region, and there is a strong attractive force between the two.

Electrons that happen to be moving in the same direction as the plasmon and at roughly the same speed become trapped in the plasmon somewhat like a surfer riding a wave. Since they move with the same speed as the wave, they experience its electric field as a nonoscillating field. The field accelerates them to suprathermal velocities, which enable them to escape the plasma layer before collisions can slow them down.

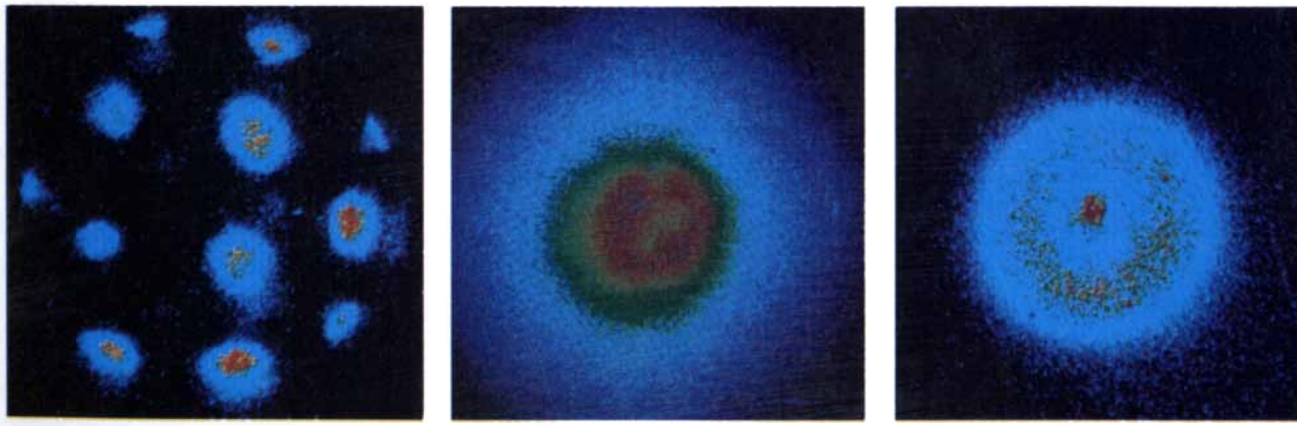
Plasmons can be generated in many ways. One important mechanism is called resonance absorption. As a laser ray penetrates the plasma, much of the energy of its oscillating electric field is given up to surrounding electrons. Near the critical-density surface, however, the natural oscillating frequency of plasmons equals the laser frequen-



ROCHESTER FREQUENCY-TRIPLING SCHEME (*top*) **employs two crystals whose *o* and *e* axes are mutually perpendicular. The incident infrared laser beam is polarized at an angle of 35 degrees to the *o* axis of the first crystal. This ensures that two-thirds of the photons incident on the first crystal are aligned with its *o* axis and one-third are aligned with its *e* axis. One *o* photon combines with an *e* photon to give an *e* photon at the second harmonic (*green*), or at twice the frequency of the incident light. The green *e* photon combines in turn in the tripler crystal with the remaining infrared photon to give one ultraviolet photon at the third harmonic (*blue*). In the authors' laboratory the technique has been shown to be 80 percent efficient, and it is now employed in the NOVA laser at the Lawrence Livermore National Laboratory. A simple reorientation of the two crystals (*bottom*) also makes it possible to generate the second harmonic in the NOVA laser with improved efficiency. The input beam is polarized in such a way that the numbers of infrared *o* and *e* photons are equal. The doubling process is less than 100 percent efficient, and some residual infrared photons remain unconverted after passing through the first crystal. When the second crystal is tilted through a small angle from its position for frequency tripling, it becomes a doubling crystal, and residual infrared photons have a second opportunity to be converted to green.**

cy. The energy that reaches this depth in the plasma can drive plasmons resonantly to large amplitudes, much as a child on a swing can go progressively higher by pumping in synchrony with the swing's natural motion. The energy that is pumped into resonantly driven plasmons is eventually released as the kinetic energy of suprathermal electrons.

There are two further important mechanisms whereby plasmons can form and generate suprathermal electrons. Although both mechanisms appear to be less significant causes of preheating than resonance absorption, it is worthwhile to explain them briefly. Both of them arise out of a phenomenon called three-wave mixing, which also plays a major role in harmonic-frequency conversion. In three-wave mixing two waves can interact through the matter in a plasma or a crystalline solid to produce a third wave. In general the mixing is strongest when the amplitudes of the interacting waves are large. The frequency of the third wave equals the sum of the frequencies of the two input waves. The direction of the process is also re-



X-RAY IMAGES can determine the uniformity of laser irradiation on a fuel pellet. The color-enhanced image at the left shows the X-ray emission when all 24 beams of the ultraviolet OMEGA laser are focused onto small areas of the pellet. Such illumination does not lead to symmetric compression, but it is useful for testing the pointing and focusing of the laser beams. The other images are from imploding pellets irradiated with overlapping beams. In the middle is a large pellet with a thin glass shell and at the right is a smaller pellet with a thick glass shell; both pellets are filled with deuterium and tritium gas. The red parts of each image are the regions of highest X-ray emission from the compressed part of the shell. The thick shell maintains the fuel at a lower temperature than the thin shell does, and it also generates a higher compression of the fuel. The X-ray emission extends just beyond the initial pellet surface.

versible: from a "product" wave two new "factor" waves can emerge.

The first mechanism for the formation of plasmons is known as the two-plasmon instability. Here the laser beam functions as the product wave, and it splits into two plasmons, which function as the two factor waves. The frequency of each plasmon is half the frequency of the incoming beam. Since the critical density of the plasma electrons varies inversely with the square of the input wavelength, the two plasmons are driven resonantly when the electron density in the plasma is a quarter of the critical density for the input laser. Hence the two-plasmon instability arises near the so-called quarter-critical surface [*see illustration on page 70*].

The second mechanism for plasmon formation is called Raman scattering; again the laser beam is the product wave, and its factor waves are a plasmon and a reflected light wave. Raman scattering can take place at the quarter-critical surface and in more rarefied regions of the plasma. The plasmons created by the two-plasmon instability and by Raman scattering generate suprathermal electrons in much the same way as they do in resonance absorption.

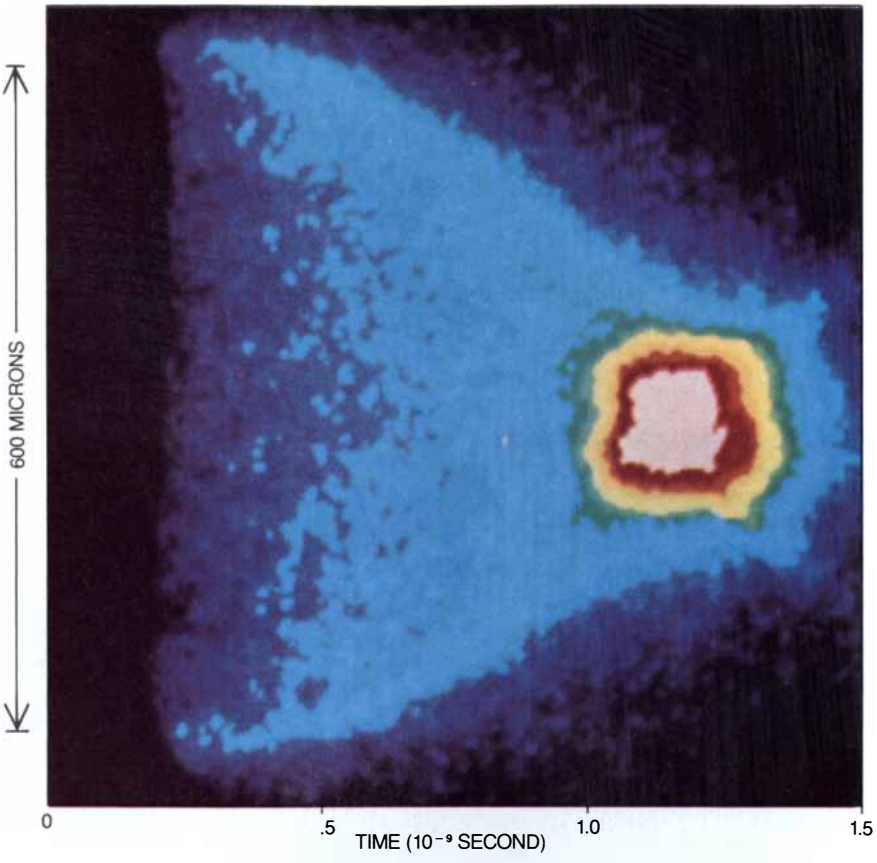
When the first suprathermal electrons escape from the corona, they leave a net positive charge on the pellet. The charge attracts later, outward-moving suprathermal electrons back toward the pellet, whereupon they overshoot their original positions and pass on into the pellet core. Some of the most energetic electrons may oscillate several times through the pellet before collisions with the fuel finally slow them; the transfer of collisional energy to the fuel preheats the fuel. The attractive electrostatic force between the suprathermal electrons and the ions in the corona can also lead to the outward acceleration of highly energetic ions. The loss of energetic ions from the ablation layer also drains away laser energy that might otherwise drive the implosion.

Experiments show that for the range of beam intensities relevant to laser fusion—generally from 10^{14} to 10^{15} watts per square centimeter—collisions are weak at infrared wavelengths and resonance absorption is dominant. For wavelengths shorter than .5 micron, however, a variety of experimental results suggest that collisional absorption is the most important absorption mechanism. Moreover, at such wavelengths neither the two-plasmon instability nor Raman scattering have been found to preheat the fuel significantly, although further experimental tests of these effects are needed for the larger pellets required in a commercial laser-fusion reactor. Thus empirical study to date has confirmed the significant advantages of short-wavelength lasers for fusion.

Frequency-Conversion Crystals

The desire to investigate how laser energy at submicron wavelengths can be exploited to compress fuel pellets has led to the recent interest in frequency-converted lasers. Since frequency conversion can readily shorten the wavelength of a laser to a half or a third of its ordinary emission value, the wavelength range of interest is experimentally accessible with existing neodymium-glass lasers.

Optical frequency conversion may become less critical as new kinds of lasers appear. For example, the krypton-fluoride laser, a gas laser under development at the Los Alamos National Laboratory and elsewhere, can also



IMPLOSION HISTORY of a spherical, glass fuel pellet uniformly illuminated by the OMEGA laser is captured in this color-enhanced photograph. An X-ray pinhole camera forms an image of the pellet similar to the ones in the bottom illustration on the opposite page, and a mask with a slit is placed over the X-ray image (*left*). A streak camera then displaces the part of the image seen through the slit a distance to the right that is proportional to the time elapsed during the implosion. The resultant image (*upper image*) thus portrays the implosion in both space and time. Soon after the laser begins illuminating the pellet, X rays are emitted from the plasma near its initial surface. As the pellet is compressed the region emitting X rays moves closer to its center, until a bright flash of X rays is visible at peak compression, about one nanosecond after the implosion begins. The speed of the imploding shell is about 200 miles a second.

deliver the high energies needed for fusion implosion; its emission wavelength is only .25 micron. Yet although the krypton-fluoride laser is widely considered to be a promising candidate for a reactor laser, questions related to the efficiency with which it can deliver the short pulses of energy needed in a reactor are still under study. The neodymium-glass laser remains the major research tool for the study of short-wavelength, laser-fusion implosions.

Many kinds of crystal have been incorporated into small laser systems for converting laser light to its higher harmonic frequencies. The crystal KDP, however, is the only one that has so far been grown large enough for current work. KDP crystals are grown from solution at a rate of only a few centimeters per month. The largest available crystals of acceptable quality are more than 30 centimeters in diameter and take as long as a year to grow. Circular slices about one centimeter thick are cut from the crystal, polished and mounted in the path of a laser beam.

One crystal can generate the second harmonic, a wave at twice the frequency of the fundamental beam. A second crystal mounted in the path of the first output beam can then mix the generated second harmonic with the residual fundamental to produce the third harmonic [*see top illustration on page 74*], or the crystal can generate the fourth harmonic by a second doubling process. In principle there is no upper limit to the frequency that could be generated by successive stages of doubling, but unfortunately suitable crystals transparent to wavelengths shorter than .2 micron are not available.

Harmonic Generation

The capacity of some crystals to generate higher harmonic frequencies of a laser beam was discovered soon after the invention of the laser. Investigators at the University of Michigan found that a laser beam focused into a crystal of quartz emerged from the crystal with some light at the second-harmonic frequency. As we mentioned above, the phenomenon is an example of three-wave mixing; in this case two components of the input laser beam make up the two factor waves, and the higher-harmonic output is the product wave. The two components of the input beam arise as the beam enters the crystal because of asymmetries in the alignment of the atoms that make up the crystal lattice.

When the oscillating electric field of the laser enters the crystal, it displaces electrons from their equilibrium locations, just as it does in the corona of the fuel pellet. In a KDP crystal the electrons are harder to displace along a preferred axis called the optic axis, or z axis, than they are along any direction lying in the x-y plane perpendicular to the z axis. The crystal is "stiffer" to the force of the electric field along the optic axis, and the electromagnetic wave whose electric field is aligned with the optic axis propagates the fastest through the crystal.

Imagine that the crystal is cut so that its face makes an oblique angle to the z axis, and suppose a light ray strikes the crystal at a right angle to its face [*see illustration on page 72*]. The electric field of the ray, which then oscillates parallel to the face of the crystal, has two mutually perpendicular components. The first component, which lies

BEAM SPLITTING, amplification and frequency tripling of the OMEGA laser are carried out in the large room shown. The bright areas on the far wall (160 feet from the camera) are caused by flash lamps that energize the laser amplifiers. The laser beams are infrared, and so they are invisible. They are converted into invisible ultraviolet light by crystals in the six boxlike modules. The visible light emerging from some of the modules comes from heat lamps, switched on intermittently to keep the crystals at a constant temperature. Residual green light from the conversion causes the green glow. The beams emerge from the modules in groups of four, and their energy is measured. Subsequently the beams are reflected by the mirrors visible in the foreground into the adjacent target area.

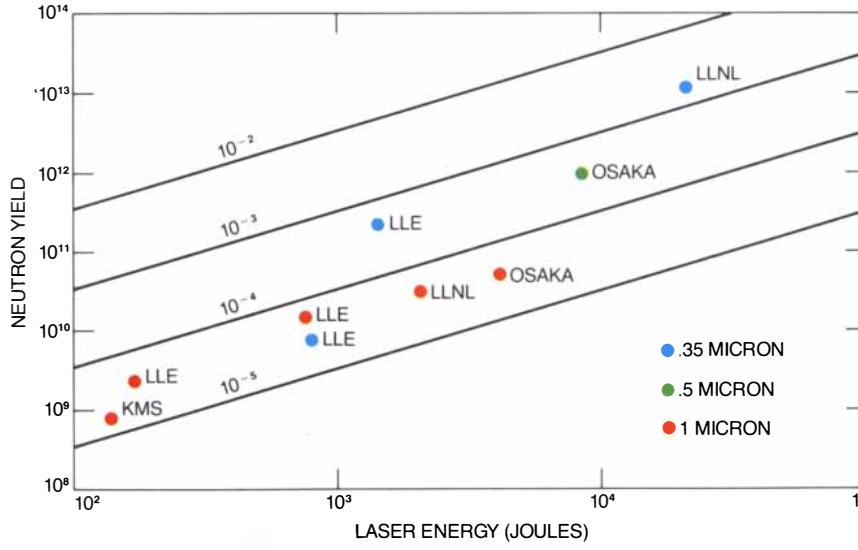
in the *x*-*y* plane and is parallel to one of the crystal axes (by convention the *y* axis), is called the ordinary, or *o*, ray. Because it has no component in the *z* direction, it propagates relatively slowly through the crystal. The second component of the incident ray is perpendicular to the first one in the plane of the crystal face. This second component is called the extraordinary, or *e*, ray, and it does have a component in the *z* direction. Because of its *z* component, which depends on the orientation of the crystal face with respect to the *z* axis, the *e* component of the incident beam moves through the crystal faster than the *o* component.

When high-intensity laser light travels through a crystal, the electrons displaced by the electric wave tend to feel a restoring force that is neither in the same direction as their displacement nor proportional to it. It is the nonlinear, or nonproportional, response of the electrons to the *e* and *o* components of the electric field that gives rise to a so-called nonlinear current wave. The current wave is made up of electrons oscillating in the *x*-*z* plane at twice the frequency of the input laser beam, and it moves through the crystal at a velocity equal to the average velocity of the *e* and *o* components of the input laser. More precisely, the refractive index of the current wave is the average of the refractive indexes of the *e* and *o* components.

The current wave of oscillating electrons generates a wave of laser light at the second harmonic of the laser beam, just as oscillating electrons in an antenna generate radio waves. The second-harmonic light wave can be generated efficiently, however, only if it propagates at the same velocity as the current wave: it must "surf," or be in phase with, the current wave. Because the second-harmonic wave is an *e* wave, its velocity and therefore its refractive index vary with the angle between its direction of propagation and the optic axis. Hence the velocity of the second harmonic can be matched with the velocity of the current wave by cutting the crystal in such a way that the direction of propagation of the laser through the crystal makes a predetermined angle with the optic axis [*see illustration on page 73*].

Attaining High Compression

Two more features of laser-fusion design, complementary to the requirement for a short-wavelength input beam, are crucial for reaching high fuel compressions in the pellet. They are the uniformity of the illumination and absorption of the incident laser beams and the hydrodynamic stability of the imploding pellet.



NEUTRON YIELD obtained with several laser systems depends primarily on laser wavelength, energy and the symmetry of the incident beams. The sloping lines on the graph indicate the energy gain of the pellet: the fusion energy generated divided by the input laser energy. Results are from KMS Fusion, Inc., of Ann Arbor, Mich. (KMS); Lawrence Livermore National Laboratory (LLNL); the Laboratory for Laser Energetics at the University of Rochester (LLE), and the Institute for Laser Engineering in Osaka, Japan (OSAKA).

The uniformity of the illumination and absorption is important for purely geometric reasons. If the radius of the compressed shell is to be decreased by a factor of 30, the implosion velocity must be uniform over the surface of the shell to within about one part in 60. That requirement limits the allowable fluctuations of the laser irradiation intensity to one or two parts in 100 over the surface of the pellet.

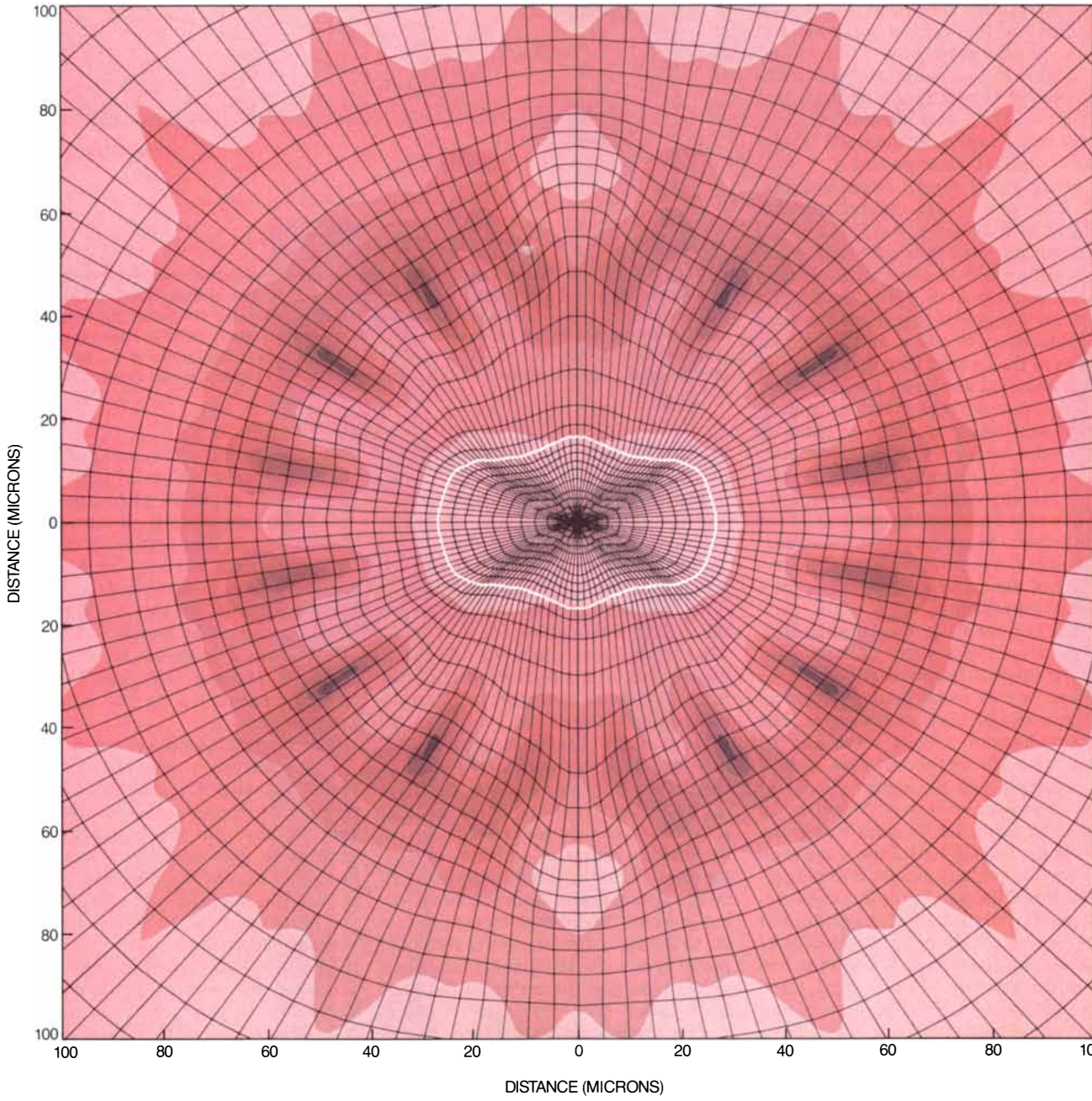
How many beams are needed, and how should they be arranged around the pellet? For laser systems employing four, six, eight, 12 or 20 beams, each beam can be positioned as if it were at the center of a face of one of the five Platonic solids (tetrahedron, cube, octahedron, dodecahedron or icosahedron) and directed perpendicular to the face. The beams are then disposed symmetrically around the pellet if it is placed at the center of the Platonic solid. The frequency-doubled VULCAN laser at the Rutherford Appleton Laboratory in England, and the GEKKO-XII laser in Osaka both employ a dodecahedral geometry. When the number of beams is greater than 20, or in other words greater than the number of faces in the regular icosahedron, suitable geometric configurations that maximize the uniformity of the illumination can be calculated. For example, a 32-beam system can be based on the configuration of the 32 pentagonal and hexagonal faces of a soccer ball.

Symmetric placement of the surrounding beams does not guarantee uniform irradiation of the pellet, however: not all points on a hemisphere of the pellet are equally irradiated by a given beam. The attainable irradiation uniformity increases as the number of beams is increased, although the complexity of the laser system also increases. The number of beams is therefore a trade-off between uniformity and complexity. The 32-beam system appears to be an attractive compromise.

The optical quality of each beam and the balance of energy among the beams are also important to the uniformity of irradiation. In our laboratory we have put much emphasis on generating laser beams with a uniform cross section, and we have been able to reduce the intensity variations across the profile of a beam to less than 10 percent. Other methods of enhancing the illumination uniformity have been proposed by workers in Osaka and at the Naval Research Laboratory in Washington. In both methods optical systems near the focusing lens fragment each beam that is split from the master beam into many—say 400—smaller beamlets. Each beamlet is given a random phase, and each spreads its energy over a full hemisphere of the pellet. Any spatial nonuniformities in any of the beams at the focusing lens are then averaged out over the pellet surface. Initial results from both laboratories show that the methods improve the uniformity of irradiation.

Another major goal in the study of

laser fusion is to understand and control hydrodynamic instabilities of the fuel pellet. When the imploding shell is decelerating just before the peak compression of the fuel, the dense fluid of the plasma shell can fall in spikes and mix with the deuterium and tritium fuel. At the same time the fuel can rise through the dense plasma in bubbles. The instability is similar to a phenomenon of incompressible fluids called the Rayleigh-Taylor instability, in which a dense fluid falls through a light one. It can be triggered by nonuniformities of irradiation or by imperfections in the shape or composition of the pellet, particularly when such imperfections are small. The ablation surface outside the fuel layer can also become hydrodynamically unstable as the implosion accelerates.



COMPUTER SIMULATION of the implosion of a fuel pellet designed for a reactor illustrates the effect of irradiation nonuniformities on pellet performance. The nonuniformities are assumed to be symmetric for rotations about the north-south axis of the pellet. They are characterized by the number of maximums of irradiation intensity encountered around the great circle represented by a line of longitude on the pellet. The pellet is a plastic shell with a layer of cold, liquid deuterium and tritium fuel on its inner surface and a small amount of deuterium and tritium gas inside. When the gas is compressed, it forms a so-called spark-plug region in which thermonuclear reactions are ignited. An irradiation nonuniformity of 1 percent is imposed in the simulation. The illustration shows the compressed core just before ignition; darker regions correspond to higher densities. The grid lines indicate the resolution with which the evolution of the pellet implosion was calculated. The boundary between the main fuel and the gas, which was the inner layer of the shell before the compression, is shown as a white contour line. Most of the material shown is compressed to a density of between 100 and 500 grams per cubic centimeter, or from 500 to 2,500 times its liquid density. The calculation was done with the program ORCHID, written by Charles P. Verdon of the University of Rochester; it required tens of hours on a CRAY X-MP, one of the world's fastest supercomputers. In spite of the nonuniform compression the fusion yield in the simulation was nearly 100 times the laser-energy input.

That instability can lead to a breakup of the shell, a mixing of the fuel and the shell and a smaller peak compression. Ultimately such effects eliminate or reduce the thermonuclear yield.

One-dimensional spherically symmetric simulations show that the highest compressions can be obtained for pellets whose shells are thin compared with their radii. Hydrodynamic instability, however, is potentially most dangerous for pellets with thin shells. A realistic pellet design must therefore represent a trade-off between the desirability of relatively thin shells and the constraints imposed by uniformity and stability considerations. It seems desirable to ensure that the shell never becomes thinner than a few percent of the radius of the pellet throughout the implosion.

Two designs currently assume major importance. The shell in our simulation is made out of plastic, and it contains an inner, cryogenic layer of liquefied deuterium and tritium. The shell interior is filled with low-density deuterium and tritium gas at the vapor pressure of the cryogenic liquid. The use of plastics made out of elements of low atomic number reduces the emission of radiation from the plasma corona, which could preheat the fuel. Since the cryogenic pellet is essentially empty, the shell is free to accelerate inward until the final moments when its kinetic energy is converted into heat.

An important alternate pellet design, which we can mention only briefly, exploits the response of materials of high atomic number to laser irradiation. The fuel pellet is placed inside a so-called radiation case made out of a material such as gold. When the interior of the gold case is irradiated by the laser, a large fraction of the incident energy is transformed into X rays. The X rays then irradiate the pellet and cause it to implode in a highly uniform way. Since the laser beams do not irradiate the pellet directly as they do in the first design, this alternate scheme is called indirect drive.

Computer Simulations

We shall conclude by describing computer simulations of a fuel pellet that might be considered for a commercial reactor, irradiated by a krypton-fluoride laser whose energy is 1.6 million joules and whose wavelength is a quarter of a micron. (A joule is the energy needed to lift one kilogram approximately 10 centimeters.) In our simulations the laser energy is deposited with an average power of 3×10^{14} watts in a pulse whose intensity increases steadily for five to 10 nanoseconds. The intensity of the pulse is kept low enough to avoid generating significant numbers of suprathermal electrons in the plasma corona.

One important measure of the response of the pellet to the laser is called the spark-plug convergence ratio, the ratio of the initial radius of the spark-plug material to its final radius. To maintain the stability of the pellet the ratio should not be too high; we aim for a ratio of about 50. In our models we then impose nonuniformities from point to point in the laser illumination of the pellet. As the interface between the shell and the fuel decelerates near the time of peak compression, pressure variations and the growth of hydrodynamic instabilities cause the density contours to follow a complex pattern around the core [*see illustration on opposite page*].

The alpha particles emitted by thermonuclear burning in the spark-plug region begin to deposit their energy in the main fuel layer. When the product of the density and the radius of the spark-plug region rises to values large enough to stop most of the alpha particles, thermonuclear ignition is achieved. The temperature in the spark-plug region reaches approximately 100 million degrees C.; the temperature of the main fuel layer is about 30 million degrees. The thermonuclear burn propagates radially outward into the main fuel layer, the temperature profile becomes smoother and the burn becomes nearly spherical as the main fuel is consumed. The pellet then disassembles and the thermonuclear fire goes out.

According to our simulations, which have considered a variety of irradiation nonuniformities, the thermonuclear energy output is typically 100 times the energy input of the laser. These results inspire renewed confidence in the feasibility of laser fusion. A power plant capable of generating almost a billion watts might be feasible if 10 such fuel pellets were ignited every second and if the overall efficiency of the laser were approximately 15 percent. On the basis of these calculations we expect that a laser emitting more than 1.6 million joules of energy but less than 10 million joules would be appropriate for a reactor.

In summary, considerable progress has been made in understanding the physics of laser fusion, and the need for short wavelengths is now firmly established. The results of the next decade of experimentation should resolve the basic scientific question for laser fusion: Can ignition be achieved? An affirmative answer would bring the promise of abundant fusion energy for the 21st century one important step closer to its practical realization.

The Origin of Corn

Modern corn, the author argues, had not one ancestor but two: it is derived from a cross between a primitive corn and a perennial form of the wild grass teosinte. Breeding experiments support the theory

by Paul C. Mangelsdorf

There has always been a central mystery about corn: where and when did it arise, and—most important—from what? Corn as we know it is a thoroughly domesticated plant. It is the most efficient of all grasses at producing grain, but it is unable to survive without human help because it has no way of spreading its seed. The kernels on an ear of corn cling tightly to the rigid cob, and if the ear were simply allowed to drop to the ground, so many competing seedlings would emerge that in all likelihood none would grow to maturity. The ear in its hundreds of modern varieties was created by human beings for human purposes through centuries of selective breeding; it has no counterpart anywhere in the wild, or for that matter among other cultivated plants. Yet corn must originally have evolved from a wild plant. Which plant? The question has puzzled botanists and anthropologists for more than a century. I myself have devoted much of my time to it for nearly 50 years.

During that period there have been two main theories on the mystery of corn's origin. The older and still the more widely accepted theory holds that cultivated corn is descended from its closest living relative: annual teosinte, a wild grass found in Mexico and Central America. In 1939 Robert G. Reeves and I first suggested an entirely different view [see "The Mystery of Corn," by Paul C. Mangelsdorf; SCIENTIFIC AMERICAN, July, 1950]. We proposed that the ancestor of cultivated corn was a wild corn: specifically, a wild race of pod corn, which is a peculiar, primitive form whose kernels are individually enclosed in chaffy pods, or glumes. The wild pod corn, now probably extinct, would also have been a popcorn, meaning its kernels were small and hard. In this scenario annual teosinte, far from being the ancestor of corn, was itself derived from the hybridization of corn with *Tripsacum,* a wild grass that is corn's only other American relative.

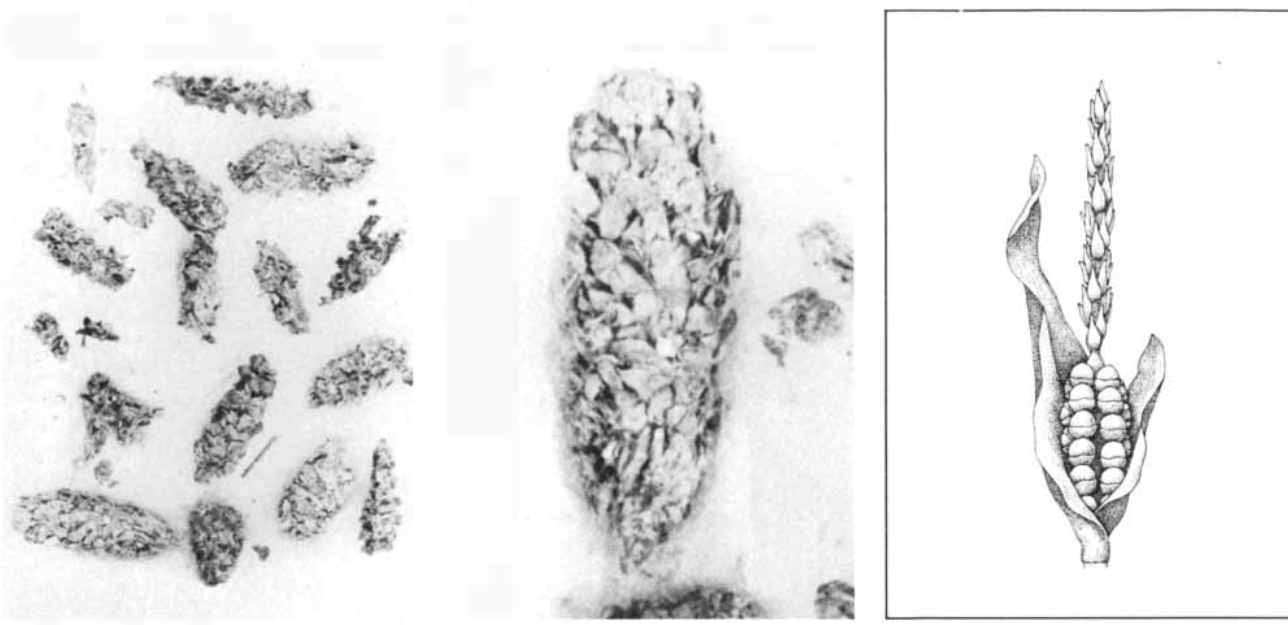
Forty-seven years later I still believe modern corn acquired its distinctive botanical characteristics, notably the ear, from a wild pod-popcorn. Two developments, however, have forced substantial modifications in the theory Reeves and I originally put forward. First, it has become clear, primarily from a comparison of pollen structures, that annual teosinte cannot be a hybrid derivative of corn and *Tripsacum*. Second, in 1979 a new form of teosinte was discovered that by virtue of its underground stems (rhizomes) and robust roots is a perennial plant rather than an annual one. (A perennial plant survives from year to year, whereas an annual plant such as corn completes its life cycle in a single season.) In my view this perennial teosinte, called *Zea diploperennis,* was the crucial missing link in the genealogy of both corn (*Zea mays*) and annual teosinte (*Zea mexicana*).

As I now see it both modern corn and annual teosinte are descended from the hybridization of perennial teosinte with a primitive pod-popcorn. This revised theory is strongly supported by the results of crossbreeding experiments. The experiments show that annual teosinte is, as I have always maintained, the hybrid progeny rather than the progenitor of corn, but that its other parent is *Z. diploperennis* rather than *Tripsacum*. And modern corn, although it inherited its distinctive botanical features from a wild corn parent, has received an equally significant contribution from *Z. diploperennis*. Indeed, it was the hybridization with perennial teosinte, probably some 4,000 years ago, that triggered corn's explosive evolution as a cultivated plant.

Corn was the most important crop in the Americas even before then, as it still is today. Moreover, there can no longer be any doubt that it is American in origin. The best evidence is archaeological. More than any other plant, corn documents its own history, because its hard cobs are well designed for preservation under a variety of conditions. The remains of prehistoric corn, including some 25,000 cobs, have been found in numerous sites from Arizona to Guatemala; in contrast, not a single cob dating unmistakably from before 1492 has been found in any part of the Old World. It was Columbus who discovered corn in Cuba and brought it back to Spain. Within a generation it had spread throughout Europe; within two generations it was being grown around the world in every region suitable for its cultivation. Only two other New World products, tobacco and syphilis, spread with equal rapidity.

By the time Columbus encountered corn, Indians were growing it from Canada to Chile. Corn was the basis of all the advanced pre-Columbian cultures, including those of the Incas of

MODERN CORN EARS vary widely from race to race in size, shape and color. The tremendous diversity of modern corn is the result of centuries of selective breeding and crossbreeding. Ultimately, the author suggests, the diversity is attributable to the hybridization some 4,000 years ago of a primitive cultivated corn with wild perennial teosinte. The hybridization gave human breeders a rich gene pool to work with, and it made corn a more mutable (and therefore more malleable) species. The ear in the lower right-hand corner is U.S. Corn Belt dent, the world's most productive corn. To the left of it is an ear of Cuzco Gigante, a Peruvian race with the world's largest kernels. Above the Cuzco is a tiny Lady Finger popcorn, and above that is the whitish ear of a Brazilian pod corn, whose kernels are enclosed in chaff. The specimens belong to the Botanical Museum of Harvard University.



FOSSIL CORNCOBS from a cave in the Tehuacán valley in Mexico date from about 5000 B.C. and are the oldest yet discovered. The cobs are shown life-size at the left, and the largest one is enlarged in the middle photograph. The long, chaffy glumes, which enclosed individual kernels, identify the cobs as those of a pod corn; because the kernels were small, the corn was probably also a popcorn. It may have been wild. Modern corn cannot survive in the wild because its ear is tenacious and is tightly wrapped in husks; as a result the seeds (kernels) cannot spread naturally. The tiny ear of wild corn, which is reconstructed in the drawing, would have been frailer, and at maturity the husks would have opened, enabling the kernels to disperse. The drawing is based on one by Walton C. Galinat.

South America and the Maya and Aztecs of Central America and Mexico, but it was first domesticated well before those civilizations flourished. The oldest corn specimens date from about 5000 B.C. No more than a few centimeters long, they were found in the dry, once inhabited caves of the Tehuacán valley, southeast of Mexico City, by the archaeologist Richard S. MacNeish of Boston University.

In the 1960's I studied these specimens intensively with Walton C. Galinat, who was then my assistant and is now at the Suburban Experiment Station of the University of Massachusetts in Waltham. We concluded that the oldest cobs were those of wild corn and that corn had therefore probably been domesticated a little after 5000 B.C. The tiny cobs, although unsegmented like those of modern corn, seemed frail enough to enable the seeds to spread and the plant to survive in the wild. Other workers later argued that the Tehuacán cobs were early cultivated corn. But everyone agreed that they were indisputable evidence of the existence 7,000 years ago of a plant with all the distinctive botanical characteristics of modern corn. The principal difference between the Tehuacán corn and modern corn is one of size.

Hence one can now give definite answers to the first two questions I posed, the questions of where and when corn arose in a modern, domesticated form. As MacNeish put it in a recent review article, "corn was domesticated in a relatively small area of Mexico north of Chiapas and south of Mexico City." Whether the most ancient archaeological cobs are wild corn or early cultivated corn, it seems safe to say, as MacNeish does, that "corn was domesticated well before 4000 B.C."

No similar consensus exists on the answer to the third question, that of corn's ancestry. The theory that corn is descended from annual teosinte has been the most popular one since the late 1960's, when it was revived in part through the work of George W. Beadle of the University of Chicago [see "The Ancestry of Corn," by George W. Beadle; SCIENTIFIC AMERICAN, January, 1980].

The main reason for believing this theory is the close genetic relation between cultivated corn and annual teosinte, which is now unquestioned. Both plants have 10 pairs of homologous (functionally identical) chromosomes, and the plants hybridize freely. The hybrids are usually highly fertile (unlike mules, say, which are hybrids of more distantly related species). When a corn-teosinte hybrid forms sex cells through the process of cell division called meiosis, each pair of homologous chromosomes—one from corn, the other from teosinte—match up virtually gene for gene. Furthermore, paired chromosomes exchange genes almost as often in the hybrids as they do in corn itself. The occurrence of such "crossing over" indicates that there must be a very close match between the genetic sequences of the two chromosomes.

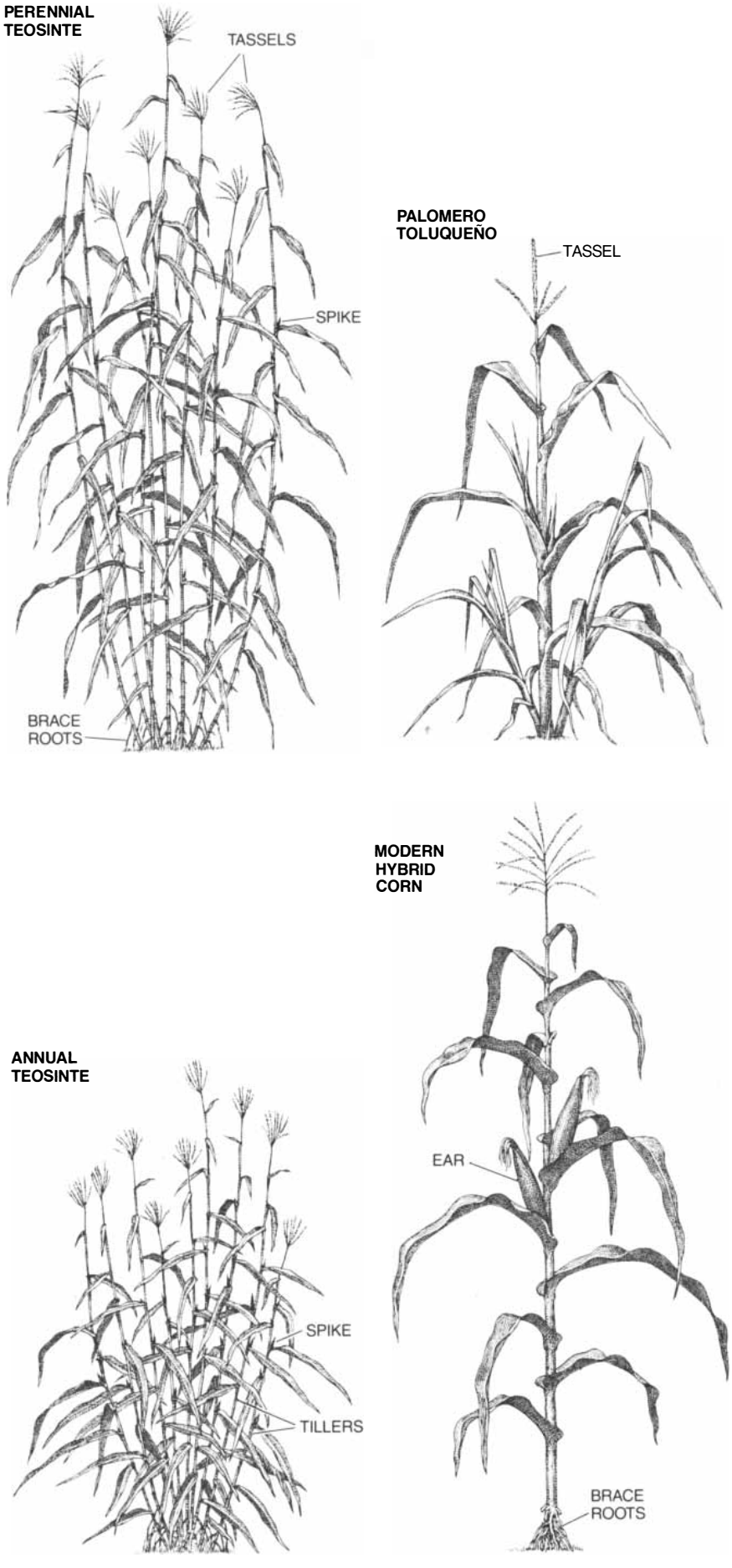
Nevertheless, the teosinte theory of corn's origin has serious flaws. Most of its various versions hold that teosinte was domesticated by Indians and gradually converted into corn through selective breeding. Yet there is no tangible evidence of any kind—archaeological, ethnological, linguistic, ideographic, pictorial or historical—to show that Indians ever grew teosinte as a crop. Again the most significant data are archaeological. In contrast to the thousands of corn specimens, only a few fragments of teosinte have ever been turned up at any of the early archaeological sites. Moreover, in those sites where both corn and fragments of teosinte or of teosinte hybrids have been found, the corn has consistently been older than the evidence of teosinte. "One might say there is not a grain of teosinte before that of corn," MacNeish has written, "nor a grain of evidence for the assertion that teosinte evolved into corn. . . ."

The most recent variant of the teosinte theory, advanced by Hugh H. Iltis of the University of Wisconsin at Madison, also runs afoul of the archaeological record. Iltis' hypothesis is brilliantly imaginative in trying to explain how the rigid ear of corn, with

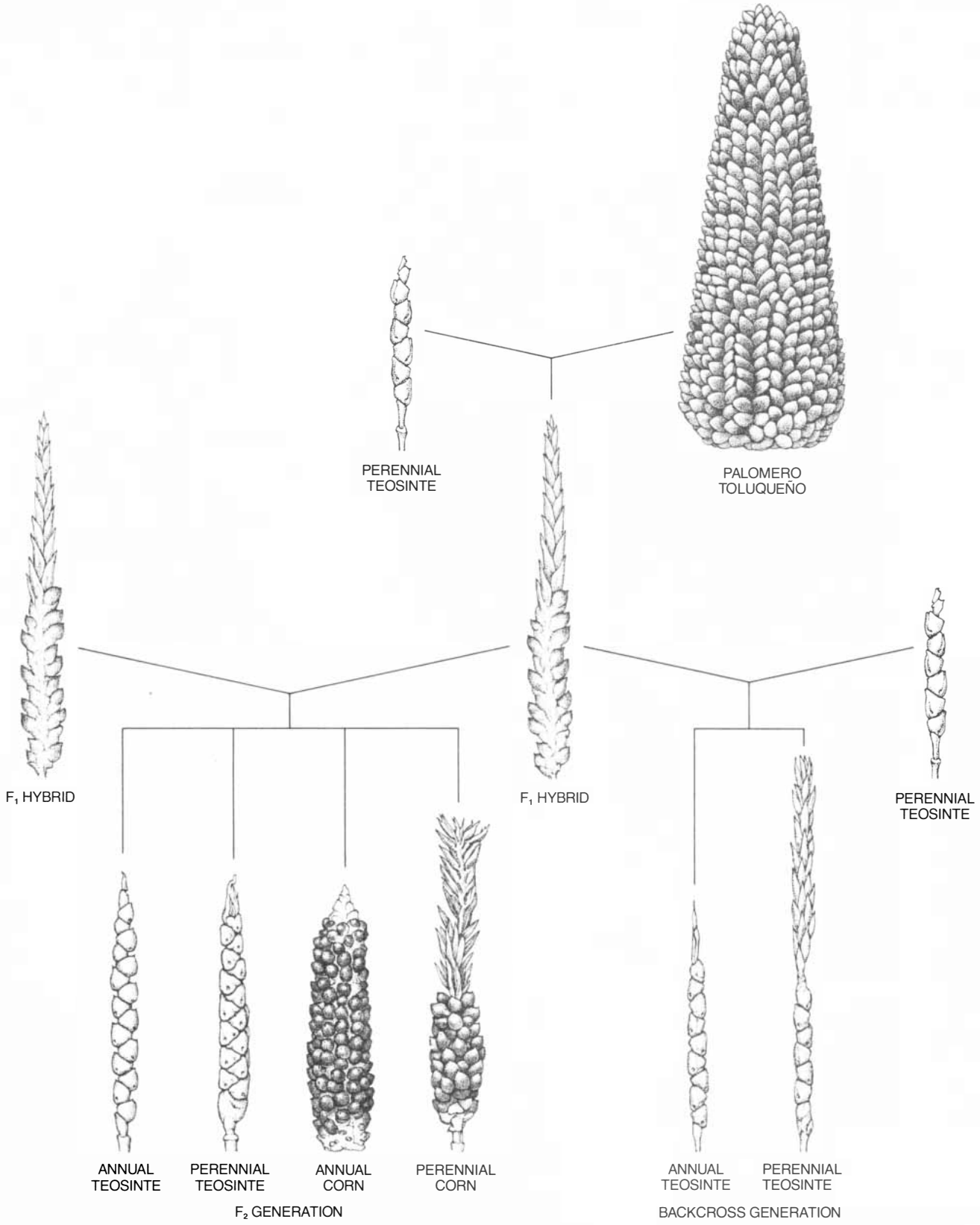
its many longitudinal rows of kernels, could have evolved from teosinte, whose own female seed-bearing organ, the spike, is small, segmented and brittle and has only two rows. Iltis thinks the corn ear evolved not from teosinte's spike but from its male, pollen-bearing flower cluster, the tassel, through a "catastrophic sexual transmutation." The problem is that the archaeological record contains no evidence that annual teosinte even existed when the catastrophic change is supposed to have occurred.

A further problem with all variants of the teosinte theory is that they do not take into account the discovery of *Z. diploperennis,* a discovery in which, ironically, Iltis played a major role. The plant was found growing wild in Jalisco, Mexico, by a Mexican graduate student, Raphael Guzmán, but it was Iltis who had organized the search for perennial teosinte. Although a different species of perennial teosinte had been known previously, that species was a tetraploid rather than a diploid, meaning that its cells contained four copies of each chromosome rather than the more usual two. John F. Doebley, a colleague of Iltis', demonstrated that *Z. diploperennis* is a diploid. Most important, it has the same chromosome number as corn and is cross-fertile with corn (whereas hybrids of corn and the tetraploid perennial teosinte are "triploids" that tend to be sterile).

The discovery of the new perennial teosinte changed rather drastically the terms of the debate on the genealogy of corn and teosinte. When the discovery was reported in 1979, H. Garrison Wilkes, a former graduate student of mine who is now at the University of Massachusetts at Boston, immediately recognized its importance. He thereupon postulated that the hybridization of the diploid perennial teosinte with a race of cultivated corn in an early



CORN AND TEOSINTE are closely related: they have the same chromosome number and hybridize freely. On both corn and teosinte the tassels (male flower clusters) are separate from the female, seed-bearing organs. The plants differ chiefly in the nature of their female organs. Whereas corn has two or three large, tenacious ears, teosinte has numerous small, brittle spikes that shatter when they mature. Another difference is that teosinte has numerous tillers, or side stalks. Many workers believe modern cultivated corn is descended from annual teosinte, which grows wild in Mexico and Guatemala. The author argues instead that both modern corn and annual teosinte are descended from a cross between an early cultivated corn, probably the Mexican popcorn Palomero Toluqueño or its precursor, and the perennial teosinte *Zea diploperennis*.



BREEDING EXPERIMENTS support the author's hypothesis for the origin of modern corn and annual teosinte. Examples of the spikes or ears of the various plants in the experiments are shown at roughly two-thirds their actual size. The primitive popcorn Palomero Toluqueño was crossed with perennial teosinte. The F_1 (first generation) hybrid was then crossed with itself and also backcrossed with perennial teosinte. Both the F_2 and the backcross generations included annual teosinte, whose segmented spikes are brittle and have two interlocked rows of kernels. The F_2 generation also included annual corn, which has unsegmented ears with multiple rows of kernels. A by-product of the experiments was perennial corn, which has been touted as a valuable new crop. The author considers its value doubtful: growing continuously on the same land, it would be particularly vulnerable to disease and insect damage.

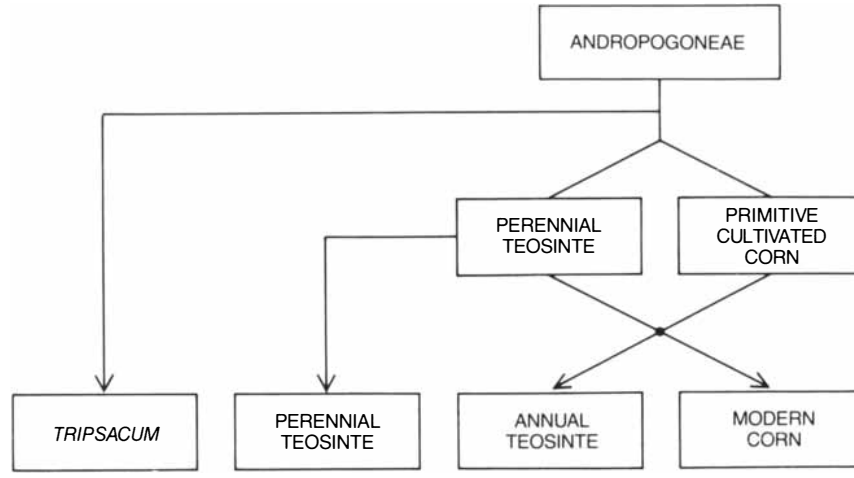
stage of domestication could have created the various races of annual teosinte. The Wilkes hypothesis appealed to me at once, in part because it was consistent with the archaeological evidence and in part because it was testable by means of hybridization experiments. It provided me with a powerful incentive to resume the experimental research I had retired from more than a decade earlier.

The early corn with which *Z. diploperennis* hybridized would have to have been a pod corn. All the earliest corncobs, particularly those found in the Tehuacán caves, bear the long, soft, chafflike glumes characteristic of pod corn. In having its kernels enclosed by chaff, pod corn resembles all other major cereals and virtually all other grasses, domesticated or wild; modern cultivated corn is unique in bearing naked kernels. This fact alone indicates that pod corn is primitive, as the French botanist Auguste de Saint-Hilaire first suggested in 1829. Saint-Hilaire thought the pod feature tended to be lost during domestication, and it seems clear that he was right. Even those investigators who believe corn evolved from annual teosinte generally accept the fact that the earliest corns were pod corns.

To test the Wilkes hypothesis I decided to cross *Z. diploperennis* with Palomero Toluqueño, a Mexican popcorn. Palomero Toluqueño is not a pod corn—its small, flinty kernels are not enclosed—but it is a primitive race that appears to be not far removed from pod corn. It or its pod-corn precursor are likely to have been grown in Jalisco, where *Z. diploperennis* was found, at the time the cultivation of corn began.

In spite of being primitive, Palomero Toluqueño has the four distinctive botanical characteristics of modern corn. First, it is an annual plant. Second, it has tenacious, unsegmented rachises, or cobs. Third, the "spikelets" that develop into kernels on the ear and into pollen-bearing flowers on the tassel occur in pairs. Fourth, the ear has many longitudinal rows of kernels. In contrast, *Z. diploperennis* is perennial; like annual teosinte it has brittle, segmented rachises that fall apart when they are ripe, and it has solitary spikelets and only two rows of kernels.

My experimental testing of the Wilkes hypothesis began early in 1979, when I grew plants of both *Z. diploperennis* and Palomero Toluqueño in my small garden in Chapel Hill. When the plants flowered in June, I successfully crossed them, employing the teosinte as the female parent and fertilizing it with pollen taken from the Mexican popcorn. The hybrid seeds ripened in the late summer, and to gain a generation I sent some of them to Argentina, where spring was just beginning. There a former associate, Julián Cámara-Hernández of the University of Buenos Aires, planted the seeds.



GENEALOGY diagrams the author's view of the ancestry of corn and of its relation to the wild grasses teosinte and *Tripsacum*. All three plants are descended from a remote ancestor that probably belonged to the Andropogoneae, a tribe that also includes sugar cane and sorghum. Sometime before 4000 B.C. corn was domesticated in Mexico from a wild ancestor that already had the botanical characteristics—notably a firm ear with multiple rows of kernels—of modern corn. In about 2000 B.C. a primitive cultivated corn hybridized with perennial teosinte. The results included annual teosinte as well as vigorous and productive races of corn. Wild corn was eventually swamped by cultivated corn and became extinct.

When the hybrid plants, members of the F_1 (first filial) generation, flowered early the following year, Cámara-Hernández crossed them with one another to produce an F_2 generation. In addition he backcrossed F_1 hybrids with the perennial teosinte parent. Since the annual teosintes differ from perennial teosinte chiefly in being annual and otherwise share the same botanical characteristics, they are likely to have arisen in a backcross population, which would tend to have a higher proportion of teosinte genes than either an F_1 or an F_2 population.

In fact, both the backcross and the F_2 populations contained significant numbers of annual teosinte plants. These results were subsequently confirmed by backcross and F_2 plantings in Texas, in which I was joined by my former graduate students Lewis M. Roberts and John S. Rogers (who are also now retired). Together the experiments gave strong support to the Wilkes hypothesis that annual teosinte is a hybrid derivative of perennial teosinte and corn. In the process they confirmed my long-held belief that annual teosinte could not be corn's ancestor.

Instead, a completely different view of corn's ancestry emerged from two observations we made in the breeding experiments. The first observation was that perennial teosinte's rhizomes, the underground stems that enable it to endure from year to year, are accompanied by large, fleshy roots (called prop or brace roots); to some degree this robust root system is transmitted to the hybrid progeny. The second observation was that the F_2 generation included plants that had all the botanical features of modern corn.

These findings suggested that the historical hybridization of *Z. diploperennis* with a primitive cultivated corn could have produced not only annual teosinte but also new races of corn more vigorous and productive than any that had preceded them. If Palomero Toluqueño is an example, the early cultivated corn was a puny plant with weak roots and stalks; it would have been highly susceptible both to drought and to excessive moisture, as well as to insect damage and disease. It may have been cultivated by the Mexican Indians only because it was better than other cereals available at the time, including perennial teosinte. The brittle teosinte spikes would have been difficult to harvest, and about half of the meager harvest would have consisted of the hard, indigestible shells that enclose the grains.

Yet whereas perennial teosinte is inferior as a crop, its robust roots make it resistant to drought, and it also resists many diseases to which corn is susceptible. In my view it transmitted some of these qualities to the offspring of its hybridization with corn. It was perennial teosinte that gave modern corn a good root system, strong stalks and resistance to some diseases. The

distinctive ear, with its tenacious cob and numerous paired rows of kernels, came from the primitive cultivated corn and ultimately, in rudimentary form, from wild corn. But without the traits supplied by perennial teosinte, corn would never have become the productive crop it is today.

The flow of teosinte genes into corn did not take place instantaneously. Rather, the various races of annual teosinte, which were also derived from the hybridization (hence their close genetic relation to corn), continued to transmit genes from *Z. diploperennis* to corn. Growing as weeds in cornfields, as they still do, the annual teosintes intercrossed with corn in generation after generation. Although they are not themselves the ancestors of modern corn, they may be regarded as carriers of valuable genes from corn's perennial-teosinte ancestor.

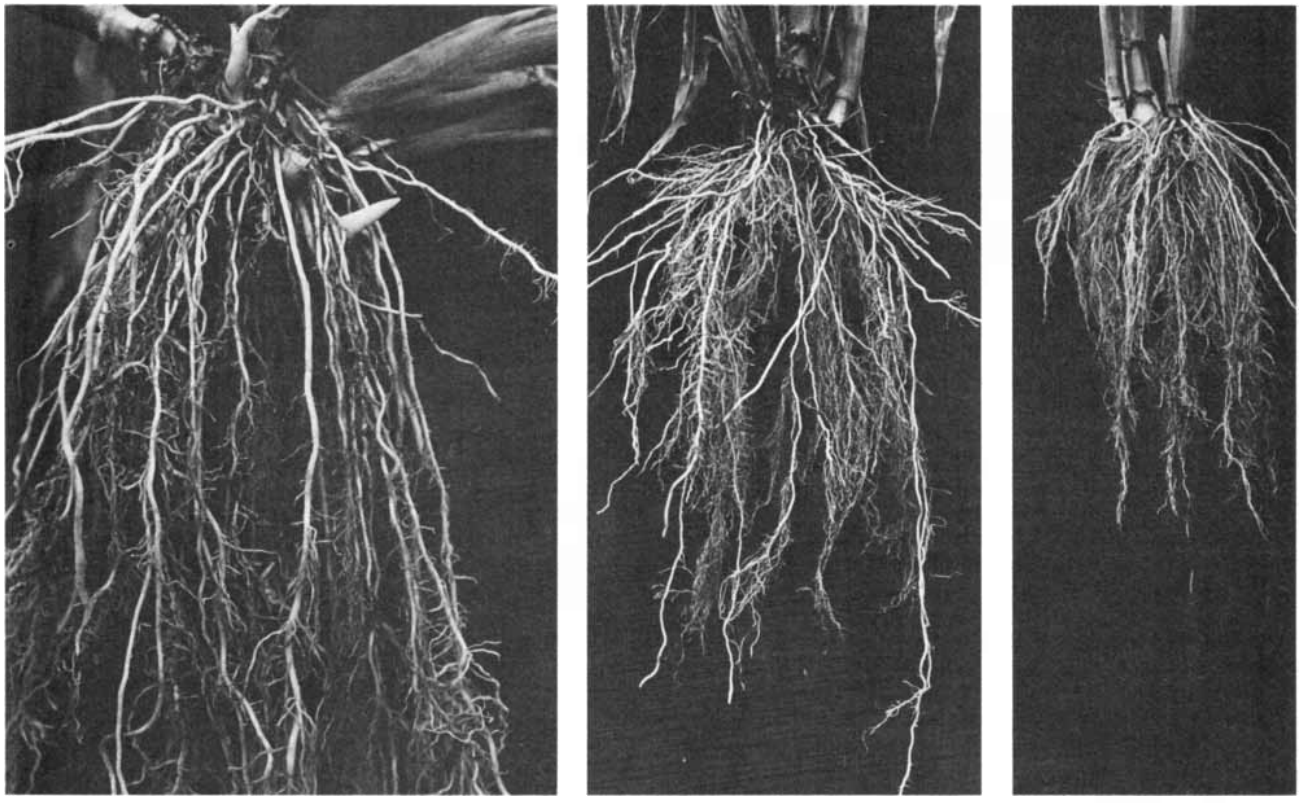
The confluence of the two genetic streams had profound effects. Not only did corn acquire some of the characteristics of perennial teosinte but also it became a more mutable species. Although the chromosomes of the two parents paired almost regularly in the course of meiosis, the paired chromosomes were not in perfect register with each other. The misalignment led to the creation of new genes when homologous chromosomes exchanged segments through crossing over. The ultimate result was a gene pool so extensive and so rich in variation that almost any kind of corn could evolve from it through natural and artificial selection. Over the centuries human breeders have skillfully exploited this resource to create hundreds of corn varieties.

When did the fateful meeting between *Z. diploperennis* and *Z. mays* take place? An approximate date can be assigned to the encounter based on the earliest appearance in the archaeological record of a corn showing signs of hybridization with teosinte. The oldest such corn is no more than 3,800 years old. Assuming it took a couple of centuries for the hybridized corn to spread, the hybridization may have occurred as recently as 4,000 years ago.

Long before they met in the Tehuacán valley, corn and perennial teosinte must have diverged from a common ancestor. About this ancestor almost nothing can be said other than that it probably belonged to the Andropogoneae, a grass tribe that includes the sorghums, the millets, sugar cane and a number of important pasture and forage grasses. How wild corn evolved from its remote ancestor, and in particular how the ear evolved, is still largely a matter of speculation.

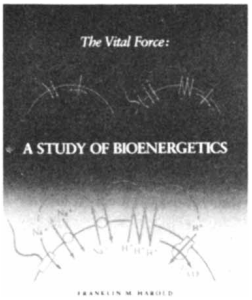
Corn has never been found growing in the wild, and I suspect it never will be. It would have become an endangered species at the latest when the Spaniards introduced grazing animals, notably the agile and voracious goat, into the New World. Even before then it might have been swamped by pollen spreading on the wind from thousands of acres of cultivated corn. Through cross-fertilization the wild corn would gradually have acquired large, solid ears that doomed it to extinction.

With the discovery of *Z. diploperennis,* however, one of corn's wild parents has at last been found. And with the recognition that modern cultivated corn has two parents, the long debate over its ancestry can be laid to rest. The question "Which is the ancestor of cultivated corn—teosinte or wild corn?" is now beside the point: both are. Other investigators would disagree, but in my opinion the mystery of corn has essentially been solved.



ROOT SYSTEM of perennial teosinte (*left*) includes thick, fleshy brace roots and is much more robust than that of the primitive corn Palomero Toluqueño (*right*). A hybrid of the two plants (*middle*) has roots of intermediate strength. This suggests that modern corn may have acquired its own robust roots, including brace roots, from the hybridization with perennial teosinte some 4,000 years ago. The short, pale shoots emerging from the roots of the teosinte are its rhizomes, which enable the plant to survive from year to year.

JUST PUBLISHED...
from W. H. Freeman and Company



THE VITAL FORCE
A Study of Bioenergetics

Franklin M. Harold
National Jewish Center for Immunology and Respiratory Medicine and University of Colorado Health Sciences Center

"Bioenergetics is the scientific study of vitality, not a discipline so much as a point of view, complementary to that of molecular biology."
—from the Preface

Thanks to the advances made during the past quarter century, the principles that govern biological energy transduction are now fairly clear, but their significance to cellular and molecular biology is not widely appreciated. The object of *The Vital Force* is to make bioenergetics accessible to a wider audience by means of an integrated, comprehensible survey of energy transactions in metabolism, membrane transport, movement, and information processing.

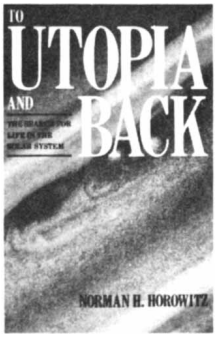
1986, 557 pages, 286 illustrations

BASIC CONCEPTS IN POPULATION, QUANTITATIVE, AND EVOLUTIONARY GENETICS

James F. Crow
University of Wisconsin at Madison

This new text is a brilliant integration of classical population genetics theory and modern molecular ideas. Written by one of the premier scholars in the field, it covers significant topics of current interest such as molecular polymorphism and evolution. Emphasizing theory, but supplying enough experimental data to place the subject in a real-world context, *Basic Concepts in Population, Quantitative, and Evolutionary Genetics* makes population genetics accessible to students whose primary interest is biology, not mathematics.

1986, 304 pages, 48 illustrations

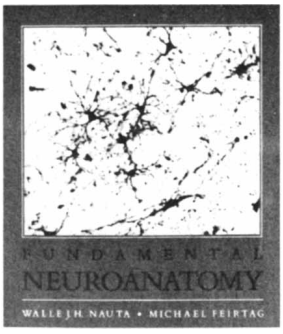


TO UTOPIA AND BACK
The Search for Life in the Solar System

Norman Horowitz
Professor Emeritus, California Institute of Technology

In this fascinating and highly readable work, the former head of the Jet Propulsion Laboratory's bioscience section for the six Mariner and Viking missions to Mars (1965–1976) provides a rare insider's view of the search for life on Mars and extends the lessons learned from the missions to the rest of the solar system. In addition, *To Utopia and Back* presents the author's first-hand knowledge and insights into the political atmosphere surrounding this scientific endeavor, some details of which are revealed in this book for the first time.

1986, 159 pages, 23 illustrations



FUNDAMENTAL NEUROANATOMY

Walle J. H. Nauta
Massachusetts Institute of Technology
Michael Feirtag
Board of Editors, *Scientific American*

Here, for the first time, is an account of the broadscale structure of the brain detailed and rigorous enough to serve the needs of medical students and researchers, yet accessible enough so that anyone who wants to learn about the brain can share in the modern understanding of the machineries underlying perception, thought, emotion, and behavior. The product of a decade-long collaboration between both authors, *Fundamental Neuroanatomy* is accompanied by more than one hundred dramatic photographs and drawings, many displayed full-page.

1985, 340 pages, 116 illustrations

Please send me

Quan.	ISBN	Title		Price
	1734-4	The Vital Force	(cloth)	$37.95
	1760-3	Basic Concepts in Population, Quantitative, and Evolutionary Genetics	(paper)	$15.95
	1766-2	To Utopia and Back	(paper)	$11.95
	1723-9	Fundamental Neuroanatomy	(paper)	$26.95
			Subtotal	$
		NY, UT, and CA residents add appropriate sales tax		$
		Add $1.50 postage and handling charge ($1.00 for each additional book)		$
			TOTAL	$

☐ I enclose a check made payable to W. H. Freeman and Company
☐ MasterCard ☐ VISA Account #_____
Expiration Date___/___ Signature_____
(CREDIT CARD ORDERS MUST BE SIGNED)

Name_____
Address_____
City_____
State_____ Zip_____
0227

Mail to: **W. H. FREEMAN AND COMPANY**
4419 West 1980 South
Salt Lake City, Utah 84104

The Solar Cycle in Precambrian Time

Periodic signals encoded in layers of Australian rock some 680 million years ago promise to advance the understanding of the sun's cyclical activity and its terrestrial consequences

by George E. Williams

In an age of satellites and interplanetary probes, it is perhaps ironic that some waterworn rocks in a eucalyptus-lined creek bed in South Australia promise to stimulate an important advance in the study of the sun and how it influences the earth. The rocks are red siltstones and fine sandstones, known collectively as the Elatina formation. To the casual eye they look undistinguished. Yet they incorporate a complex pattern of laminations whose thicknesses may well record cyclic variations in the relative values of mean annual temperature or mean summer temperature over many thousands of years of Precambrian time, 680 million years ago. If that is the case, the complex patterns of climatic cyclicity recorded by the Elatina formation are clearer than any previously recognized in geologic records or for that matter in other proxy, or indirect, climatological data such as tree-ring records. The patterns may have been caused by cyclic variations in the activity of the sun—variations strikingly similar to the cyclic activity the sun exhibits today, almost 700 million years later.

The effort to establish such a conclusion of course must have two foci, the study of the rocks and the study of the sun. The rocks (to take them first) were deposited near the end of the Precambrian era, in a world dramatically different from the earth as it is today. Much of the planet was in the grip of a severe ice age; indeed, the climate was so extraordinary that ice sheets and permafrost (soil frozen throughout the year) lay near sea level at equatorial latitudes. No plants grew on the barren land surface, and animal life from the seas of that time is unknown: the oldest recorded fossil remains of multicellular animal life are preserved in rocks deposited after the Precambrian ice sheets had disappeared. The world was still the domain of primitive algae and bacteria. Man could not have survived; the oxygen content of the Precambrian atmosphere probably was only a few percent of the present level.

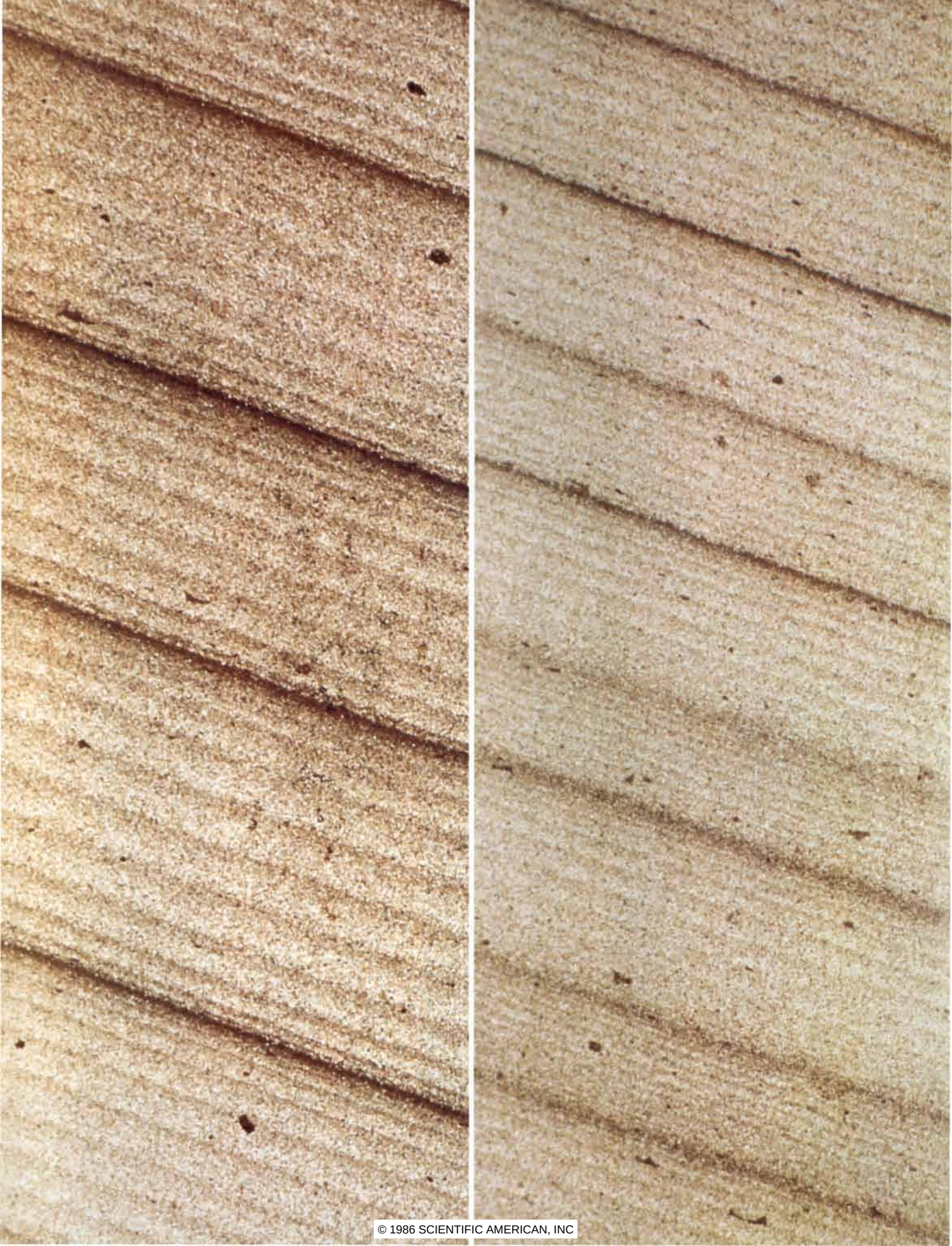
The ice age left its signature in the strata of most of the continents. In South Australia beds of boulders were deposited by melting icebergs in a gulf that marked the position of the Adelaide geosyncline, an elongated, sediment-filled depression in the crust. The adjacent terrain was marked by permafrost, by desert dunes and by a scattering of large lakes. In the cold, arid, strongly seasonal climate the lakes received periodic influxes of meltwaters turbid with sediment, and in several of the lakes cyclic variations in the volume of the meltwater left complex signals encoded in the thicknesses of successive silty laminations deposited on the lake beds. The Elatina formation almost surely arose in that way.

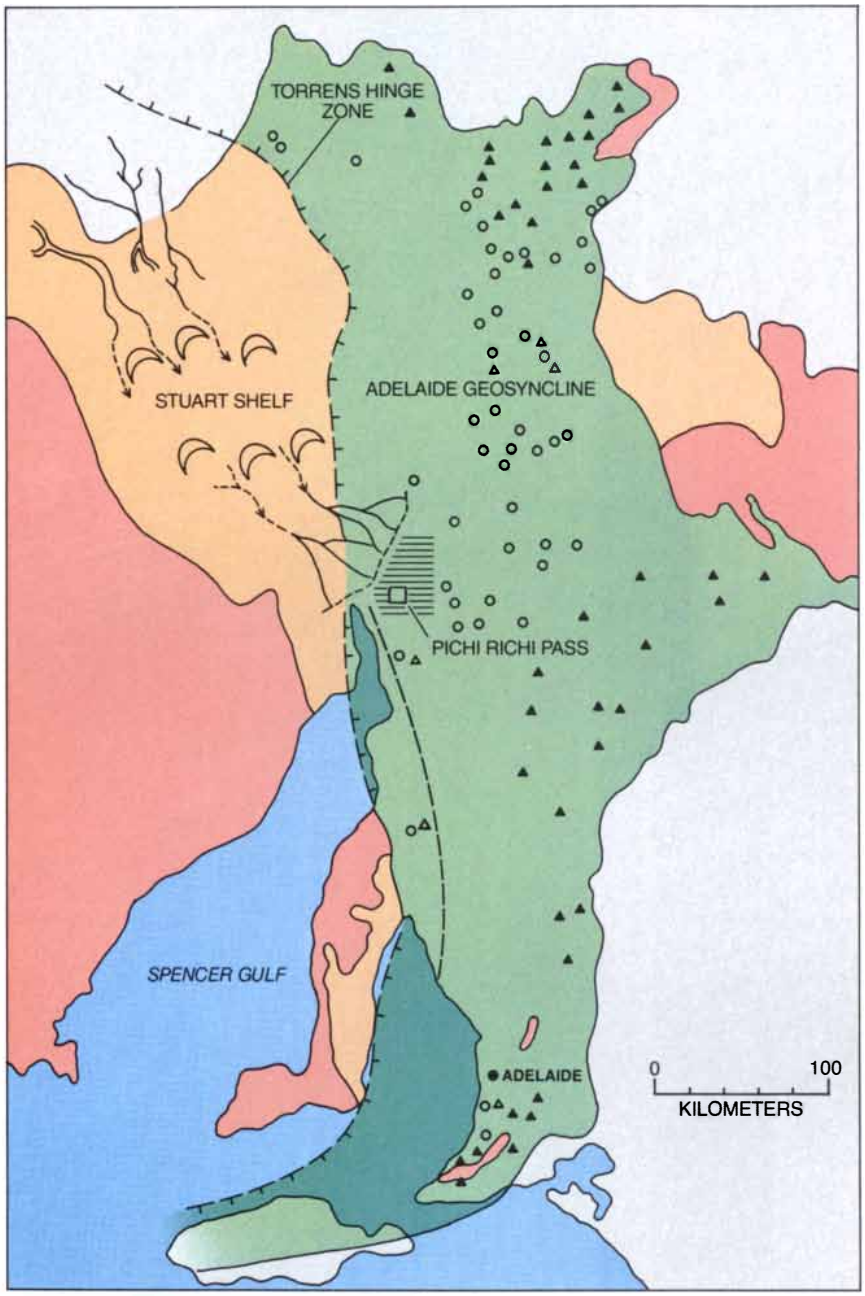
Eventually the climate improved. The permafrost thawed, the glaciers disappeared and the lake deposits were buried, lithified and preserved under the strata deposited by warm latest-Precambrian and Cambrian seas. Subsequent earth movements and erosion exposed the lake deposits; they can be found today in a creek bed in the Flinders Ranges, a ragged belt of folded sedimentary rocks that extends from near Adelaide northward into the arid interior of South Australia.

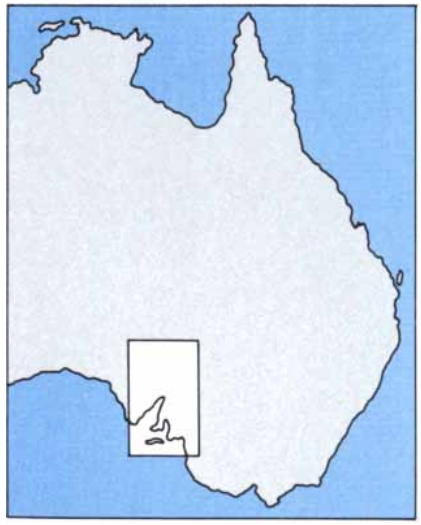
I became intrigued by the Elatina formation in 1979; its cyclicity seemed much too clear to have arisen through processes normally operating on the earth's surface. The modern study of periodicities in the activity of the sun had begun some 150 years earlier. In 1843 Heinrich Schwabe, a German amateur astronomer, pointed out an apparent 10-year cycle in the abundance of sunspots. He based his conclusion on 17 years of his own observations. The astronomical community was both startled and embarrassed; the cyclicity had escaped notice in spite of a telescopic watch of the sun for more than two centuries. Rudolf Wolf of the Zurich Observatory initiated an international program of observations of the sun; a comparable program continues today. Wolf also searched widely through earlier data on sunspots and concluded that a cycle averaging 11.1 years had existed for the years 1700 to 1848.

Wolf's reconstruction of sunspot

LAMINATED SANDSTONES AND SILTSTONES of the Elatina formation in South Australia may reveal the sun's influence on the climate of the earth late in the Precambrian era, a time when diverse animal life apparently had not yet evolved. The photographs show thin sections of two samples of the formation. They are remarkable for their cycles: groups of from 10 to 14 laminations bordered by darker bands in which the laminations are thin and closely spaced. The thicker laminations are as much as three millimeters wide; the thickest ones are found near the center of a cycle. The laminations, moreover, are graded: they consist of relatively coarse grains of quartz and feldspar toward the bottom of each lamination and finer, more clayey material toward the top. Similar graded laminations, called varves, are often deposited at the bottom of modern glacial lakes by the annual runoff of meltwaters from glaciers and snow. Each year the coarser grains settle during the summer; the finer material settles later, when the water is calm during winter. Cycle doublets (the alternation of a thick cycle and a thin one) are evident in the sample at the right.







GEOLOGY of South Australia in the late Precambrian era does much to establish the circumstances in which the Elatina formation was deposited. The geology was dominated by the Adelaide geosyncline, a sediment-filled crustal depression (*green*) bordered by platforms (*orange*) such as the Stuart shelf. The shelf was covered by permafrost and by dunes consisting of windblown sand. Annual meltwaters probably derived from winter snows. A number of lakes received the meltwaters, including Lake Elatina (*horizontal bars*). Geologic deposits in the geosyncline include sandstones and siltstones characteristic of deltas, lakes and streams (*circles*); beds of boulders deposited by icebergs, perhaps in marine environments (*closed triangles*), and isolated stones dropped into fine sediments from melting ice (*open triangles*). Stable areas of crystalline rock (*red*) such as granite were the source of much of the glacial debris.

numbers before 1850 became, as John A. Eddy of the National Center for Atmospheric Research in Boulder has noted, an accepted truth of solar history. Actually, however, Wolf's data are much inferior to the data collected by observations made in more recent years, and they deteriorate considerably with increasing distance into the past, toward 1700. The unreliability arises both from gaps in the observation records and from uncertainty as to the correction factors needed to calculate standardized sunspot numbers. For example, different observers have telescopes of different power, so that some observers detect sunspots others cannot see. It is noteworthy that charts of sunspot numbers beginning in 1700 show a change of structure in the mid-1800's. Such change can be attributed to the changes in the sources and reliability of the data from that time on.

It is now well established that the sunspot cycle is not strictly periodic but ranges from about nine to 14 years in duration. Moreover, the sunspot cycle is now known to be only one expression of solar variability: the incidence of solar flares and the flux of solar cosmic rays, ultraviolet radiation and X-radiation all vary directly with the sunspot cycle. Further still, the polarity of magnetic fields on the sun (the general field and local fields associated with sunspots) reverses systematically from cycle to cycle, producing a double cycle of 22 years called the Hale magnetic cycle after the pioneering investigator of solar magnetic fields, George Ellery Hale of the Mount Wilson Observatory. Although solar variability was first recognized by means of sunspot observations, it is more appropriate today to group these various phenomena as components of a "solar-activity cycle."

Does the cycle influence terrestrial weather and climate? After more than a century of investigation and debate, the relation remains unclear and controversial. The imprint of the 11-year sunspot cycle and the 22-year Hale solar-magnetic cycle has often been reported, in the form of periodicities discerned in records of such meteorological and climatological variables as rainfall, temperature, storm tracks, thunderstorm activity, surface atmospheric pressure and high-level pressure and winds. Invariably, however, the putative periodic signals are weak, and commonly the signals are of dubious statistical significance.

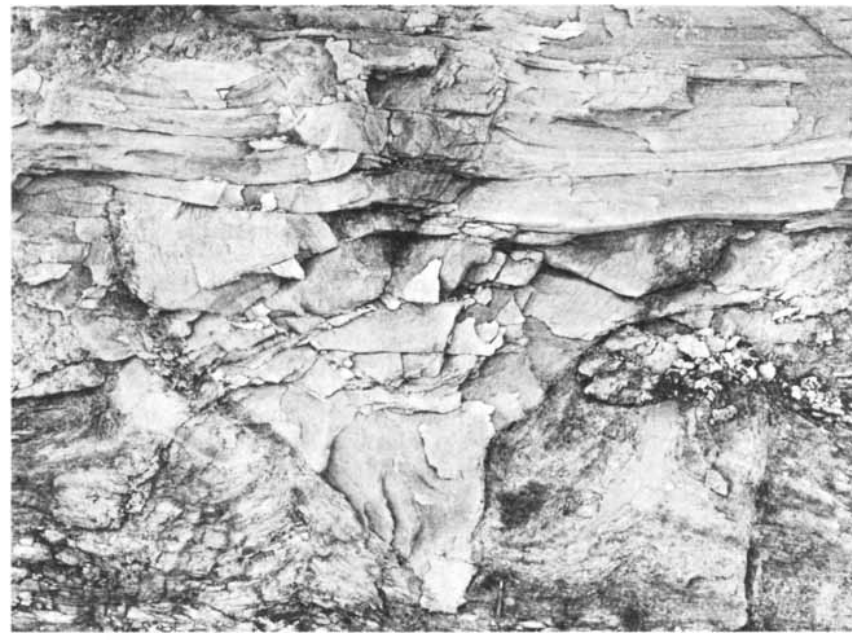
Effects of solar variability over longer terms have also been sought. In particular, the apparent cessation of sunspot activity in the late 17th and

early 18th centuries (a period called the Maunder minimum) has been linked to the coldest part of the Little Ice Age, a span of unusual cold in Europe extending from the 16th to the early 19th century. The veracity of the Maunder minimum has yet to be fully established, however, and so its suggested link with climatic deterioration must be viewed with caution. Many other studies have sought to find solar signals in such data as records of the thickness of tree rings. In general the studies have failed to produce unequivocal evidence of a link between terrestrial climate and the solar-activity cycle, or even evidence that the solar-activity cycle existed in past time.

Such evidence would yield insight into important problems in solar physics and solar-terrestrial science, the study of solar influences on the earth. In the first place, indisputable evidence of solar signals in proxy climatological data would settle the question of the existence of a mechanism linking solar and terrestrial events and might provide valuable pointers to its nature. None of the proposed mechanisms has won wide acceptance, although any solar-terrestrial link must involve the effects of the changing flux of solar cosmic rays, X-radiation or ultraviolet radiation on the earth's atmosphere.

Second, a record that allowed the operation of the solar-activity cycle to be traced prior to the brief era of visual observation might reveal the origin of the cycle. At least two explanations of solar activity are attracting attention today. The first supposes that the internal motions of the sun (caused by rotation and convection) interact with the large-scale magnetic field in the sun to produce a dynamo, the general term for a device in which mechanical energy is converted into the energy of a magnetic field. In short, the magnetic field in the sun is taken to be self-sustaining over a very long interval of time. It follows that the solar-activity cycle would be maintained with little overall change for perhaps billions of years.

The alternative explanation supposes that the large-scale magnetic field in the sun is the remnant of the field the sun acquired when it formed. In this model the field is not sustained against decay, so that solar activity runs down much more quickly than it does in the dynamo model; the character of the solar-activity cycle therefore could be expected to change over a long interval of time. The direct observations of the sun accumulated since 1850 span too short a time to reveal whether cyclical solar activity is a long-lived feature of the sun, as it would be in the dynamo model, or merely a transient phenomenon, as it would be in the competing model.



V-SHAPED SANDSTONE WEDGE about three meters high is good evidence that a cold, arid, strongly seasonal climate prevailed in the catchment area of Lake Elatina and governed sedimentation in the lake. The wedge was found in late Precambrian rock at the Cattle Grid copper mine, 120 kilometers northwest of a site where the Elatina formation comes to the surface, in a creek bed at Pichi Richi Pass. Such wedges form over hundreds or thousands of years as permafrost contracts and cracks each winter and the crack fills with windblown sand. They are forming today in the dry valleys of Antarctica, where the climate is arid, cold and strongly seasonal. The field of view includes a second, smaller wedge in flat-bedded sandstone of windblown origin immediately above the large wedge.

If the evidence in the Elatina formation is to be brought to bear on these questions, the late Precambrian geography and climate of the formation's surroundings must be reconstructed as accurately as possible. In the Flinders Ranges I have studied the formation at a place called Pichi Richi Pass. There the formation consists of sandstones and siltstones, which are granular sedimentary rocks. In siltstone the size of the grains is less than a sixteenth of a millimeter. The formation is about 60 meters thick, including a rhythmically laminated silty unit some 10 meters thick. The laminations accumulated on a delta at the western side of a large, late Precambrian lake, which I shall call Lake Elatina, situated between a frigid, windswept plain to the west and the shallow waters of the gulflike Adelaide geosyncline to the east. The dimensions of the lake are not known, but rocks comparable to the ones at Pichi Richi Pass crop out over a distance of about 30 kilometers. The laminated unit was deposited when the lake was unusually deep and widespread, perhaps during an interval of extreme cold.

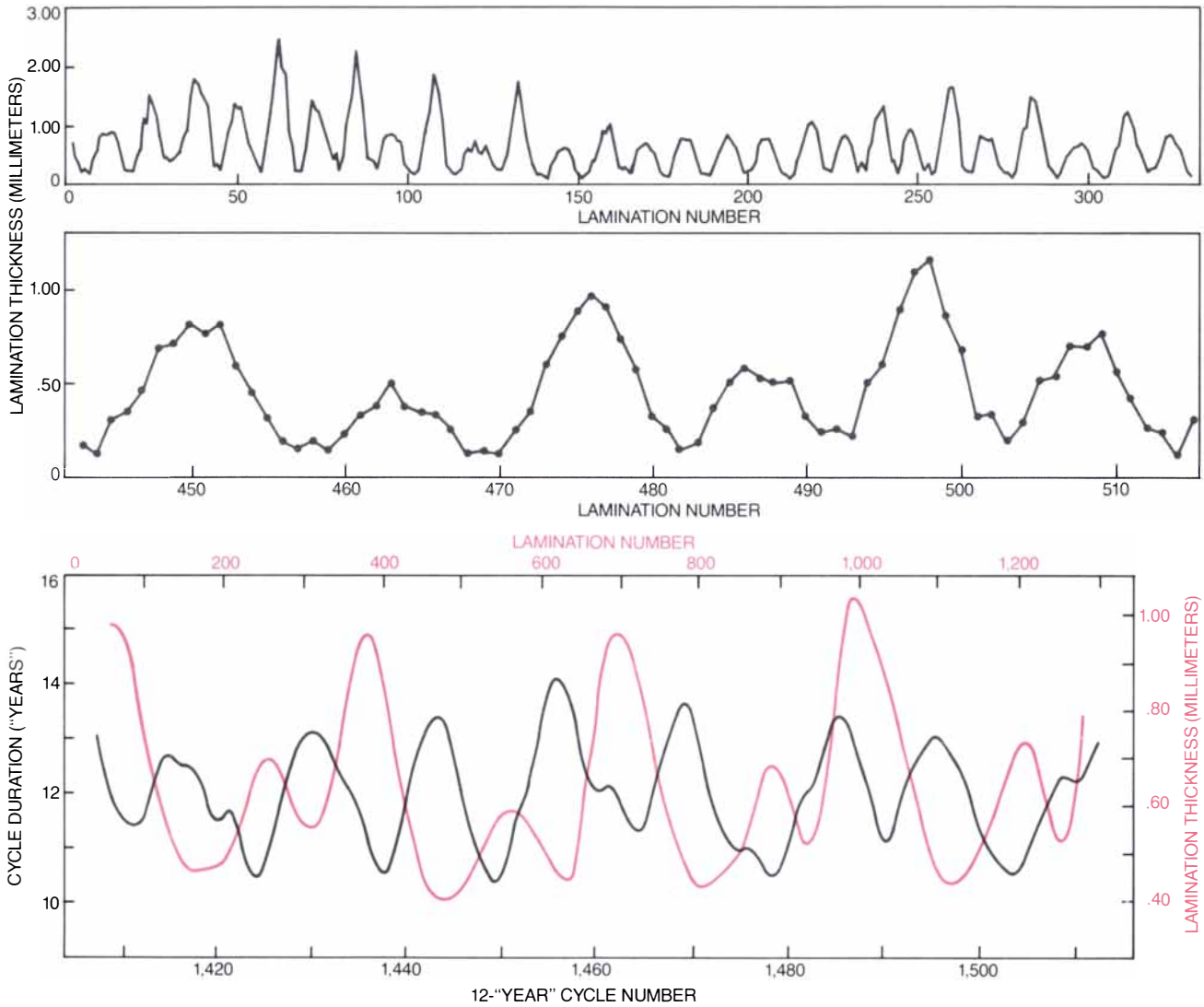
Evidence independent of the laminations themselves does much to show that the climate was cold, arid and strongly seasonal. The extreme aridity is confirmed by the presence of dune sands that interdigitate with the late Precambrian lake deposits. Moreover, geologic features indicative of a cold, arid, seasonal climate are distributed extensively in late Precambrian rocks throughout what must have been the catchment area of Lake Elatina. An example is afforded by large, V-shaped sandstone wedges within the Precambrian rocks. Similar wedges are found in modern periglacial areas, such as the dry valleys of Antarctica. They are formed by the winter contraction and cracking of the upper levels of permafrost and the filling of the cracks by windblown sand. Through annual repetitions of the process large wedges build up over periods of hundreds or thousands of years. The evidence suggests that the mean monthly temperature at Lake Elatina may have ranged from as low as −30 degrees Celsius, or even −40 degrees, in midwinter to a little above freezing (0 degrees) in midsummer.

The individual layers in the laminated unit of the Elatina formation range

from about .2 millimeter to three millimeters in thickness. Most of the laminations are graded, that is, they show a gradual reduction in grain size toward the top of the lamination. The laminations form conspicuous cycles, or groups, each consisting of some 10 to 14 laminations. The thickness of the laminations varies systematically within each cycle, attaining a maximum for the laminations near the center of the cycle. A given cycle is usually bounded by darker bands that turn out to be thinner, more clayey laminations crowded together. A cycle "doublet" consisting of a thick cycle and an adjacent thin cycle is common. Longer rhythms are also evident; they take the form of regular changes in cycle thickness over spans of many cycles.

The beds of modern glacial lakes suggest a way to interpret the regular, graded laminations of the Elatina formation. In such beds it is common to find that an extremely regular pattern of lamination in silts and fine sands is caused by the seasonal control on sedimentation. During the spring and summer thaw, meltwaters carry abundant suspended matter to the lake. The meltwaters are colder, and therefore denser, than the water of the lake; thus they form an underflow and deposit a thin, usually graded layer of silt and fine sand. Later, during the winter, a still finer, clayey material may settle when the lake has frozen over or a temperature inversion has overturned the lake waters. The material is so fine that wave action at the surface of the lake had kept it in suspension until winter came.

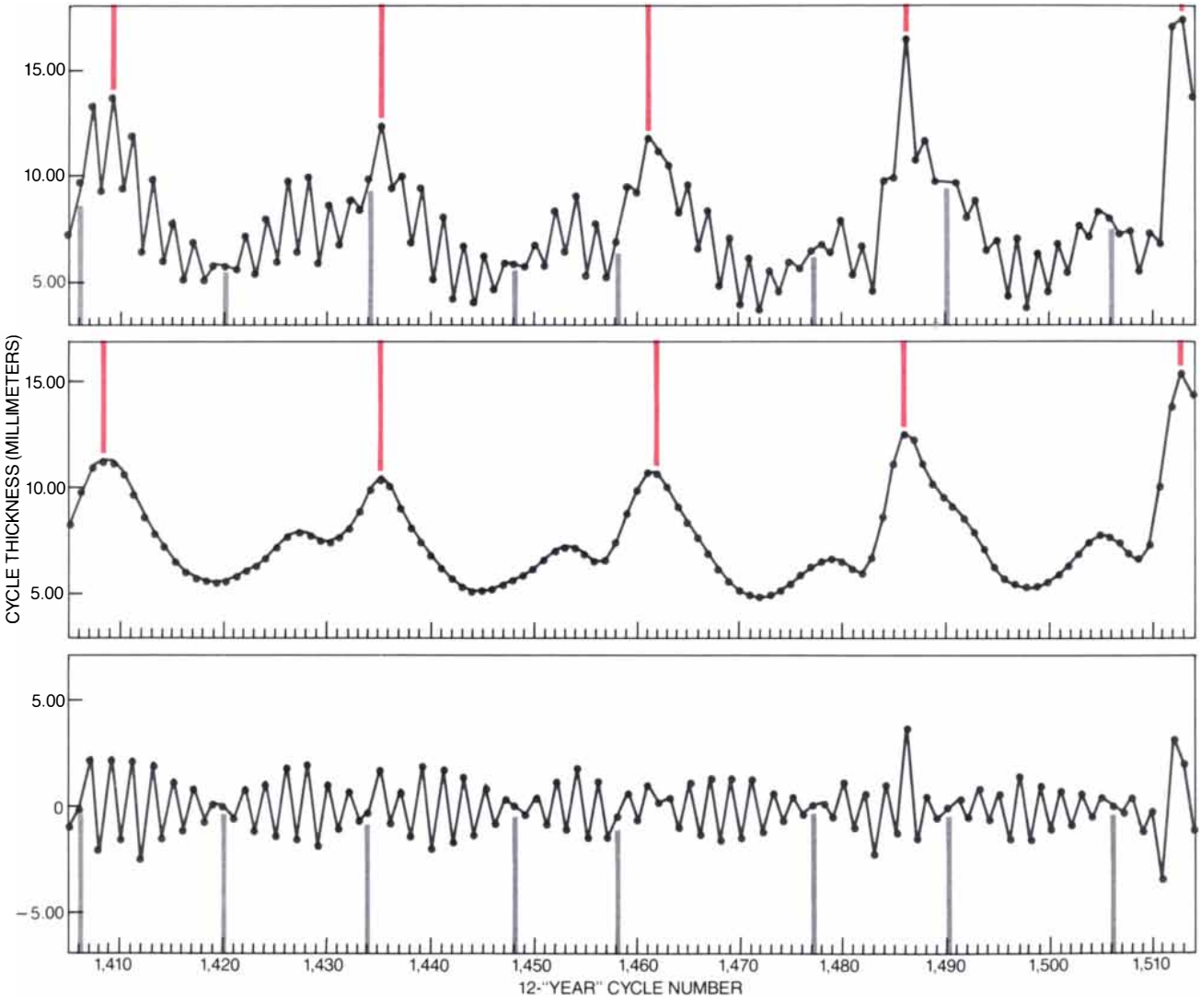
An extensive and regular or rhythmic sequence of either graded laminations or coarse-fine couplets is usually interpreted, therefore, as a set of varves, the geologic term for an annual deposit. For a number of reasons I believe the graded laminations in the Elatina formation are best interpreted in the same way. The fact that Lake



COMPLEX CYCLICITY in the laminations of the Elatina formation emerges from analysis of the "detailed sequence," an unbroken core sample with 1,337 laminations. The two top charts show thickness of successive laminations at two scales. Cycles are defined by thickness minimums; on the average each cycle contains 12 laminations. Cycles of relatively high and low amplitude tend to alternate. Moreover, the minimum lamination thicknesses tend to have roughly the same value, whereas the maximum thicknesses, at the center of each cycle, have a wide range of values. The bottom chart compares lamination thickness (*color*) and cycle duration, or number of laminations per cycle (*black*), over the entire detailed sequence. Both variables show systematic variation: the thickest laminations occur about every 26 cycles, whereas the cycles with the greatest number of laminations occur about every 13 cycles. The thickest laminations tend to occur in the shortest cycles. The curves have been filtered, or smoothed, to emphasize the long-term periodicities.

Elatina received meltwaters from a catchment area that had a markedly seasonal, arid climate implies that the discharge of meltwater into Lake Elatina was strongly controlled by the climate; disruptions of the cyclic seasonal pattern by storms must have been few and intermittent. Hence the seasons would have exerted strong control on sedimentation in the lake. In fact, the graded laminations resemble the delicate but distinct modern glacial varves that typically are laid down in relatively deep water some distance from the source of the sediments. The thinnest laminations in the Elatina formation presumably reflect the coldest part of the climatic spectrum. Their clayey composition can be attributed to the deposition of very fine material because of notably widespread freezing or major overturn of the cold surface waters of Lake Elatina.

My conclusion, by 1981, that the Elatina laminations are varves opened up a remarkable prospect. To the extent that the thicknesses of varves at a particular site reflect the volumes of meltwaters (and their suspended matter) entering a lake, varves may provide a relative measure of a succession of summer or annual mean temperatures. Thus in the Elatina formation the thicknesses of the presumed varves and varve cycles might constitute proxy climatological data recording mean annual or summer temperatures over many thousands of years of late Precambrian time, almost 700 million years ago. Moreover, the cyclicity in the Elatina laminations put me in mind of the solar-activity cycle. More information was needed than the limited rock outcrops could provide. Thanks largely to the enthusiasm and support of the late Ronald G. Giovanelli, former chief of the division of physics of the Commonwealth Scientific and Industrial Research Organization in Sydney, the CSIRO agreed to

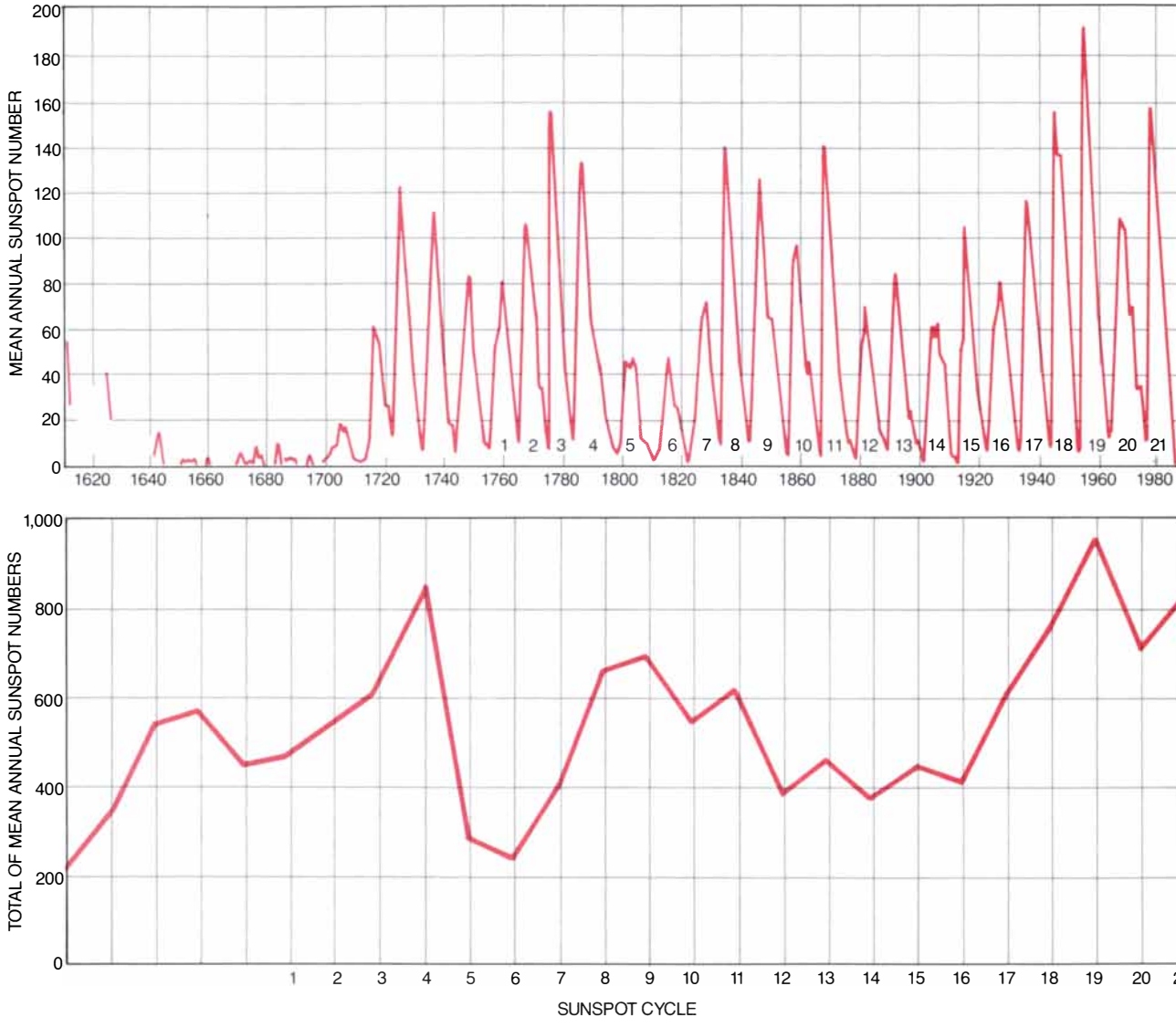


FURTHER CYCLICITY emerges from an analysis of the "long sequence," a record of 1,580 cycles, or roughly 19,000 laminations, spanning nearly the full 10-meter thickness of the laminated rock in the Elatina formation. The top chart plots the thickness of entire cycles in a sample of the long sequence. Distinct maximums appearing every 25 to 27 cycles (*colored lines*) define a periodicity called the Elatina cycle. The data also show a sawtooth pattern reflecting the alternation of relatively thick and relatively thin cycles. In the middle chart the data have been filtered, as in the illustration on the opposite page, to emphasize long-term periodicities. The highest peaks represent the Elatina cycle; the lower peaks represent one of its harmonics. The bottom chart was made by subtracting the values in the middle chart from those in the top one. What remains is a sawtooth pattern in a series of envelopes. Periodic 180-degree phase shifts, or reversals in the alternations composing the sawtooth pattern (*gray lines*), lie at or near the necks between envelopes.

sponsor drilling in the Elatina formation. The goal was to obtain a complete cored sequence of the 10-meter thickness of rhythmic lamination. The drilling was done in December, 1982. Since some core would be lost through rock fracturing, a group of three vertical holes was drilled to allow the matching of cores and the construction of the longest possible continuous varve record.

During 1983 I logged the drill cores and compared and correlated the three sequences. I began by placing small, overlapping strips of clear tape on the cores and carefully marking on the tapes the thicknesses of cycles that lay between conspicuous darker bands in the rock. I then transferred the strips of tape to large transparent sheets. Superposing the sheets established correlations among the three cores and made it possible to bridge stratigraphic gaps caused by core breakage. The result was a stratigraphically continuous log 9.39 meters long, spanning 1,580 cycles. The conspicuous cycles comprised from eight to 16 varves; the average was 12. Hence the entire log covered an estimated 19,000 "years." (The word years in quotation marks signifies varve time.) I shall refer to this log as the "long sequence." In addition overlapping photographic enlargements were made of thin sections from a single piece of core nearly a meter long. These prints constitute what I shall call the "detailed sequence." It is an unbroken record of 1,337 varves whose thicknesses can be readily measured.

In the next part of the project the thickness of varves and cycles in the two sequences was measured with a precision of .01 millimeter in the Laboratory of Tree-Ring Research at the University of Arizona in Tucson by means of equipment designed by William J. Robinson, the director of the laboratory, for the precise measure-



CYCLIC ACTIVITY OF THE SUN has a pattern much like the one encoded in the Elatina formation. The top chart shows the mean annual sunspot number since 1610. A change in the pattern occurs between cycle 7 and cycle 8, or at about the time the amateur astronomer Heinrich Schwabe began making systematic sunspot observations. The earlier data become progressively less reliable toward the beginning of the record, casting doubt on the reality of the Maunder minimum, a period of low solar activity said to have begun about 1650. From 1850 onward the data show alternate cycles of relatively high and low amplitude, creating the sawtooth pattern displayed in the bottom chart. The pattern is much like that of the Elatina formation's cycle thickness. The pattern persists in subdued form across the peak of cycle 19; the value for cycle 21 is based on the prediction that the cycle will have ended by about June, 1986.

ment of tree rings. The data were then transferred to a computer at the Lunar and Planetary Laboratory of the University of Arizona for analysis, which I did in association with Charles P. Sonett.

The measurements of varve thickness in the detailed sequence revealed a complex interweaving of cyclicities [*see illustration on page 92*]. The cycles themselves show an overall tendency to positive skewness, that is, the rise to maximum varve thickness within a cycle tends to be steeper than the fall to minimum thickness. The mean ascent period is 5.57 "years" and the mean descent period is 6.45 "years." The succession of cycles displays further rhythms. The cycles containing the thickest varves occur on the average every 26 cycles. Intermediate peaks in varve thickness occur between these highest peaks. In contrast, the cycles containing the greatest number of varves occur on the average about every 13 cycles. Intriguingly, the two rhythms are negatively correlated: the thickest varves tend to occur in the cycles having the fewest varves, that is, the cycles of shortest duration. In other words, strong cycles tend to be brief and weak cycles tend to be long.

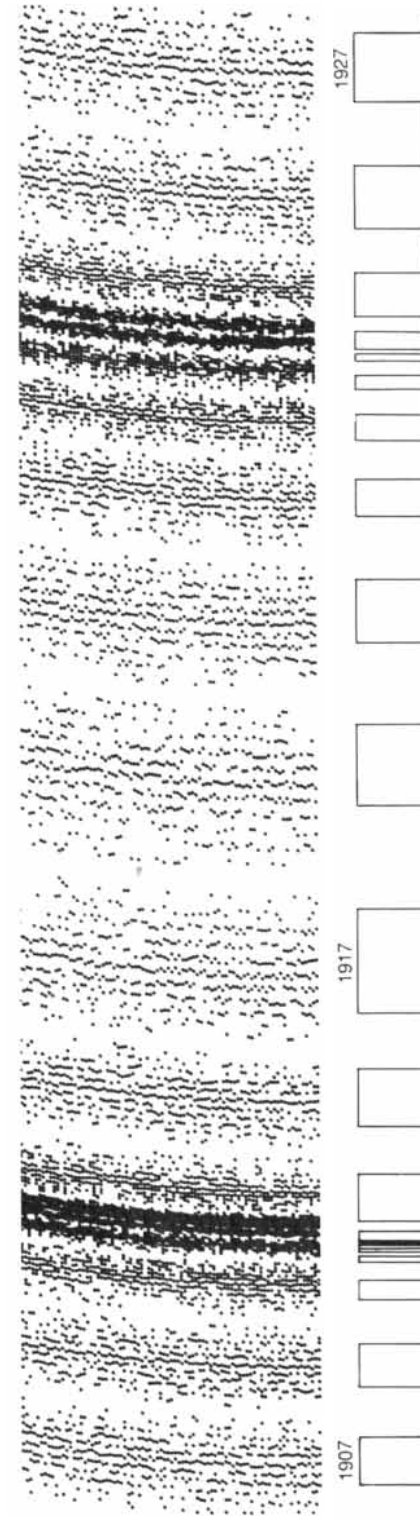
The long sequence of 1,580 cycles likewise contained a complex pattern of rhythms [*see illustration on page 93*]. Here, however, the periodicities identified by visual inspection and mathematical analysis of the thicknesses of the cycles prove to range from about four cycles to as many as 720 to 780. (That is, each rhythm has a characteristic duration, or repetitive length, and these lengths vary greatly from one type of periodicity to another.) The identified periodicities include the "Elatina cycle," a prominent rhythm averaging about 26.1 12-"year" cycles in duration, or about 275 to 335 "years." The periodicities also include a number of rhythms that are harmonics of the Elatina cycle: rhythms whose frequencies (the reciprocal of the duration) are whole-number multiples of the frequency of the Elatina cycle. The mean durations of the identified harmonics are 314, 157, 105, 79, 63, 52 and 45 "years."

Superposed on the Elatina cycle is a sawtooth, or zigzag, pattern resulting from the characteristic alternation of relatively thick and relatively thin 12-"year" cycles. The pattern exhibits 180-degree phase shifts (that is, the thick-thin alternation reverses) at intervals ranging from nine to 23 cycles; the average interval between shifts is 14.6 cycles. The mean period for a 360-degree change of phase is therefore 29.2 cycles, or 3.1 cycles longer than the mean period of the Elatina cycle. Hence the positions of features of the sawtooth rhythm gradually change along the sequence with respect to the features of the Elatina cycle.

Armed with this statistical analysis of the signals in the Elatina formation, I was in a position to compare the Elatina record with the modern record of solar activity. The two show similar features. Both records (analyzed mathematically) have periods of between nine and 14 years, as well as periods of about 22 to 25 years and periods of about 90 to 110 years. (The difference in mean period between the Elatina 12-"year" cycles and sunspot 11-year cycles may be of little significance, since both records show a tendency to systematic variation in cycle duration over many cycles; the reliable solar record may be too brief to yield a truly representative mean period.) The longer Elatina harmonics of about 157 and 314 "years" are comparable to climatic periods (possibly of solar origin) that have been identified for recent times in tree-ring studies. In both records the minimums (the thinnest varves in the Elatina formation, the smallest mean annual sunspot numbers in modern solar-activity records) occur at a more or less uniform level, whereas the maximums have a wide range of values. In both records the amplitudes of cycles vary systematically over a number of cycles. In both records sawtooth alternations between high and low cycles are common. In both records the individual cycles show a tendency to positive skewness. In both records the strong cycles tend to be of short duration.

These similarities argue for a direct connection, through climatic temperature variability, between varve thicknesses in the Elatina formation and solar activity in late Precambrian time. In other words, an increase in solar activity caused a corresponding increase in climatic temperature, which in turn entailed a greater annual discharge of meltwaters and the deposition of thicker varves on the bed of Lake Elatina. The study of recent varves in Skilak Lake, a glacial lake in Alaska, suggests that a solar signal, albeit a very weak one, has been recorded in this way. Further support is provided by the work of Ronald N. Bracewell of Stanford University, who has shown that an Elatina-like varve-cycle pattern can be generated artificially from a sequence of mean annual sunspot numbers by positing a direct relation between varve thickness and sunspot number.

The apparent strong influence of



COMPUTER-PLOTTED PATTERN resembling that of the laminations in the Elatina formation was generated from sunspot abundances for the years 1907 through 1927 by Ronald N. Bracewell of Stanford University. Bracewell assumed that solar activity, as represented by the mean annual sunspot number (*left*), has a direct relation to the thickness of laminations. In the parts of the pattern corresponding to sunspot minimums the layers coalesce into dark bands.

solar cyclicity on the earth's climate nearly 700 million years ago carries important implications for solar-terrestrial science and solar physics. The findings immediately encourage one to ask how the solar cycle could have affected the climate so markedly in late Precambrian time when today a solar signal is barely if at all discernible in patterns of weather and climate. One promising idea is that the great sensitivity of the Precambrian climate to solar variations resulted from the small oxygen content of the Precambrian atmosphere.

It is thought that in Precambrian time, before plant life began to liberate oxygen on a wide scale as a by-product of photosynthesis, the oxygen content of the earth's atmosphere was only a small fraction of the present value. The small oxygen content would have enabled solar ultraviolet radiation to penetrate deep into the atmosphere before being absorbed by molecular oxygen, converting the oxygen into ozone (O_3), and so would have given the atmosphere's ozone layer a far lower altitude than it has today. As a result the changes in temperature and other properties of the ozone layer brought on by cyclic variations in solar ultraviolet flux would have occurred at low altitude and might have directly influenced ground-level temperatures. With the increase in oxygen content that came in Paleozoic time, the terrestrial effects of the solar-activity cycle would have been pushed to a height where they are now barely able to influence weather and climate. The idea champions solar ultraviolet flux as the dominant link between solar cyclicity and climatic response.

The Elatina data also imply that the sun's activity has not changed greatly over the past 700 million years. The data therefore lend support to solar models that invoke a self-sustaining dynamo in explanation of the solar-activity cycle. The data may further discriminate between dynamo models that predict the random recurrence of Maunder-type minimums and models requiring only cyclic modulation of solar activity. In particular the Elatina data show deep minimums in cycle heights about every 275 to 335 "years," but nowhere in the 19,000 "years" of continuous signals is there a cessation of cyclicity comparable to the alleged Maunder minimum. This finding may imply that the Maunder minimum is an artifact of insufficient or inaccurate data. Indeed, naked-eye observations made in China during the 17th and early 18th centuries provide no support for an absence of sunspot activity in that period.

The nature of solar activity before about 700 million years ago cannot be inferred with any reliability from the known rock record. The apparent stability of the sun's behavior since that time accords well with observations of the activity of stars resembling the sun. Such observations indicate that variations in activity similar to those of the sun are largely independent of the star's age and its period of rotation. The findings suggest that when a star early in its life settles down to a fairly low rotation rate (comparable to the sun's rate of roughly one rotation per month), it can be expected to maintain a relatively stable activity cycle on a billion-year time scale. Still, the evidence of complex rhythms in the late Precambrian solar cycle implies that solar processes are more complex than any existing model would suggest. The prospect, therefore, is that the long, detailed record of solar activity provided by the Elatina formation could prove valuable in developing new models of the physics of the solar-activity cycle.

One important practical application of the Elatina data may be the forecasting of the relative amplitude and duration of future solar-activity cycles. If the solar maximum that occurred in 1957 is equated with the maximum of the 314-"year" periodicity in the Elatina record, a number of predictions emerge. Future solar-activity cycles should display an overall decline into a deep trough in the course of the next nine or 10 cycles. A 180-degree phase shift in the sawtooth (strong cycle, weak cycle) pattern should come, probably within the next few cycles. (The past 13 cycles have maintained the phase of the alternating pattern, and the mean duration between phase shifts in the Elatina record is 14.6 of the record's 12-"year" cycles.) Finally, the mean duration of the solar-activity cycle should increase to a value exceeding the mean for the past 100 years. This increase in duration will accompany the decrease in amplitude.

The data encoded in the Elatina formation promise, then, to provide a stimulus in several related fields: the study of the earth's atmosphere and its evolution; the search for mechanisms by which the sun might affect the earth's climate; the modeling of the physics of the solar-activity cycle; the prediction of future solar activity. Ancient rocks bearing annual layers, particularly those of Precambrian age, should be viewed as potential solar observatories that may greatly advance the understanding of solar processes and the sun's linkages to the earth.

No other system of keeping up can compare with ours.

YOUR SYSTEM: a time-consuming, futile struggle to keep up with the information explosion.

The classic texts are convenient references—but the information they contain is obsolete before publication.

Like many physicians, you probably rely on the texts you first used in medical school. But even using the most recent editions, you find material that no longer reflects current clinical thinking—and they lack the latest information on such topics as herpes, oncogenes, AIDS, and photon imaging.

Reading stacks of journals alerts you to recent developments—but can't give you quick answers on patient management.

Struggling through the hundreds of journal pages published each month—even on only the really significant advances in the field—is arduous and memory-taxing. And it's a task that costs physicians valuable time—their most precious resource.

Review courses cover clinical advances—but, months later, do you recall the details of a new procedure or unfamiliar drug?

Seminars can also be costly and make you lose valuable time away from your practice—expenses that may amount to several thousand dollars. And, the speaker's skill often determines how much you learn.

OUR SYSTEM: a rewarding, efficient way to keep yourself up-to-date—and save hundreds of hours of your time for patient care.

A comprehensive, 2,300-page text in two loose-leaf volumes, incorporating the latest advances in medical practice as of the month you subscribe.

This superbly designed, heavily illustrated resource, called "the best written of all [the internal medicine] books" by *JAMA* (251:807, 1984), provides a practical, comprehensive description of patient care in 15 subspecialties. And, because the text is updated each month, the clinical recommendations reflect all the current findings. A practice-oriented index and bibliography of recent articles further enhance the efficiency of the text.

Each month, six to nine replacement chapters to update your text *plus* new references, a four- to five-page news bulletin, and a completely new index.

You'd have to read hundreds of journal pages each month—and memorize the contents—to get the same information SCIENTIFIC AMERICAN *Medicine* contains. With our updated text, you read only the information you really need. Our authors, largely from Harvard and Stanford, sort through the literature and monitor developments, incorporating the significant advances into our chapters.

At no additional cost, a 32-credit CME program, to save you valuable patient-care time and the expense of attending review courses.

Earn 32 Category 1 or prescribed credits per year with our convenient self-study patient management problems; each simulates a real-life clinical situation. Choose either the complimentary printed version or, at a modest extra charge, the floppy-disk version for your IBM® PC/PC Jr., Macintosh,™ or Apple® IIe or II+ (or compatible).

☐ **Yes,** I'd like to try the SCIENTIFIC AMERICAN *Medicine* system. Please enter my subscription at a first-year price of US$245.*

☐ Also enroll me in the CME program. I'd prefer:

☐ the printed version at **no additional charge.**

☐ the floppy-disk version at $US97* additional. Computer type or compatible:

☐ IBM® PC/PC Jr. (128K RAM, DOS 2.0/2.1) ☐ Macintosh™ (512K RAM)

Apple® II+ (64K RAM) 80-col. card by: ☐ Apple® ☐ Videx® ☐ Apple® IIe (64K RAM)

☐ Enroll me in the floppy-disk CME **only** at US$152.* (Note model above.)

☐ Check enclosed* ☐ Bill me Signature _____

☐ VISA ☐ MasterCard Exp. Date _____ Account No. _____

Name _____ Specialty _____

Address _____

City _____ State _____ Zip Code _____

Or, call toll-free: 1-800-345-8112 (in Penn. 1-800-662-2444). 7E

*Add sales tax if for Calif., Ill., Mich., or N.Y. Allow 8 weeks for delivery. Add US$10 for shipping to Canada. IBM is a registered trademark of International Business Machines Corporation. Apple is a registered trademark of Apple Computer, Inc. Videx is a registered trademark of Videx, Inc. Macintosh™ is a trademark of McIntosh Laboratory, Inc. and is used by Apple Computer, Inc. with its express permission.

Your subscription includes:

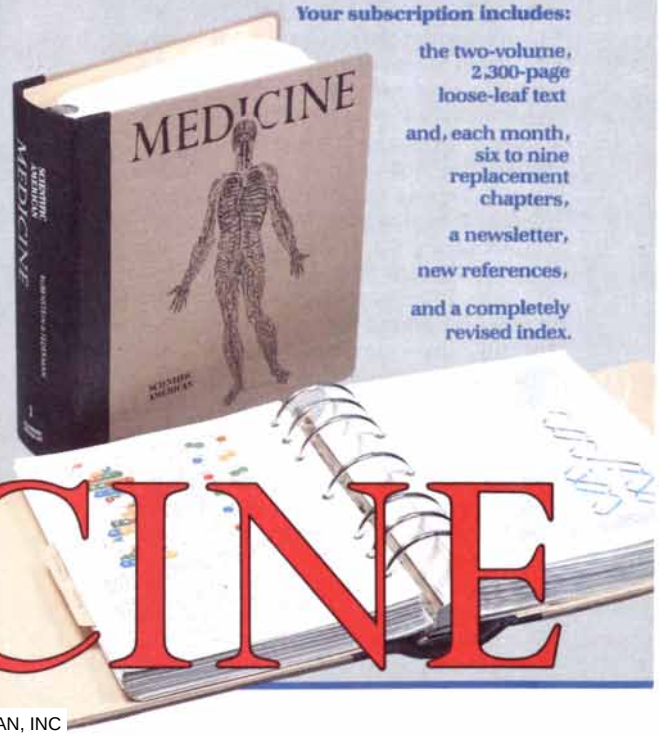
the two-volume, 2,300-page loose-leaf text

and, each month, six to nine replacement chapters,

a newsletter,

new references,

and a completely revised index.



SCIENTIFIC AMERICAN 415 MADISON AVENUE NEW YORK, N.Y. 10017

MEDICINE

Thermal Niches of Striped Bass

It appears that the preferred water temperature changes with the age of the fish. Overcrowding in optimal zones may be the reason that certain populations fail to thrive

by Charles C. Coutant

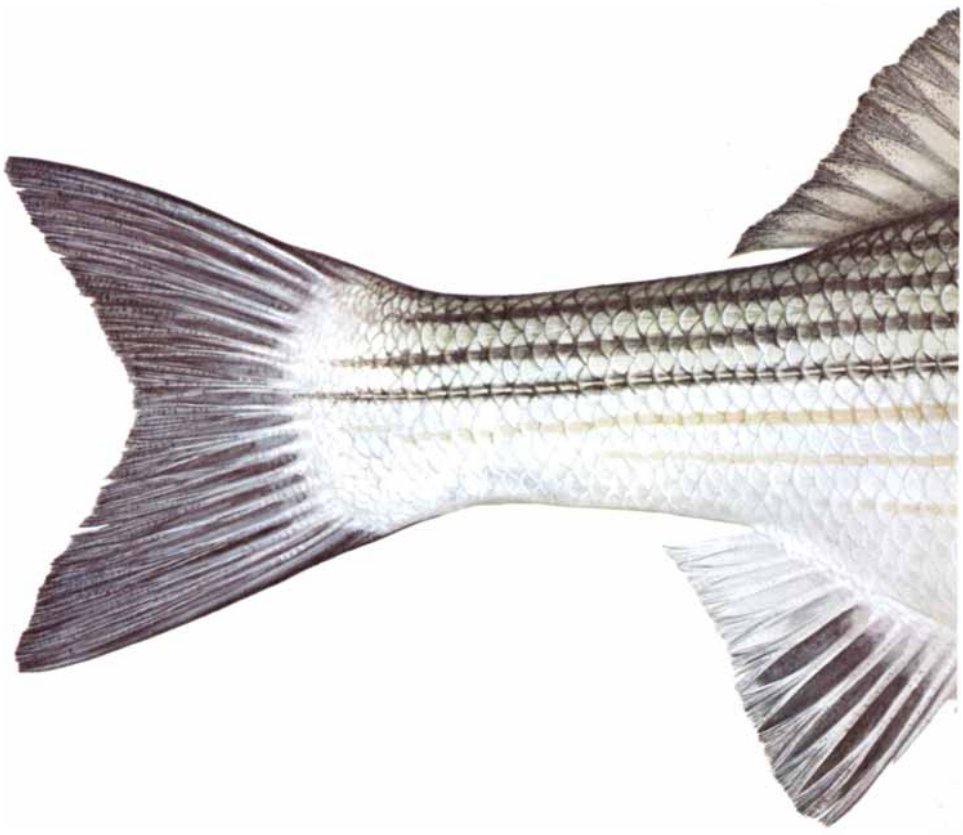
The striped bass (*Morone saxatilis*), prized by commercial and sports fishermen alike, is by nature a coastal fish that migrates up rivers to spawn. Some 30 years ago it was discovered that the fish can thrive in fresh water, and so it has now been stocked in many inland rivers and reservoirs.

Today the striped bass is an object of concern because the coastal populations have been declining and even some of the mostly successful freshwater populations have displayed such problems as excessive deaths in midsummer among the largest adults and failure of some stocked fish to attain the expected size. Examining these problems as part of my ecological work at the Oak Ridge National Laboratory, I have found that the water temperature preferred by the striped bass changes with the age of the fish. At each major stage of its life an individual striped bass occupies a different thermal niche. Presumably some of these thermal niches can become overcrowded occasionally, either because there are too many fish or because the volume of water at the proper temperature is insufficient to support the population. It may turn out that the problems affecting the striped bass can be reduced by applying this information in managing the coastal and freshwater populations.

Such a strategy would supplant the tendency to look for such local causes of difficulty as toxic pollution, spawning barriers, predation and overfishing. A particular focus of controversy, and therefore a stimulus for much research on the striped bass, has been the power-plant cooling systems that discharge warm water at places (such as in the Hudson River) frequented by striped bass.

Now that more is known about the fish, attention is turning to common features seen in striped bass populations throughout the U.S. One can now view the ecology of striped bass as a mosaic of habitats and resources apportioned among age groups. The basis for the view is the demonstrably different environmental requirements of adults and juveniles. In particular the requirement of adults for cool water seems to explain many otherwise paradoxical features of success and failure in striped bass populations.

Male striped bass usually mature at the age of two years, when they are from 30 to 35 centimeters (12 to 14 inches) long and weigh from .3 to .6 kilogram (.6 pound to 1.3 pounds).



STRIPED BASS (*Morone saxatilis*) is normally anadromous, meaning that it lives in coastal waters and migrates up rivers to spawn. In the 1950's it was discovered that the fish can

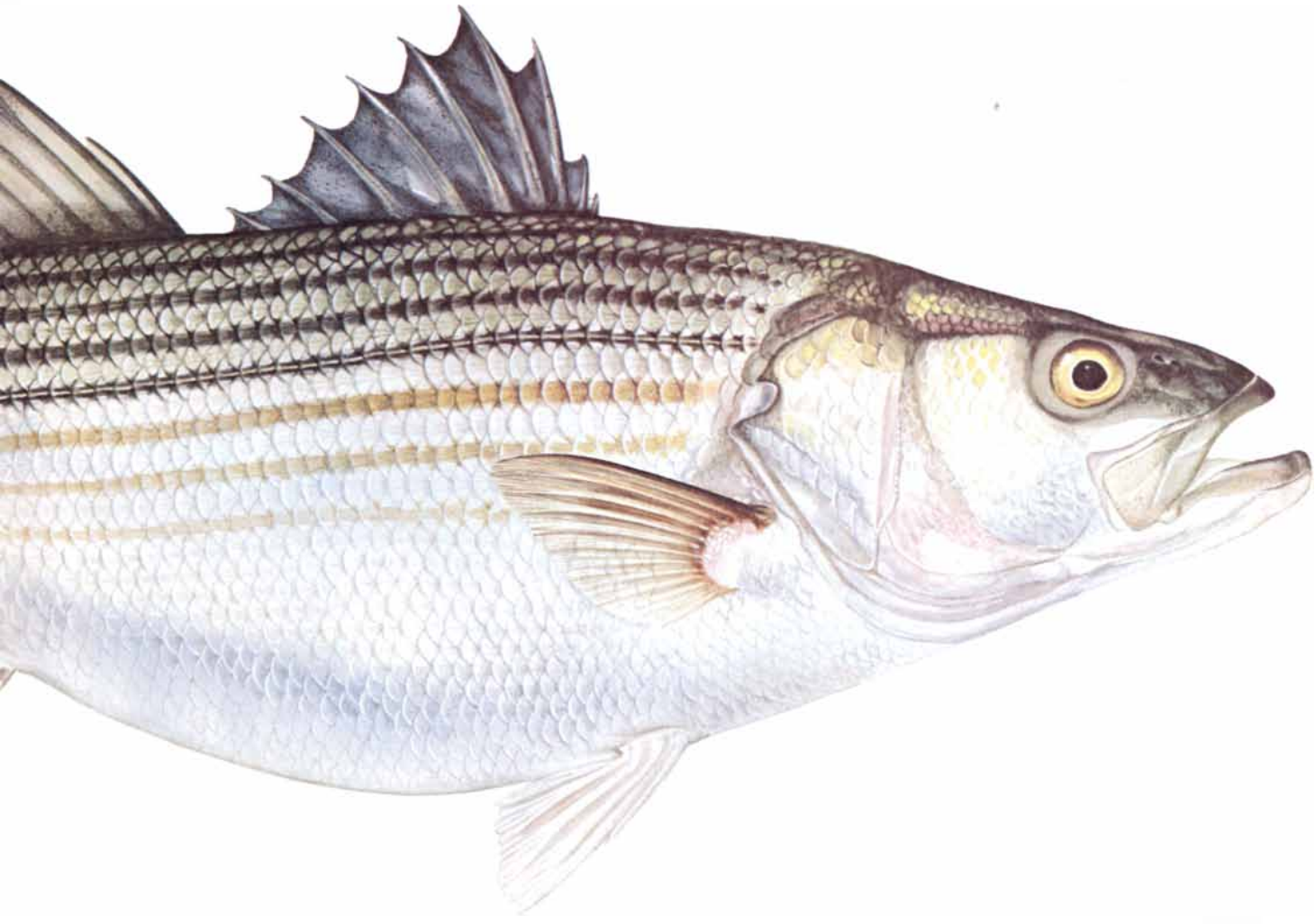
Females typically spawn for the first time at the age of four or five years, when they are from 45 to 55 centimeters long and weigh from 1.5 to two kilograms. Spawning takes place in the spring; individual fish spawn in several years (but not necessarily every year) of the normal 10-year life span. A spawning female releases about a million semibuoyant eggs. After three or four days, depending on the water temperature, the eggs hatch as nearly transparent larvae that continue to drift with the currents. A month later the larvae that survive the rigors of planktonic life grow into free-swimming juveniles that cluster in shallow areas along the shoreline. They usually remain in the brackish river mouth and coastal estuaries until they are two or three years old.

From Cape Hatteras to New England substantial numbers of adults (but not all) then leave the estuaries of their birth and migrate in schools along the coast, traveling generally north in summer and south in winter. South of Cape Hatteras, in the Gulf of Mexico, in the St. Lawrence River and on the West Coast striped bass rarely make coastal migrations; they remain closely associated with tidal rivers. Striped bass in freshwater reservoirs also display seasonal movements. Adults spend the winter in a deep part of a bay, away from currents.

Even this brief synopsis indicates the major differences in the habitats occupied by striped bass at different ages. The shifts of habitat are evidently not simple responses to local conditions but instead reflect age-specific genetic adaptations by the species. A useful hypothesis in guiding research has been that the life-cycle pattern developed through natural selection in native estuaries of the East Coast and fixed the environmental requirements of the species. These requirements would be manifested in any environment to which striped bass were transplanted, although further natural selection may have diversified regional stocks somewhat. Because many of the new environments do not fully reproduce the conditions of the native estuaries, they provide a means of obtaining more detailed knowledge about the habitat requirements of the species.

The new environments, particularly freshwater reservoirs in eastern Tennessee, have been the focus of my research on the thermal niches of striped bass. Fishes generally cannot control their body temperature, and so they function best when they can find water at a temperature that enables them to operate at a physiological maximum. Fish of a given species or at a particular stage of life will often congregate in water that is at the preferred temperature.

Juvenile striped bass (between one month and one year old) are small enough to be studied easily. Early research by several investigators supported the idea that the juveniles perform best at fairly high temperatures, meaning from 24 to 29 degrees Celsius (75 to 84 degrees Fahrenheit). My col-



thrive in fresh water. Since then many inland rivers and reservoirs have been stocked with striped bass. The coastal populations have been declining for several years and unusual numbers of the largest adults in freshwater populations have been dying in midsummer.

league David K. Cox and I did a detailed study of the growth rates of juvenile striped bass at different temperatures and feeding rates and found that growth was maximal at temperatures of between 24 and 26 degrees C.

In contrast, striped bass during the first month of life do best at appreciably lower temperatures, near the spawning and larval-development temperature of 20 degrees C. or less. The preferred temperature tends to rise as the fish grow, probably reflecting an adaptation to the typical springtime increase in water temperature.

It was much harder to get good data on the thermal niches of larger striped bass. Although records obtained in the 1930's by Daniel Merriman of Yale University on striped bass along the Atlantic coast indicated that adults are seldom found in water warmer than 27 degrees C., the notion persisted that the data on juveniles were typical of all older ages. A few laboratory experiments that successfully handled large adults suggested a decline in performance at a water temperature of about 27 degrees and a tendency among the fish to avoid water that warm. These data were helpful, but they did not form a basis for confident judgments about the effect of warm-water discharges by power plants. Since one of my laboratory's tasks was to assess such discharges, I sought to clarify the importance of temperature for the larger striped bass.

These fish do not lend themselves to laboratory experiments, being as much as three feet long and weighing as much as 70 pounds. (The biggest striped bass ever caught by a sport fisherman weighed 78.5 pounds.) We therefore had to find a way to study them in their natural habitats. Our success is due largely to my colleague James M. Rochelle, who devised a temperature-sensing transmitter that can be attached to free-swimming fish. It tells us where a fish is and what the temperature of the water is there.

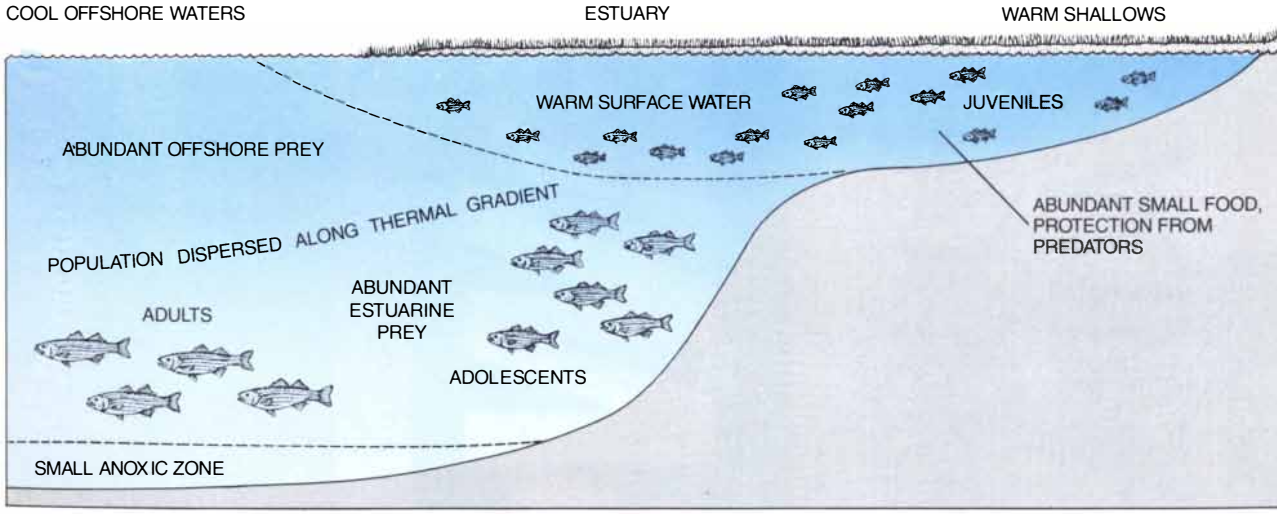
Our initial field studies with transmitter-equipped striped bass involved two-year-old subadults inhabiting two small lakes that had once been rock quarries. The lakes show a strong vertical stratification of temperature in the summer. Like most lakes, they get warm at the surface, but the depths stay cool. Our bass could select temperatures ranging from 10 to 30 degrees C. in midsummer. The lakes are well oxygenated, and so we could be certain the movements of the fish did not involve avoidance of low-oxygen areas.

The results that we (my colleague David S. Carroll, my student Larry Little and I) obtained over two spring and summer periods were striking. In early spring the fish tended to select the warmest water available 1.5 meters or more below the surface. As the surface water warmed above about 24 degrees C., tagged fish followed the zone of 20 to 24 degrees to increasing depths. Although they often made excursions to cooler water and occasionally went into warmer water, our statistical analysis showed clearly that this age group seeks a thermal niche centering near 22 degrees.

During the same years we were trying with little success to tag and monitor adults in an open reservoir. I had chosen Cherokee Reservoir, a large impoundment (87 kilometers long and 12,000 hectares in area) of the Holston River in Tennessee. My reasons were that a power plant warmed the inflow at the headwaters and that the people participating in the reservoir's sport fishery were dismayed because the larger striped bass had been dying mysteriously each summer since 1971, some seven years after a program of stocking the reservoir annually with juveniles had begun. The situation was ideal: our laboratory needed information about the effect of power-plant discharges on striped bass and the Tennessee Wildlife Resources Agency wanted to find out what was killing the large adults.

Our first problem was that we had trouble finding fish to tag in the summer. Notwithstanding our advanced university degrees, we did not solve the problem until we consulted a fishing guide who showed us the right places to find striped bass. Our next problem was that all but one of the many fish we tagged during our initial studies died within about two hours after we released them. In each case the fish made a number of rapid movements between the surface, where the temperature was above 25 degrees C., and middle depths at a temperature of 23 or 24 degrees. The one fish we tagged successfully in that series stayed in cool water (about 21 degrees) after its initial dive.

We did not solve the problem of this perplexing behavior until we realized that the one successful fish had been returned to the water near the capture



THERMAL GRADIENT separating striped bass by age is inferred from available data on populations in reservoirs, coastal waters and unpolluted estuaries in summer. Differences in the preferred temperature, ranging from warm (about 26 degrees Celsius) for juveniles to cool (about 20 degrees C.) for adult females, allow space and food resources to be partitioned in such a way as to achieve a maximal survival at all stages of life. Limitations on this selection of habitats by striped bass populations have arisen through increasing depletion of oxygen in deeper water and the concentration of toxic substances in zones of preferred water temperatures.

point, whereas the others had been released some distance away because the boat had drifted while we were tagging them. Looking into that phenomenon, we found that the capture area had cool water because a stream from a spring entered the reservoir there. The cool water sank, since it was denser than the warm water at the surface.

Tracking the one surviving fish for several days as it meandered up and down the cool, submerged stream channel, we found it sought temperatures that ranged from 18 to 25 degrees C. and centered at 22.5. In the main reservoir we mostly found temperatures of 25 degrees or more. There were a few other cool spots, some well known to the local fishing guides, but most of the deeper, cooler parts of the reservoir tended to be low in dissolved oxygen.

This pattern of midsummer residence in cool refuges in Cherokee Reservoir was documented in detail during studies in 1977 and 1978 by my graduate students Harold R. Waddle and Barbara A. Schaich. In each year the larger tagged fish, weighing up to 12 kilograms, chose cooler temperatures than the smaller ones.

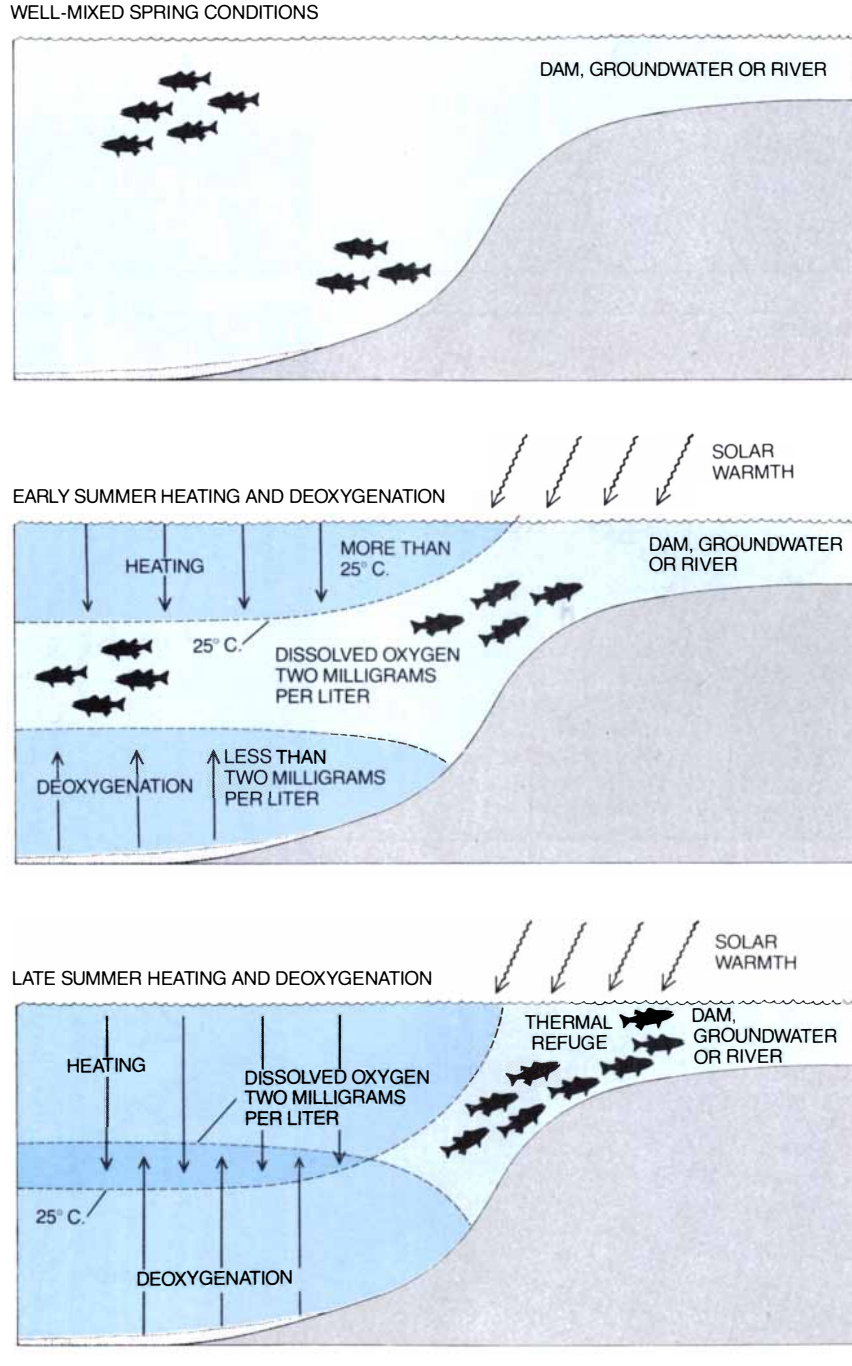
Our early failures in tagging appear in retrospect to have been valuable chronologies of a fatal squeeze affecting the fish that died. The temperatures were too high at the surface and there was not enough oxygen at the bottom. A tagged fish released outside its thermal refuge could survive for only a couple of hours. Its death was attributable to the alternation between water that was too hot and water that did not have enough oxygen, together with the stress on the fish from its frantic up and down movements in search of a suitable environment.

Did this finding have any bearing on the mysterious natural deaths of large striped bass every summer? The history of those deaths in Cherokee Reservoir and in a number of other reservoirs in the Southeast is consistent with a pattern of decreasing preferred temperature as the fish age and progressively fill the limited number of thermal refuges available to adults.

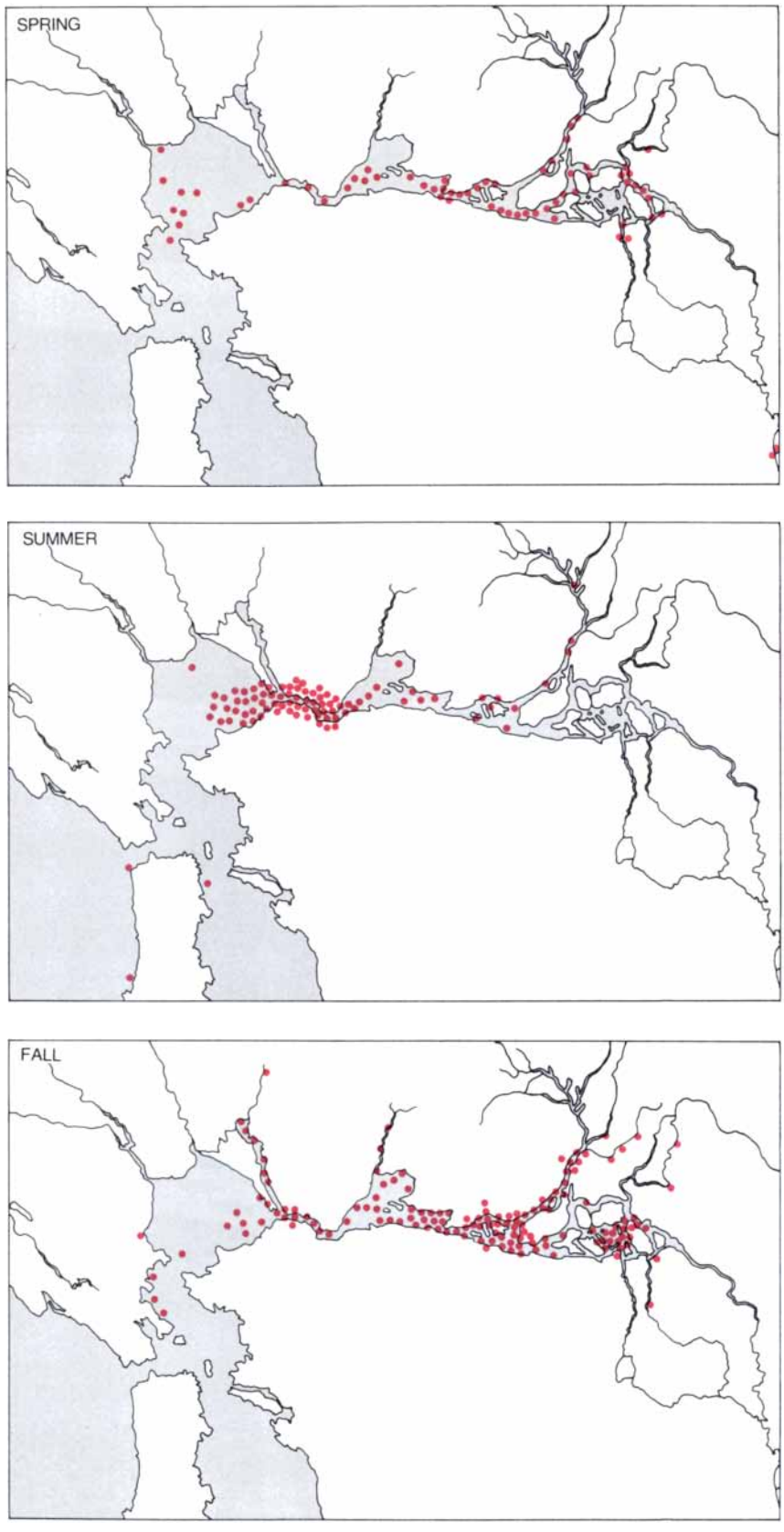
While we were doing our tagging experiments we saw a number of large (five kilograms or more) striped bass that had died. We autopsied many of them and found poor conditions that worsened as the summer went on. The conditions included empty stomachs (notwithstanding the abundance of small prey fish in surface waters), swollen and dark gallbladders (indicative of a low level of digestive activity) and numerous body sores.

We see these conditions as the results of crowding in the thermal refuges. Under those conditions the food supply dwindles rapidly. Furthermore, the close contact among the fish enhances the transmission of disease at the very time that resistance to infection is reduced by poor nutrition.

The requirement for cool temperatures must be quite important. Starving striped bass will not go into warm water two or three meters above them to pursue the abundant prey there. As they become weaker they may lose the ability to select appropriate tempera-



ADVERSE CONDITIONS for adult striped bass can develop in a freshwater reservoir in summer. The example here is Cherokee Reservoir in Tennessee. Because water near the surface gets warmer and water near the bottom becomes low in oxygen owing to the decomposition of organic matter, the habitats suitable for the fish shrink. A suitable habitat for an adult has a temperature below 25 degrees C. and at least two milligrams of dissolved oxygen per liter of water. Such cool water as there is comes from a dam, a river or a spring. It stays below the warm surface water, so that a thermal stratification develops. As juveniles mature and require cool water, they may saturate the space in thermal refuges, forcing larger adults into unsuitable niches, where they are physiologically stressed and may die.



SEASONAL MOVEMENTS of estuarine striped bass are consistent with a cool-water thermal niche for large adults. They are shown in the region of San Francisco Bay and the Sacramento–San Joaquin river delta. The colored circles represent tagged fish returned by fishermen from 1947 through 1952. The distribution is uniform in the spring; a tight concentration develops in the summer and shifts upstream in early fall. The groupings correspond to the movement of the zone where the water temperature is from 18 to 21 degrees C.

tures. They may also be forced out of refuges by competition from stronger individuals.

We made another discovery that is quite disconcerting: the stressful summer conditions in Cherokee Reservoir make females less able to reproduce. The Cherokee stock, compared with striped bass in two other reservoirs, showed a lower percentage of females capable of spawning, poorer fertilization of eggs, reduced success in hatching and poorer survival of larvae. The reason may be that an undernourished female cannot supply each egg with food reserves to carry the embryo and larva to the free-feeding stage.

Surprisingly, the heated water from the power plant in the headwaters of the reservoir had no discernible effect on the striped bass in summer. The human activity that did affect them was the hydroelectric plant at Cherokee Dam. The water withdrawn to generate power depleted the volume of cool water in the main reservoir, increased the volume of the upper layers being warmed by the sun and brought oxygen-depleted bottom water from the productive upstream regions of the reservoir down to the lower reaches. A challenge for ecologists is to improve their understanding of the physical and chemical processes that deplete oxygen and raise temperatures in reservoir habitats otherwise suitable for striped bass.

Does the age-specific behavior of the striped bass in Cherokee Reservoir represent a more widespread phenomenon? We investigated this question in two ways. We did a telemetry study of the striped bass in another reservoir that is physically quite different from Cherokee but offers the same thermal niches. The other approach was to review what has been written about the response of striped bass across North America to various conditions of temperature and dissolved oxygen.

The site of the telemetry study was Watts Bar Reservoir on the Tennessee River. It differs from Cherokee Reservoir in that it receives cool water discharged from an upstream dam. The cool water enters in a major tributary, the Clinch River, before it mixes with the warmer water of the main reservoir, and together they form a horizontal temperature gradient of considerable length. Terry E. Cheek, a graduate student, Michael J. Van Den Avyle of Tennessee Technological University and I found in a 16-month study that tagged adult striped bass showed the same pattern of temperature selection as the fish in Cherokee Reservoir. They left the warm, low-oxygen

water of the main reservoir for the cooler tributary water in summer and dispersed in the fall when the main reservoir cooled.

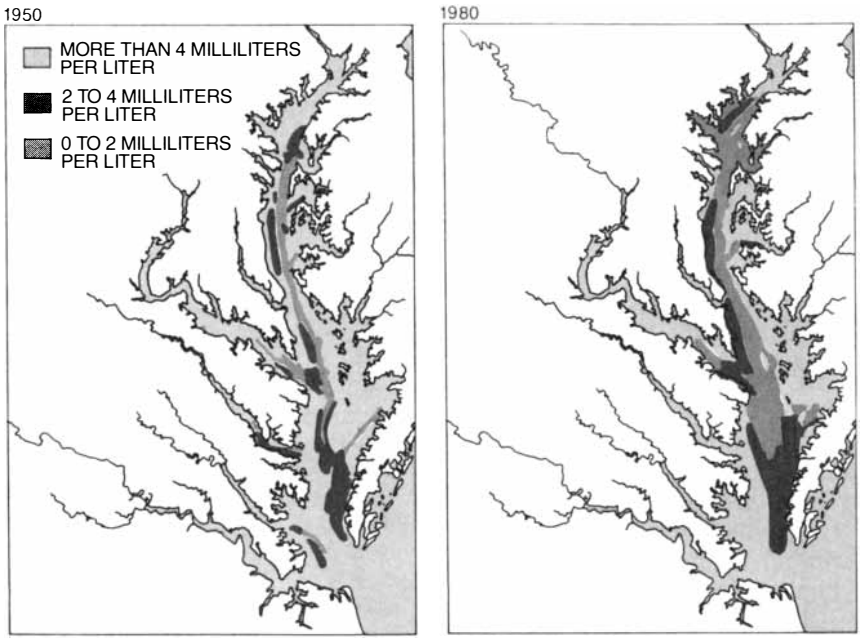
The written material, covering numerous studies in both fresh and salt water, also supports the notion that a temperature gradient from 18 to 25 degrees C. forms a thermal niche for adult striped bass. For example, the native striped bass in Florida, which remain in rivers, congregate at large springs (temperature near 21 degrees) when river temperatures rise above 26 degrees. In those cool refuges they have been seen crowded at high density, in poor condition and diseased. Movement of large striped bass toward cool waters discharged from dams seems to be a common phenomenon in the Southeast. The seasonally changing distribution of striped bass in estuaries of the East Coast and in San Francisco Bay is also in general agreement with the preferred temperatures seen in our reservoir studies.

Much other evidence can be cited in support of the hypothesis that the temperature preference of striped bass changes with the age of the fish. Growth rates of juveniles in wild populations are maximal at the relatively high temperatures of southern latitudes, whereas adults grow faster at the relatively low temperatures of northern latitudes. Striped bass introduced to reservoirs that have good thermal diversity and enough oxygen in summer grow well at all ages.

The evidence seems to be persuasive, and it gives perspective to a number of other observations. For example, the prominent and age-specific seasonal migrations of striped bass in estuaries and coastal waters now seem to be better explained by temperature preferences than they are by the vague notion of "chasing food" that had been advanced previously. The tendency for the largest adults to travel farthest into cooler coastal waters is consistent with a decreasing preferred temperature as fish age. In the East the annual migrations that put adult striped bass off southern New England in the summer and off the Carolinas in winter match the seasonal locations of water near 20 degrees C. It is also significant that striped bass along the coast of the Gulf of Mexico live in rivers rather than estuaries. The water of the gulf is uniformly warm, whereas the rivers have cool spots arising from springs.

Can the newly recognized importance of temperature for striped bass, particularly the critical need of adults for cool water, contribute to the management of populations in decline or distress? Our data apply most directly to reservoir stocks. Armed with the understanding that juveniles need warm water and adults need cool water with an adequate amount of oxygen, one can choose reservoirs for stocking on the basis of their annual cycle of physical and chemical conditions. Moreover, it is sometimes possible to manipulate the seasonal water temperature, say by releasing water from a dam. The supply of oxygen can be improved by controlling the oxygen-depleting materials that enter the water. Reduced rates of stocking and increased harvests, particularly of two-year-old fish, can prevent the saturation of limited summer refuge space in reservoirs that are warm and oxygen-poor.

It is less certain that the thermal-niche information can be applied to the management of coastal stocks. Substantial rewards can be expected if the information turns out to be useful. The Emergency Striped Bass Research Study, coordinated by the U.S. Fish and Wildlife Service and the National Marine Fisheries Service, estimated that the reduction of 90 percent in the commercial catch along the East Coast since 1973 had by 1980 cost 7,000 jobs and a loss in income of $220 million per year. Fishing moratoriums imposed by Maryland and Delaware in 1985 to stem the decline in population have increased the economic effects of the striped bass problem. If temperature is a key factor in the ecology of the species, looking for changes in the available thermal-niche space where populations are in trouble may help to explain the losses and to suggest remedial measures.



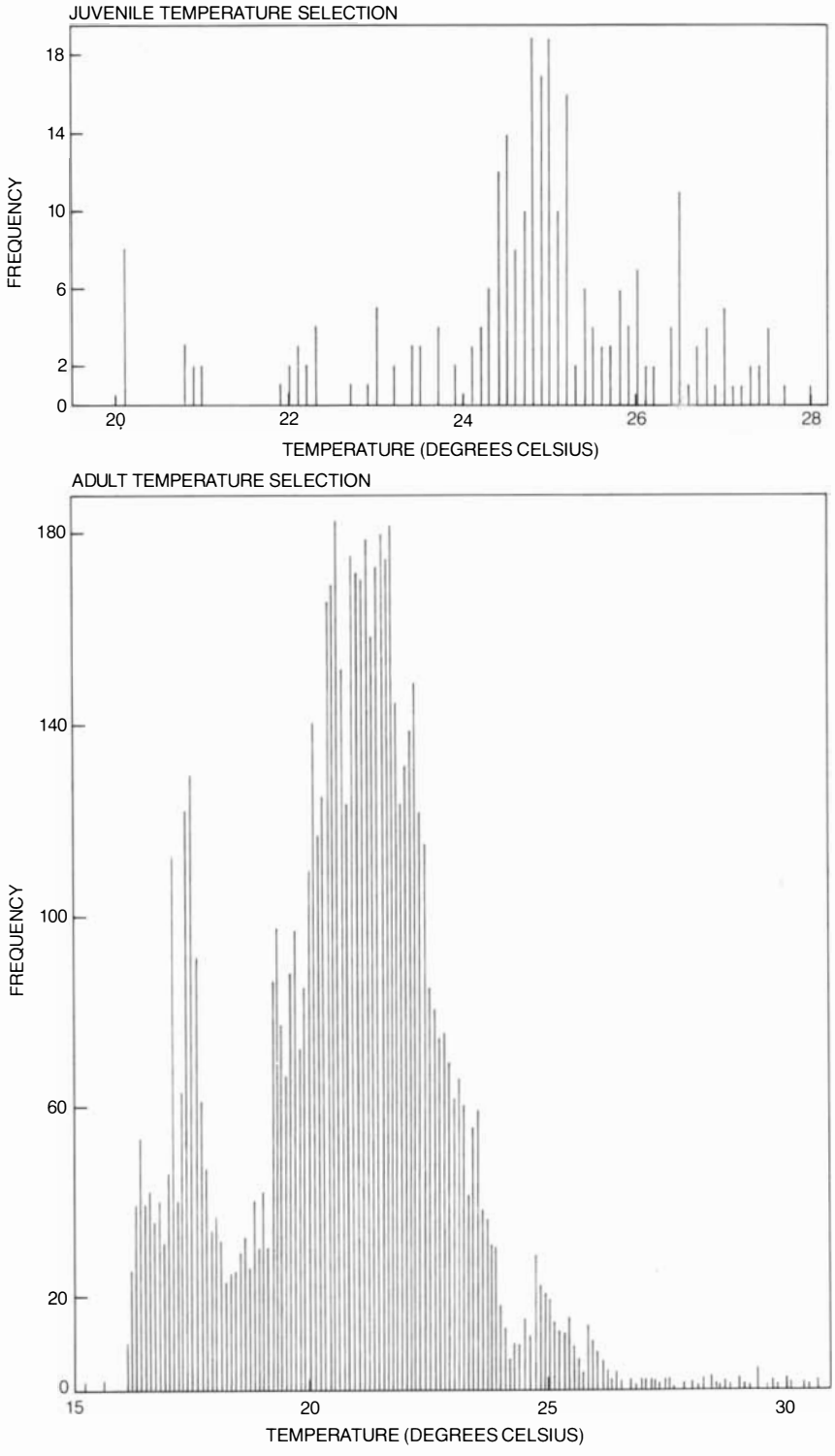
LOWERED OXYGEN in the cool waters of Chesapeake Bay may have contributed to the coinciding decline in striped bass populations. Since 1950 low-oxygen zones have expanded considerably. The data were assembled for the U.S. Environmental Protection Agency.

One possibility is that striped bass congregate in heavily polluted zones of estuaries because of the favorable temperatures there and so become exposed to high levels of toxicants. This pattern could explain the summer aggregations of striped bass in the San Francisco Bay area and the finding by Jeanette A. Whipple and Maxwell B. Eldridge of the National Marine Fisheries Service that those fish carry considerable toxic material in their bodies. Such areas could be given priority in pollution control.

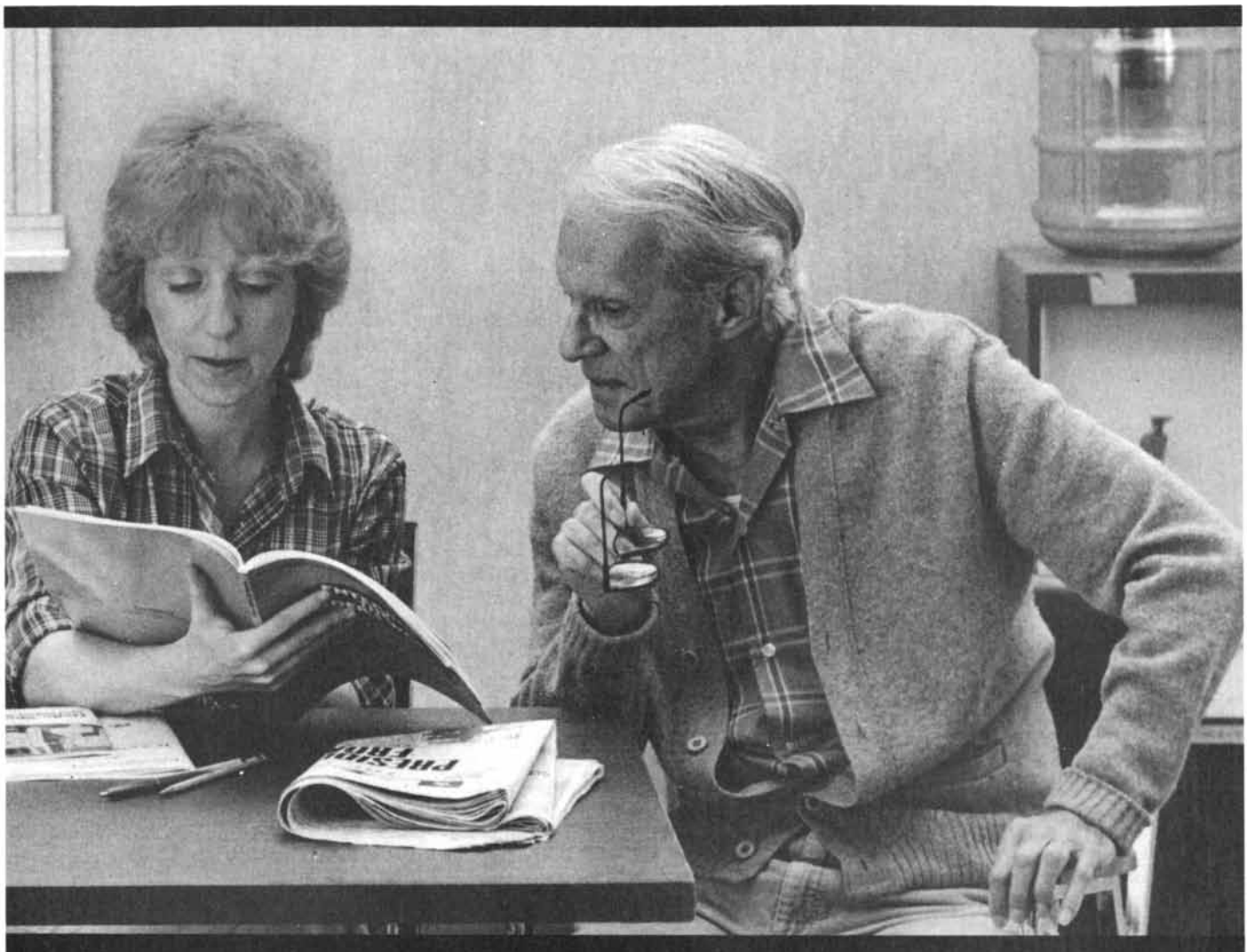
One of the most striking changes in the habitat of striped bass along the East Coast is the declining oxygen content of the cooler waters in Chesapeake Bay during the summer. The Chesapeake has long been the supplier of most of the young striped bass found later as adults in the bay and in coastal waters, and it is this stock that is declining most rapidly. A survey of the bay made in the summer of 1980 by Kent Price and his colleagues for the U.S. Environmental Protection Agency showed a temperature-oxygen pattern reminiscent of the worst conditions in Cherokee Reservoir; there was simply no water sampled that was suitable for large striped bass. Moreover, the volume of water containing little or no oxygen in summer has increased considerably since about 1970 and the annual onset of low-oxygen conditions now occurs earlier. The principal

causes appear to be the large growth in the human population of the area and a rising inflow of nutrients from agriculture.

Is this deteriorating temperature-oxygen condition important to striped bass that have access to the open ocean, as the Chesapeake Bay population does? Because one cannot be sure, prudence suggests that researchers and managers should examine thoroughly the possibility that the Chesapeake striped bass confront vanishing thermal niches.



WATER TEMPERATURE preferred by adult striped bass is lower than the temperature of the water most commonly frequented by juveniles. The data, which are typical for striped bass, represent the movements of several juveniles in a laboratory tank incorporating thermal gradients and of one tagged adult in Cherokee Reservoir during the summer of 1977.



We've got a job so important, it takes a retired person to do it.

After a lifetime of working, we want you to tackle one more job. Maybe the most important job you've ever done. We want you to help give new lives to 27 million Americans, the 27 million who are functionally illiterate.

Today, illiteracy is more than a problem. It's an epidemic that, believe it or not, has reached 1 out of 5 adults. It robs them of a chance for a decent living…it robs them of the chance for a decent life.

You can change that by joining your community in the fight against illiteracy. It takes no special qualifications. If you can read, you can tutor or help us in countless other ways. And because you're who you are…because you've got the patience and the experience in dealing with people, you can do it better than others.

To volunteer, call the Coalition for Literacy at toll-free **1-800-228-8813.** Beating illiteracy is a big job. But you may find it's the most rewarding job you've ever had.

Volunteer Against Illiteracy. The only degree you need is a degree of caring.

Ad Council | Coalition for Literacy

The Emergence of Maya Civilization

It was once thought to have been sudden. That view has recently been overturned: it was a gradual process. Civilization came into being well before the beginning of the Classic period in A.D. 300

by Norman Hammond

How does a great civilization arise? Does it emerge full-blown in a single episode of cultural transformation, or is it the final flowering of a long process that entails the patient accumulation of many innovations? These two possibilities are the extremes of a spectrum. Although one might expect the scholarly view of a particular culture to shift its position on the spectrum slightly from time to time, it seems unlikely that the view of any culture would move from one extreme to the other. Yet in the case of Maya civilization that is exactly what has happened. What is even more remarkable is that the transformation, which amounts to a revolution in our understanding of the origins of Maya civilization, has taken little more than a decade.

It has long been clear that the Maya, who occupied the Yucatán peninsula along with parts of Belize, Guatemala, El Salvador and Honduras, developed one of the most sophisticated of all native American cultures. At its height in the Classic period (from about A.D. 300 to 900) Maya society was highly stratified, with a ruler at the apex of six or seven clearly defined social classes. A complex cosmology held places for gods, natural forces and ancestors. An elaborate calendar provided the framework for ritual and historical events. The rituals were enacted at ceremonial centers that formed the core of great cities. This cultural superstructure rested on the cultivation of maize, and techniques had been developed for making steep hillsides and swamps cultivable.

A decade ago it was thought that Classic Maya civilization sprang into being quite suddenly during the third century A.D. The preceding period, called Preclassic or Formative, was believed to have been an age of humble village farming societies. Since 1975 many discoveries and reassessments of known evidence have radically changed the accepted picture of Preclassic culture. It has been found that intensive agriculture above the slash-and-burn level was developed far earlier than had been thought. Huge buildings that have been dated to Preclassic times imply substantial social and economic organization. Recent excavations show that standardized tools were made in large-scale workshops and traded over great distances. It even seems that some elements of the Classic intellectual culture were widespread in Formative times. As a result of such findings Preclassic culture no longer seems a lowly forerunner. On the contrary, it appears to have been (at least in its culmination) the first phase of the great Maya civilization.

One reason scholars have been able to accept the idea of advanced social development in the Preclassic period is the discovery that the ancestors of the Maya occupied their territory much earlier than was known until recently. In 1975 the earliest confirmed date of occupation rested on a radiocarbon date corresponding to a calendar date of about 900 B.C. (Before about 400 B.C. there is a discrepancy between radiocarbon dates and calendar dates. The radiocarbon dates are calibrated with the calendrical ones by means of bristlecone pine trees of known age. At 900 B.C. the discrepancy is about 200 years, and the uncalibrated date is indicated by lowercase letters: 700 b.c.) Since 1975 the span of occupation has been extended by a factor of five. It now seems likely that the forebears of the Maya and their descendants inhabited their territory continuously from around the end of the last ice age 10,000 years ago.

The pushing back of the date of occupation applies to all three of the environments in which the Maya lived: the volcanic highlands of Guatemala, the tropical forested lowlands of Belize and the low-lying Yucatán peninsula, with its scant rainfall and scrub vegetation. At Loltun Cave in the Puuc Hills of the Yucatán, Norberto Gonzalez Crespo and Ricardo Velasquez Valadez of Mexico's National Institute of Anthropology and History have uncovered a long sequence of cultural deposits. The oldest layers, at the bottom of the sequence, are said to contain the bones of animal species (such as the native horse) that became extinct soon after the end of the Pleistocene ice age. Unfortunately there is as yet only a single radiocarbon date for the Loltun sequence. That date, about 2300 B.C. (1800 b.c.), marks the end of the preceramic layers and the making of the first pottery.

In Belize several stone projectile points have been found that bear a strong resemblance to those used by big-game hunters of the Clovis culture, who roamed the Great Plains of the U.S. between 12,000 and 10,000 years ago. The best-known of the Belize finds is the Ladyville Point, which was discovered by a team from the University of Texas at San Antonio under the leadership of Thomas R. Hester. Richard S. MacNeish of Boston University has made a series of other finds that complement the Ladyville Point. MacNeish suggests that his finds can be arranged in a sequence extending from 9000 to 2500 B.C. In the highlands of Guatemala a site called Los Tapiales has yielded Clovis-type points in association with artifacts

providing radiocarbon dates of about 10,000 years ago. In the surrounding region of El Quiché a survey by Kenneth L. Brown of the University of Houston revealed many preceramic sites dating from before 2500 B.C.

The cumulative effect of the recent finds has been to demonstrate that the entire Maya area had been penetrated by human beings as long as 8,000 years before the beginning of the Classic period. Moreover, the likelihood is great that the early occupants were the direct ancestors of the Classic Maya population. The early date of occupation renders the notion of Preclassic civilization more palatable because it allows more time for social development. During the 8,000 years that preceded the emergence of the high Classic culture, the ancestors of the Maya had plenty of time to accumulate the economic and technical underpinnings of civilization.

One of the most critical of these underpinnings is agriculture, a subject on which the recent findings have had a

NINE JADE HEADS make up a pair of ritual offerings found at two Maya sites. The four heads in the upper row were found by the author at Nohmul in Belize. The four in the lower row and the one at the bottom were found 28 kilometers away at Cerros by David A. Freidel of Southern Methodist University. The author believes both rows include the same set of deities: two portraits of the sun god (*left and second from left*), one of the maize god (*second from right*) and one of a god of darkness (*right*). The two sets of heads were carved between 100 B.C. and A.D. 100, during the Late Preclassic period. The correspondence between portraits from different sites suggests that a shared intellectual culture developed long before the beginning of the Classic period in the third century A.D., which was formerly thought to mark the beginning of Maya civilization. (The head at the bottom may portray the duality of life and death, a common theme in subsequent pre-Columbian art. Its left side, which is cut away, may depict a skull and its right side a living face.)

potent effect. A decade ago it was thought that the inception of settled farming communities took place in about 900 B.C. It is now known that the actual date lies between 2500 and 1250 B.C. (2000 and 1000 b.c.). At Cuello in northern Belize, where I directed a National Geographic Society expedition from 1978 to 1980, radiocarbon dates suggest that the first sedentary agricultural settlement was established in about 2500 B.C. (2000 b.c.) [see "The Earliest Maya," by Norman Hammond; SCIENTIFIC AMERICAN, March, 1977]. The radiocarbon chronology from Cuello has aroused controversy, but even conservative critics do not put the beginning of the cultural sequence later than about 1500 B.C. (1200 b.c.).

The early corn farmers at Cuello form part of the history of the Early Preclassic phase, which lasted until about 1250 B.C. Maize farming was not the only cultural innovation that the Early Preclassic villagers at Cuello possessed. They constructed plaster-surfaced house platforms around patios and built wood-framed thatched houses on them. They made competent pottery in a variety of colors and shapes. In the succeeding period, the Middle Preclassic, which lasted from 1250 to 450 B.C. (400 b.c.) the cultural repertoire of the villagers was augmented by the expansion of interregional trading networks. The trading networks brought obsidian from sources near Guatemala City and jade from central Guatemala north to communities in the lowlands.

In spite of the growth of trade, Maya society at the beginning of the Middle Preclassic period retained a village farming character that largely corresponds with the previously accepted picture of life in the Preclassic period. Later in the Middle Preclassic, however, crucial changes took place that began to lift Maya society above the plane of the agricultural village. Those changes, which were concentrated in the three centuries preceding 450 B.C., are not yet well understood. Indeed, considerable difficulties stand in the way of understanding them, since the artifacts needed for full comprehension are in many instances buried under tons of material from later construction. Yet what happened near the end of the Middle Preclassic poses some of the most significant questions in New World archaeology, because by the beginning of the Late Preclassic in 450 B.C. a society quite different from that of the village farmers had emerged. Indeed, during the six or seven centuries of the Late Preclassic period true civilization appeared.

It is possible to find evidence of the transformation even at a site as small as Cuello, which in Late Preclassic times was a large village of about 1,000 residents. The ceremonial core of the Middle Preclassic community at Cuello was a small courtyard surrounded by wood buildings on platforms. In about 450 B.C. the courtyard was converted into a broad open platform capable of holding a large audience (or congregation); at the west end of the platform a small pyramid was erected. During the 1980 digging season we found that the conversion had been accomplished by filling the courtyard with rubble to the level of the surrounding platforms. The wood temples were burned and the facades of their substructures were ripped apart in a deconsecration ritual.

At the center of the courtyard a saucer-shaped depression was left in the rubble, and in it were put some two dozen butchered human bodies. Frank and Julie M. Saul of the Medical College of Ohio at Toledo have analyzed the remains and found that all the victims were young; it is likely that all

MAYA TERRITORY includes parts of Mexico, Guatemala, Belize, western Honduras and El Salvador. The area includes a wide range of environments, from the volcanic highlands of southern Guatemala, through the tropical forested lowlands of northern Guatemala and Belize to the flat, low-lying peninsula of Yucatán with its meager rainfall. Many of the sites that have contributed to the recent reevaluation of Maya prehistory are in the tropical forests of the northern lowlands. Maya-speaking people are found throughout this area today, occupying the same range as the civilization of their forebears, although the once populous northern Guatemala heartland is occupied only by farmers and small towns.

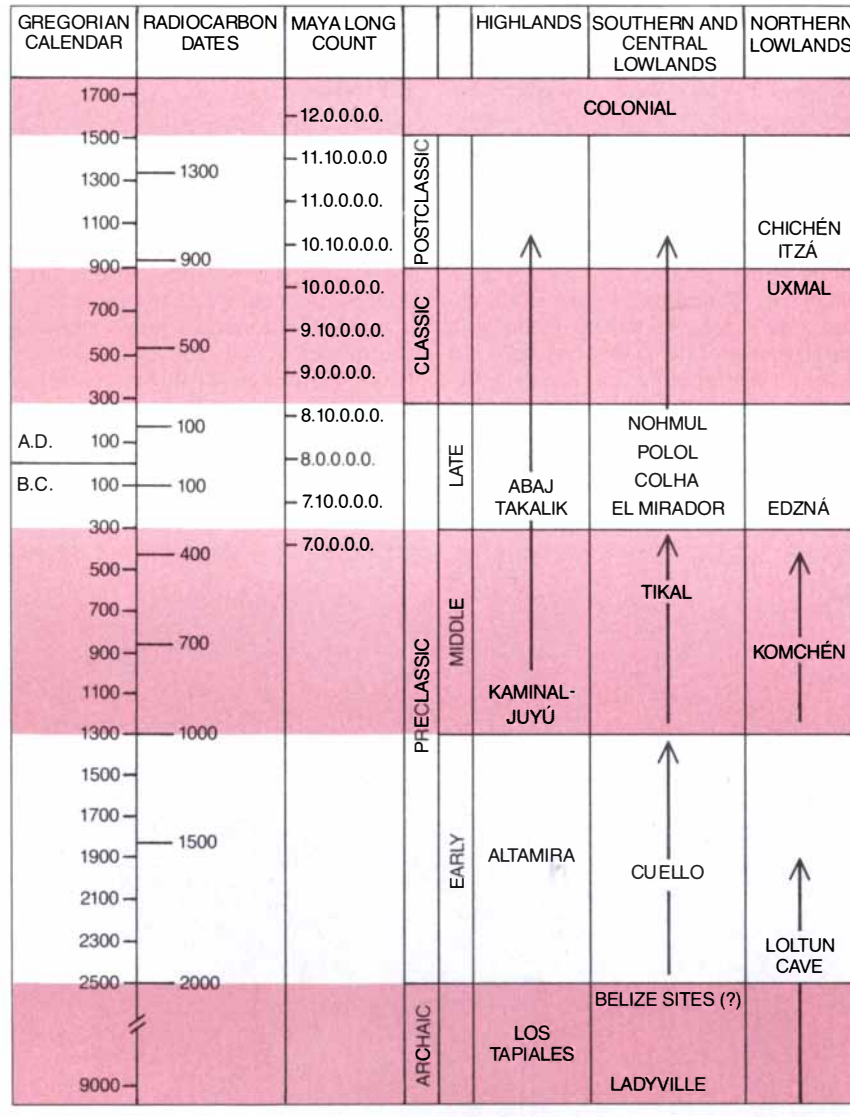
were male. Accompanying the hacked bodies were pottery vessels and half a dozen carved bone tubes. The tubes may have been handles for bloodletting implements like those employed in rituals of the Classic period; they may also have been handles for feather fans. More important than their precise function, however, is the fact that five of them bear a carved design that is clearly identifiable as the Maya *pop* (a woven mat). The *pop* was used by Maya rulers and was equivalent to a throne or a crown in modern iconography as an emblem of royal authority.

The discovery of the carved tubes suggests that by 450 B.C. there were rulers who had established both the reality of royal power and its symbolic expression. None of these rulers, however, could be interpreted as having held sway over more than a limited area. Even during the florescence of the Classic period Maya society was not unified under a central authority. Instead its political structure resembled that of Renaissance Italy: a patchwork of independent polities varying greatly in territory, riches and military power. The findings from Cuello are too fragmentary to demonstrate conclusively that the basis of this political patchwork had been established by 450 B.C. Yet a growing number of findings suggest that midway through the Late Preclassic (or not long before the birth of Christ) real political power had been forged.

One interesting piece of evidence is that the Maya population grew considerably during the Late Preclassic phase. There are many more Late Preclassic sites than there are Middle Preclassic sites, and the later ones tend to be larger and more densely packed in the landscape. Our survey of northern Belize, carried out in 1973 and 1974 under the auspices of the British Museum and the Corozal Project of the University of Cambridge, revealed that after 450 B.C. there were four times as many sites as there had been in the region before then. Of course population growth does not in itself imply the development of political power. The rapid demographic growth of the Late Preclassic, however, was accompanied by an increase in the scale of economic projects. The scope of those projects suggests that they were directed by rulers who possessed both considerable administrative skill and substantial authority.

One area where authority does seem to have been present before Classic times is agriculture. It has been known for some time that during the Classic period slash-and-burn techniques were supplemented by more intensive forms of cultivation entailing modification of the land. For example, hillsides were modeled into terraces. The terraces prevented erosion by trapping soil as it washed downhill and made otherwise uncultivable areas productive. Large areas were cleared and terraced in one operation, which suggests that the work was directed by a single source of authority. Once the land was divided into terraced fields it may have been allocated to individuals to farm. A similar level of organization and authority is implicit in the draining of swamps. In the wetlands of river valleys and in slow-draining basins such as Bajo de Morocoy on the Yucatán or Pulltrouser Swamp in northern Belize canals were cut to create drained or raised fields, transforming the damp margins into land that could be farmed seasonally or even the year round.

Recent work suggests that similar techniques were developed well before the beginning of the Classic period. The earliest example of drained fields



MAYA CHRONOLOGY can be described in several ways, as the chart shows. The first column displays dates in the Gregorian system. The second column displays uncalibrated radiocarbon dates, which differ from calendar dates by an amount that increases with the interval from 100 B.C. The third column displays dates in the Maya system. The Maya calendar was a linear progression divided into 400-year cycles called *baktunob;* it began in 3114 B.C. with the most recent of several creations of the universe. The fourth, fifth and sixth columns show the periods into which scholars have divided the development of Maya culture in the three main zones the Maya occupied. Recent finds have pushed the earliest date of human occupation in the Maya area from 900 to 9000 B.C. and shown that many of the crucial elements of Maya civilization were formulated in the Late Preclassic period.

comes from Cerros, a site on the coast of Belize. The ceremonial core of the Preclassic site at Cerros lies on a small peninsula, and it was separated from the mainland by a canal 1,200 meters long. Vernon L. Scarborough, working at Southern Methodist University, showed that the canal was dug sometime between 200 and 50 B.C. The canal was six meters wide at its broadest point and more than two meters deep in places. Running into the main canal was a feeder channel draining a small group of fields that had been cut from a scrub-filled depression. Because there are only a few drained fields at Cerros, the contribution they made to the local economy cannot have been great. Yet their existence demonstrates that the development of technology for diverting and channeling water supplies was well under way in Preclassic times.

That conclusion is supported by the discovery at Edzná, a city in the coastal plain of Campeche, of waterworks that are several orders of magnitude larger than those at Cerros. The main structure of the waterworks, which date from the Late Preclassic, is a canal 12 kilometers long extending south from the city toward the nearby Rio Champotón. At the north end of the canal is a moated complex that was probably a ceremonial precinct but was originally called a fortress. The moat, still filled today, is as much as 100 meters across; most of the 253,000 cubic meters of soil and limestone taken out of it was used in the construction of the ceremonial complex, which took place between 200 B.C. and A.D. 100. Fanning out from the north side of the city center are seven shorter canals that fed reservoirs and provided canoe transport. The scale of construction at Edzná is staggering. Raymond T. Matheny of Brigham Young University has estimated that altogether the waterworks may have provided 2.25 million cubic meters of water storage and taken 1.7 million man-days of labor to construct.

Projects such as those at Edzná and Cerros show that Preclassic Maya communities had an agricultural econ-

DRAINED FIELDS were created by the Maya in swampy areas that would otherwise have been uncultivable. The fields in the photograph are on the edge of Nohmul, a large Late Preclassic city near the Belize-Mexico border. The water was drained into the small channels between the fields and then into larger canals, which are at the left of the grid of fields. Recent excavations have shown that during the Late Preclassic period the Maya developed techniques for draining fields and terracing hillsides that raised agriculture well above the slash-and-burn level. The author is currently directing an excavation at Nohmul sponsored by the National Geographic Society and the British Museum; further work there may show that the fields in the photograph are of Late Preclassic origin.

omy based on a high degree of collective organization. Findings from Colha, a site south of Cerros, show there was also an organized manufacturing economy supplying a trading network for regional distribution. Colha lies in the coastal-swamp belt of eastern Belize on a massive seam of chert nodules. (Chert is a stone much like flint). Initial excavations from 1973 to 1975 by the Corozal Project, of which I was the director, proved that Colha was an important site in Preclassic times. Further work since 1979 by Hester and Harry J. Shafer of Texas A&M University has demonstrated the scale of the manufacturing activities carried out there.

Within the six square kilometers of the site nearly 1,000 Maya mounds have been identified by the two projects. Many of the mounds were platforms for dwellings, but Hester and Shafer have identified 89 as chert workshops and shown that at least 32 of the workshops were functioning during the Late Preclassic period. The remains found at Colha indicate that a limited range of tools were made there. The commonest were large oval axes and hoes, which served as general-purpose tools for clearing brush and working the land. Another product of the workshops was a type of adze made by the "tranchet" technique, in which the sharp edge was formed by detaching a final curved flake from the bit end of the tool.

The curved flakes, which have been dubbed "orange peels" because they resemble bits of discarded rind, make up a significant fraction of the waste left at the site of the workshops. The thick beds of waste flakes are 1.5 meters deep in some places, and they contain nearly five million flakes per cubic meter. Such copious debris provides a means of gauging the rate of production. For example, by sampling the orange peels lodged among the other stone flakes Hester and Shafer have calculated that the workshops at Colha produced a total of more than two million tranchet adzes.

Even if the factory had functioned continuously during the 650 years of the Late Preclassic period, the annual output of tranchet adzes could not have been less than about 3,000. To that sum must be added an even larger number of oval axes and other tools, which are more difficult to count because their manufacture did not leave a single striking remnant such as the orange peel. The total is clearly too large to have been used exclusively by the local population, and Colha products were indeed widely distributed.

"LOST WORLD" PYRAMID at Tikal in Guatemala suggests the scale of construction of which Preclassic Maya societies were capable. The pyramid is more than 30 meters tall and has only recently been excavated. Built in the second half of the Late Preclassic period, it is surrounded by other slightly later buildings that form a ceremonial center. When Tikal was first explored by archaeologists, it was assumed that these structures were built in about A.D. 700. Excavation has shown they were constructed at least 400 years earlier.

They have been found throughout northern Belize and as far away as El Mirador and Uaxactún in Guatemala. In spite of the fact that many Maya communities had chert-tool workshops, standardized implements from Colha captured much of the market within a radius of 40 kilometers and were in demand more than 160 kilometers away.

The findings from manufacture and agriculture leave the impression of a society on the verge of civilization. That impression is strengthened by considering the Late Preclassic ceremonial centers. During the Early and Middle Preclassic periods ceremonial complexes were relatively small units standing at the center of farming communities. By the Late Preclassic they had grown much larger and stood at the center of towns, foreshadowing the spatial and demographic organization of Classic society.

The site where such developments are evident earliest is Komchén in the northern Yucatán, which E. Wyllys Andrews V of Tulane University has estimated was a densely nucleated community of between 3,000 and 5,000 people. In the center of Komchén five large platforms surround a plaza some 80 by 150 meters. Two of the platforms are long, narrow, tall structures reached by two stairways apiece; the others are broad and lower. A raised causeway links two of the platforms. The entire complex was erected in the first century of the Late Preclassic, between 400 and 300 B.C.

As the social impulse behind the construction of such monuments became stronger, structures were built that dwarfed the ones at Komchén. At Nohmul our recent expedition, sponsored by the National Geographic Society, has uncovered a Late Preclassic center of the third century A.D. The central plaza there, 130 meters across, was flanked on the east and west by large, low mounds that probably supported wood buildings. On the north side was the "acropolis": a platform more than 6,000 square meters in area rising 10 meters above the natural soil level. That platform, built in one continuous operation of limestone blocks quarried a few hundred meters from the building site, contained as much construction material as the entire group of mounds at Komchén. The scale of social organization required in the construction of Nohmul is suggested by the fact that as much as 100,000 cubic meters of limestone went into

STONE AX found in the Rio Hondo near San Antonio on the Belize-Mexico border retains its original handle of *Annona* (custard apple) wood. Blackened by 2,000 years of immersion in river water, the ax may have been used for cutting brush and hoeing fields; microscopic examination reveals none of the marks that result from felling trees. The oval axhead is about 20 centimeters long. (Each division of the measuring bar is one centimeter.) The axhead was made at Colha in Belize, about 40 kilometers from San Antonio. In late Preclassic times Colha was the site of a factory that turned out thousands of stone tools every year. The ax was found by the late Dennis E. Puleston of the University of Minnesota.

the plaza and the adjacent platforms.

Just as the spatial arrangement and scale of construction at sites such as Komchén and Nohmul suggest transitional forms, so some of the most interesting architecture there foreshadows Classic designs. On top of the acropolis at Nohmul we found the postholes that formed the foundation of a long, rectangular hall built of timber. Across its width were three aisles spanning more than seven meters; down its length were at least eight bays covering more than 23 meters. Such a structure was far too large to have been a simple residence, and it lacked the typical plan of a Maya temple. The most plausible interpretation is that the long hall was a predecessor in wood of the stone "palaces" of the Classic period. The palaces, found at many major sites, are thought to have housed the rulers of Maya cities along with members of their administration; they comprised many house-size modules built of masonry and set end to end in an overall design not unlike that of the building on the acropolis.

Similarly, the pyramids that form a significant element of Classic period architecture began to assume monumental proportions during the Late Preclassic. At Nohmul there is a second concentration of buildings northwest of the acropolis, including an isolated temple pyramid some 17 meters high that we think dates from the Late Preclassic. Other Late Preclassic sites include much larger pyramids. At Lamanai, some 60 kilometers south of Nohmul, stands the massive Structure N10–43. That pyramid is more than 30 meters high and is surmounted by a cluster of three small temples; David Pendergast of the Royal Ontario Museum in Toronto has shown that the entire edifice is of Late Preclassic construction. The famous "Lost World" pyramid at Tikal, of comparable size, was built at about the same time, as were the deep trench and rampart around the center of Becan, which have a circumference of about 1,900 meters and were apparently defensive in function. These structures are so imposing that when they were discovered it was immediately assumed they belong to the peak of the Classic period; only careful excavation has shown that they were built from four to eight centuries earlier.

The degree of economic and social organization implicit in the monumental works at Tikal, Lamanai, Nohmul and other sites is impressive. Yet what most people instinctively associate with the word civilization is not the organized mobilization of manpower

BONE TUBE carved with a design that may symbolize royal authority was found at Cuello in Belize. The bone comes from an unidentified large mammal, perhaps a jaguar. The carved design represents a *pop*, the woven mat used by Maya rulers. The tube, which may have been the handle of an instrument for ritual bloodletting, was found in a shallow pit among the bones of two dozen young men. They were apparently sacrificed when the ceremonial center at Cuello was rebuilt in about 400 B.C. The *pop* motif suggests that by the beginning of the Late Preclassic period Maya rulers had established the reality and the symbols of royal command.

but intellectual and cultural accomplishments of a high order. One site that offers evidence of both forms of social advancement is El Mirador, which lies in the remote country near the border between Mexico and Guatemala. El Mirador was discovered by Europeans from the air in the 1920's. In 1962 it was explored and mapped by Ian J. A. Graham of Harvard University. Because of the style of construction and the lack of monuments with inscribed calendar dates, Graham concluded that El Mirador was built before the beginning of the Classic period.

Recent work by a team from the New World Archaeological Foundation under the direction of Matheny has revealed how large El Mirador is and has uncovered some intriguing examples of Preclassic intellectual culture. The heart of the site is an area of dense construction about one kilometer by 800 meters. Within the core are two pyramids called El Tigre and Monos that are among the largest buildings ever raised by the Maya. Constructed in about 150 B.C., each pyramid holds more than a quarter of a million cubic meters of construction fill. At 55 meters, El Tigre is as tall as an 18-story building. Clustered on its top are three temples in a triangular arrangement. The triangular form, which is repeated on a larger scale by two pyramids that flank El Tigre, seems to be a characteristic motif of Late Preclassic ritual architecture.

The pyramid flanking El Tigre to the south, known as Structure 34, has been excavated, and the results confirm that some elements of the Classic period cosmology had emerged in a clear form by the Late Preclassic. On Structure 34 Matheny's team found several modeled stucco masks, each depicting a jaguar more than two meters tall. Some parts of the mask are black and the background is cream-colored, but the claws and teeth are painted red.

The pairing of red and black is a common feature of Classic Maya art that first appears in the Late Preclassic. The duality has a ritual significance, because red is associated with East and black with West; the daily transit of the sun from one to the other is the dominant axis of Maya cosmography. It is my thesis that red and black are conscious transformations of each other, derived from observations by the Maya that pottery vessels fired red in an oxidizing atmosphere turn black when they are refired under reducing conditions and turn red again on reoxidation. Late Preclassic pottery, which frequently has controlled areas of red and black, shows that Maya craftsmen understood the phenomenon beginning in the fourth century B.C.

Two kilometers east of El Tigre is the largest complex of buildings at El Mirador; indeed, it is perhaps the largest in Maya history. The complex is known as Danta. It consists of two terraces, each more than 300 meters wide, carved out of a hillside like two gigantic steps. Immediately above them sits a two-tiered temple that is 10 stories tall. The temple is much higher than the foliage of the surrounding rain forest and is visible from a great distance. The visual aspects of the pyramid are closely related to its ritual significance. To an observer on El Tigre, at about the time of the vernal equinox in certain years Jupiter, Mercury, Mars and Saturn all seem to rise out of the top of the Danta pyramid. Many sites of the Classic period were designed to capitalize on celestial phenomena, and some sites constructed near the end of the Late Preclassic in A.D. 300 already show such planning. The ceremonial complex at El Mirador, built some 400 years earlier, suggests that Maya knowledge of the heavens developed long before the beginning of the Classic period.

The iconographic and astronomical competence displayed at El Mirador are not isolated specimens. Artifacts from a variety of sites show that most of the elements of the shared intellectual culture typical of the Classic period had begun to be articulated and brought into relation with one another during the Late Preclassic. Among the central elements of that shared culture were literacy and numeracy. The most important functions served by those skills included depicting the gods and showing how the relations between gods and human beings underlay the succession of rulers in the Maya com-

munity. Such functions can be seen quite clearly in artifacts from the Late Preclassic.

The shared pantheon of the Late Preclassic is illustrated by findings from Cerros and Nohmul. At Cerros an offering found in the top of Structure 6B (a temple pyramid) included four jade heads of three facial types. An identical offering was discovered at Nohmul, 28 kilometers away, in the foundation of a small temple. Moreover, examples of at least two of the types of head have been found at sites spread across the Yucatán peninsula, suggesting that the population of the entire northern lowlands understood and embraced the iconography embodied in the stone heads. What might that iconography be?

I interpret the four heads as representations of three Maya gods. Moreover, I think they are the same gods depicted on the well-known Pomona jade flare. Jade flares were circular ear ornaments worn by Mayas of high status. In 1949 a particularly large and beautiful flare was discovered in a tomb at Pomona, a small site on the coast of Belize. By examining pottery found in the same tomb as well as the notes of the first scholars to see the flare, I have been able to assign the flare a date in the second part of the Late Preclassic period, perhaps about A.D. 100. The flare is a beautiful piece of jewelry, but its main interest lies in the four profile heads and the signs engraved around its circumference.

Two of the heads depict Kinich Ahau, the sun god, who is identified by the glyph (sign) for the Maya word *kin* on his cheek. The third profile is that of the maize god; the fourth is a long-lipped deity with the sign for *akbal,* or darkness, on his cheek. The maize god and the sun god opposite him are identified by the numbers 8 and 4 respectively; those are the numbers of which they were the patrons. The other signs associated with the heads have been deciphered by John S. Justeson of Vassar College and Will Norman of Tulane as a statement in an early version of the Yucatec Maya dialect about the interaction of the gods and the accession of a ruler. Justeson and Norman think the accession may be that of an early ruler of Yaxchilán, a major Maya city on the Rio Usumacinta far from Pomona.

The signs engraved on the Pomona flare yield some significant conclusions. In the first place, the numbers associated with two of the profiles show that the Maya system of enumeration had developed by about A.D. 100. (In the Maya system a dot stands for 1 and a bar stands for 5. Additional signs indicating position make it possible to compile very large numbers from these simple elements.) We found an example of the same bar-and-dot writing at Cuello. A tomb inside the pyramid there held several seals of roughly the same age as the Pomona flare. One of the seals bore the number 9: a bar and four dots. That finding reinforces the hypothesis that numeracy was part of the emerging intellectual complex of the Late Preclassic.

CARVED SEALS found at Cuello show that the Maya were numerate in Late Preclassic times. The seals may have been used for indicating the ownership of skins or cloth. The lower one bears the Maya number 9: a bar and four dots. (A bar stands for 5 and a dot for 1. The relevant bar is the line to the right of the dots; the line to the left is part of the carved design around the number.) The seals were found in a tomb inside the Cuello pyramid, which was built in about A.D. 100, but even earlier bar-and-dot inscriptions are known.

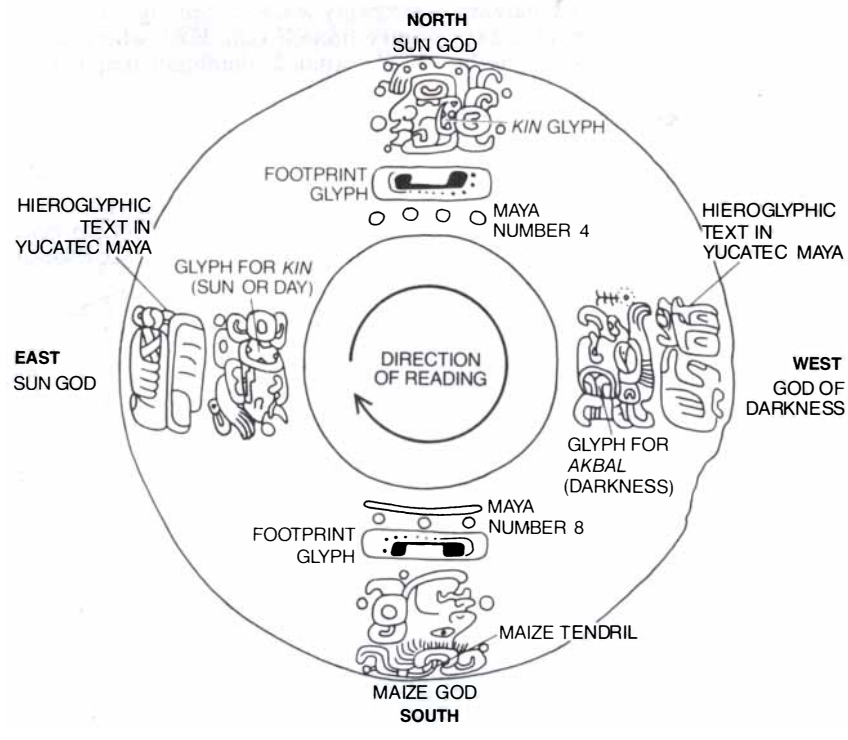
Perhaps even more significant is the conjunction of signs designating gods, rulers and numbers. One of the hallmarks of Classic Maya civilization is the recording of dates in the reign of a particular ruler. The dates, along with portraits of the ruler, were generally recorded on stone pillars called stelae. They served to commemorate the royal person and to rationalize his reign by connecting him with the gods. Indeed, in some instances the ruler is portrayed in a godlike guise. Stelae with dates and inscriptions were long taken to be one of the defining characteristics of the Classic period. Until a generation ago their appearance was dated at A.D. 300, but the discovery of Stela 29 at Tikal, bearing a date equivalent to 292, suggests the formal beginning of the Classic period should perhaps be set at 250.

Whatever date is chosen for the formal beginning of Classic times, it is increasingly clear that many elements of the "stelae cult" appeared during the Late Preclassic. At Cuello we found a plain stela that is only about 80 centimeters tall, far less than the fully developed stelae of the Classic period. The small stela was found on Platform 34 in front of the Late Preclassic pyramid, but excavation of the platform showed the stela had been erected when an earlier platform floor was in place. We believe the stone monument was put up between A.D. 50 and 100, or as much as 200 years before

the beginning of the Preclassic period. The Cuello stela lacks carving, but Late Preclassic monuments with portraits of rulers and even hieroglyphic inscriptions have also been found. Among them are works at El Mirador and a rock carving at San Diego in Guatemala.

All these Late Preclassic objects must be considered transitional in that they lack a crucial feature of stelae from the Classic period: inscribed dates in the Maya calendar. Even that final touch has been supplied by a few stone objects predating Classic times. A looted stela, currently in Seattle, has an inscribed hieroglyphic date equivalent to A.D. 199. Perhaps the earliest inscribed monument from the Maya lowlands, however, is a fragment of an altar from Polol in Guatemala. On the fragment two richly clad figures flank a column of numbers. At the top of the column is a sign known as the Initial Series Introducing Glyph, which announces a calendrical date. Unfortunately the stone is too eroded for the date to be read, but John A. Graham of the University of California at Berkeley believes that it falls within Cycle 7 of Maya time, which ended in A.D. 41. The correct date could be considerably earlier, since Graham notes striking parallels between the altar and two stelae from Abaj Takalik, a site he investigated on the Pacific coast of Guatemala. Although the latest possible date for the Abaj Takalik monuments is not long after A.D. 100, the earliest is about 300 B.C.

The development of calendrical inscriptions was only the last step in a long process of cultural advance that took place during the Late Preclassic. When the process was complete, the Maya had accumulated almost all the trappings of Classic civilization. Distinct representations of gods and rulers had been developed; those depictions were accompanied by both writing and numbers. The elaboration of a shared intellectual culture rested on a series of economic and social innovations that took up the long preceding stretches of the Late and Middle Preclassic periods. One of the most important questions remaining to be answered is why the process of social development accelerated so greatly just before the beginning of the Late Preclassic period. Perhaps in the decade to come that question, along with many others, will be answered. No matter how striking the revelations of the coming years may be, however, they will be hard put to match those of the past decade, which have radically transformed our view of the emergence of Maya civilization.



JADE FLARE found in 1949 at the Pomona Ranch in Belize bears examples of Preclassic Maya writing. The flare, 18 centimeters across, is an ear ornament meant to be worn on a headdress. As the footprint glyphs suggest, the symbols are read clockwise. Beginning at the left there appear: the sun god next to an inscription in the Yucatec Maya dialect; another depiction of the sun god, patron of the number 4; a god of darkness, and the maize god, patron of the number 8, with maize growing from his head. That the portraits on the flare match the heads found at Cerros and Nohmul shown on page 107 suggests how unified Maya culture was in Late Preclassic times. John S. Justeson of Vassar College and Will Norman of Tulane University interpreted the hieroglyphics as being in the Yucatec dialect.

THE AMATEUR SCIENTIST

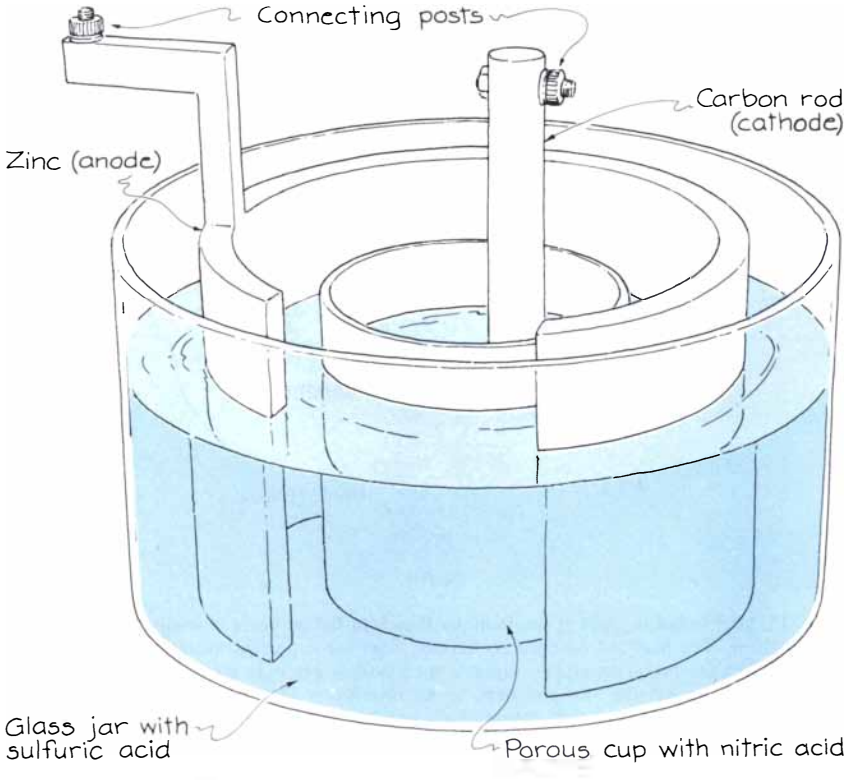
Retracing the steps by which aluminum metal was initially purified back in 1886

by Jearl Walker

The U.S. aluminum industry owes its existence to the remarkable work Charles Martin Hall did a century ago when he discovered a way to make pure aluminum metal. Norman C. Craig of Oberlin College has recently retraced Hall's steps. In describing Craig's adventure I rely on manuscripts by him and on historical research by Emily Nunn and William Bigglestone.

Hall made his discovery at the age of 22, eight months after he was graduated from Oberlin. His laboratory was in a woodshed behind his parents' house; his primitive equipment included several items he had made himself. Although he had pored over chemistry books since he was quite young, he had taken only two quarters of chemistry at college. His teacher there was Frank F. Jewett.

Jewett lectured on the properties of aluminum, displaying a lump of the metal he had brought back from Europe. The chemical methods then employed to make the metal were so costly that aluminum sold for about the same price as silver. Jewett told his students that the person who found a method for producing aluminum commercially would benefit the world and make himself rich. Hall, who had already studied aluminum metallurgy, turned to another student and said, "I'm going for that metal."



A Bunsen-Grove battery of the type that served Charles Martin Hall in 1886

At the time the method of making aluminum metal had not advanced much beyond the work done in 1825 by Hans Christian Oersted, the Danish scientist who is remembered for his pioneering research on electricity. Oersted heated a mixture of dilute potassium amalgam (potassium metal dissolved in mercury) and aluminum chloride. After removing the mercury by distillation, he found a small, gray lump of material that he supposed was aluminum metal. In 1827 the German chemist Friedrich Wöhler undertook to repeat the experiment. Failing to do so, he substituted potassium metal for the potassium amalgam. By means of this substitution the experiment yielded aluminum only as a black powder. Not until 1847 did Wöhler succeed in obtaining a sample of aluminum that was large enough to reveal the characteristics of the metal.

In 1854 the French chemist Henri Étienne Sainte-Claire Deville improved on Wöhler's procedure by substituting sodium metal for potassium metal. A sample of aluminum produced by Sainte-Claire Deville's technique was exhibited at the 1855 Paris Exposition. The metal was demonstrably strong, light in weight, durable and resistant to corrosion. Among those who were impressed was Napoleon III, who envisioned outfitting his army with aluminum equipment for the impending war with Prussia. Commissioned by Napoleon, Sainte-Claire Deville improved his techniques, but aluminum still cost too much for wide use. From 1860 to 1880 the worldwide production of aluminum was about 1.5 tons per year.

By the 1880's the world knew of aluminum's industrial possibilities. Hall started his work against this background. Initially he tried, as other experimenters had, to reduce aluminum from aluminum oxide (alumina) by purely chemical means. He and his older sister, Julia, prepared the aluminum oxide from household alum and washing soda. (Julia helped in many other ways, drawing on her college background in chemistry.)

Hall also attempted to reduce the cost of producing aluminum with the Sainte-Claire Deville method by seeking an inexpensive way of preparing anhydrous aluminum chloride. In addition he experimented with the reaction between sodium metal and the sodium aluminum fluoride called cryolite ($AlF_3 \cdot 3NaF$), trying to find a new and inexpensive chemical reduction method. None of the experiments showed promise.

While Hall was still in college he and Jewett had begun to investigate whether or not an electric current might make pure aluminum from an aluminum salt dissolved in water. The dissolved salt yields positive ions of aluminum. Hall and Jewett hoped that when current passed through the solution, electrons in the current would combine with the aluminum ions, creating aluminum metal.

These electrolysis experiments required several amperes of direct current. It was difficult to produce electricity in a small town such as Oberlin, and so Hall and Jewett resorted to a classical electrochemical cell invented earlier in the 19th century by Robert Wilhelm Bunsen, whose gas-burner design is still found in laboratories. A battery, sometimes known as a Bunsen-Grove battery, consists of one or more cells. Each cell centers on a porous ceramic cup holding concentrated nitric acid. A carbon-rod electrode is inserted in the acid. The cup rests in a container filled with a dilute solution of sulfuric acid (one part acid to 10 parts water). The electrode in the sulfuric acid is a sheet of zinc that is curved around the container until it almost forms a cylinder. The battery produces about 1.9 volts across the terminals, with the carbon rod acting as the positive electrode.

At the anode the zinc is converted into Zn^{2+} (a positive ion) with a release of electrons. At the cathode electrons combine with H^+ and NO_3^- to yield nitrogen dioxide gas and water. Hall's battery employed seven of the cells. Depending on how he connected them, they generated between five and 10 amperes of current.

The initial electrolysis experiments were carried out with aluminum fluoride dissolved in water. To Hall and Jewett's disappointment the products of the reaction were hydrogen gas and aluminum hydroxide at the cathode. These experiments were done in Jewett's laboratory during Hall's senior year. After his graduation in 1885 Hall moved most of his research to the woodshed.

Since water solutions of aluminum salts had proved unpromising, Hall began to consider dissolving aluminum oxide in fused salts: salts heated to their melting point. Most fused salts are liquids similar to water in viscosity and thermal conductivity. Moreover, they conduct electricity well. (The idea of dissolving aluminum oxide in anything was bold at the time because the material was regarded as inert. For this reason it was commonly included in firebricks.)

The coal-burning furnace Hall had used in earlier experiments could not generate and sustain the high temperature needed for the work with aluminum. Hall therefore set up a gasoline burner to heat the interior of an iron tube that he lined with clay. Although the new furnace could melt some fluoride salts such as potassium and sodium fluorides, it could not melt others such as calcium fluoride, aluminum fluoride and magnesium fluoride, each of which liquefies only at a temperature of at least 1,260 degrees Celsius. Unfortunately the pure fluoride salts that did melt in the furnace dissolved little aluminum oxide.

$$\underset{\text{anode (-)}}{Zn = Zn^{2+}(\text{aqueous}) + 2e^-}$$

$$\underset{\text{cathode (+)}}{2H^+(\text{aqueous}) + NO_3^-(\text{aqueous}) + e^- = NO_2(\text{gaseous}) + H_2O(\text{liquid})}$$

The reactions in a Bunsen-Grove cell

$$\underset{\text{cathode (-)}}{Al^{3+}(\text{melt}) + 3e^- = Al(\text{liquid})}$$

$$\underset{\text{anode (+)}}{2O^{2-}(\text{melt}) + C = CO_2(\text{gaseous}) + 4e^-}$$

Reactions in the Hall-Héroult aluminum process

Hall was an avid reader of *Scientific American,* which was then a weekly periodical. Craig believes Hall's work in electrolysis was spurred by a short item in the issue for October 24, 1885, reporting the success of a Mr. Graetzel in producing pure magnesium from a fused chloride salt through an electrolytic process. Hall later acknowledged that the process he discovered was similar to Graetzel's.

Hall still had to deal with the fact that his furnace could not produce temperatures high enough to melt some of the pure salts that might dissolve appreciable amounts of aluminum oxide. He decided to experiment with cryolite, a mixture of sodium fluoride salt and aluminum fluoride salt. Salt mixtures often have lower melting points than the pure salts from which they are made. Hall synthesized a sample of cryolite and demonstrated that its melting point was within the capability of his furnace. In later tests he added aluminum fluoride to reduce the melting point even more. When he added aluminum oxide to the hot liquid, he found that it easily dissolved. By February 10, 1886, the experiments were completed.

Six days later Hall did another crucial experiment. Could he produce pure aluminum by sending an electric current through the liquid salt in which aluminum oxide had been dissolved? He inserted two graphite rods into the hot solution, which was held in a clay crucible. Connecting one rod to the low-potential end of the Bunsen-Grove battery and the other rod to the high-potential end, he allowed the current to flow for about two hours. Then he poured the solution into a frying pan. When the material had cooled and solidified, he broke it apart. The experiment had failed. The deposit on the cathode was gray, not shiny as he knew aluminum should be.

Hall repeated the experiment with the same results. Eventually he realized the gray deposit probably came from the silica in the clay of the crucible. The silica dissolved in the hot solution, and then silicon was reduced to its elemental form by the electric current. Hall switched to a graphite crucible. The replacement was no mean feat, because large graphite rods were hard to obtain in the 1880's. Craig believes Hall got the graphite from nearby Cleveland, where the Brush Electric and Power Company had a supply of the material for its work in the new industry of arc lighting and the manufacture of dynamos.

On February 23, with the graphite crucible in place, Hall repeated the experiment. After the current had flowed for two hours, he disconnected the battery, poured out the solution and let it cool and solidify. Then he broke apart the solid material. Inside were several shiny nuggets of pure aluminum.

Hall was unaware that he had been in a race with Paul L. T. Héroult, a Frenchman who discovered a similar process and filed for a French patent in April. Today the basic electrolytic process for the production of aluminum metal goes by the name of the Hall-Héroult process.

When aluminum oxide dissolves in the hot cryolite melt, it probably dissociates into aluminum ions (Al^{3+}) and oxygen ions (O^{2-}). At the anode the oxygen ions combine with carbon in the electrode to become carbon dioxide gas, which then bubbles from the liquid. The conversion releases electrons that are pulled through the electrode to the high-potential end of the battery. Electrons are expelled from the low-potential end; they flow to the other electrode in the hot melt, where they combine with the aluminum ions to form liquid aluminum. As the aluminum accumulates it combines into globules that sink to the bottom of the melt owing to the lower density of the surrounding electrolyte. The globules cool and solidify, forming small nuggets of aluminum. Some of the nuggets obtained by Hall are currently on display by Alcoa (the Aluminum Company of America). The company refers to the nuggets as "the Crown Jewels" because of their brilliance and historical significance.

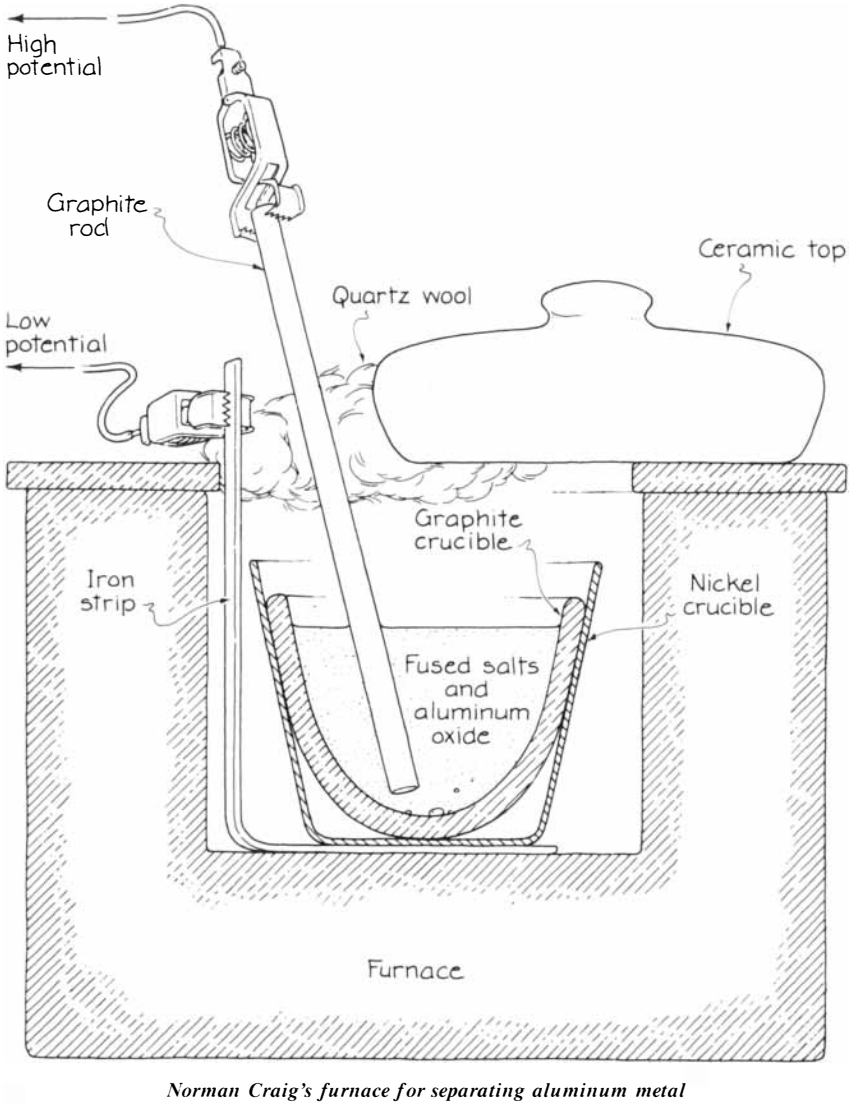
The maximum amount of aluminum reduced in the experiment can be calculated from the amount of current flowing through the liquid. Three electrons are needed to reduce one Al^{3+} ion to a neutral Al atom. Three moles of electrons are needed to reduce one mole of aluminum. A mole of electrons has a charge equal to the product of 6.022×10^{23} (Avogadro's number) and 1.6×10^{-19} coulombs (the charge of each electron). Thus a mole of electrons has a charge of 9.64×10^{4} coulombs. Suppose the current through the liquid is five amperes, that is, five coulombs per second. If the current flows for two hours, the total charge sent through the liquid is 3.6×10^{4} coulombs, or .37 mole of electrons. The current reduces a total of .12 mole of aluminum. One mole of aluminum has a mass of 26.98 grams. Hence the current produces about 3.4 grams of aluminum. Aluminum's density is 2.698 grams per cubic centimeter. If all the aluminum were collected into one spherical nugget, the diameter of the nugget would be 1.3 centimeters. In practice the amount of aluminum metal produced is about half the theoretical maximum because of inefficiencies in the process.

As Hall noted in a letter to his brother on the day of the successful experiment, the cryolite is mostly unchanged by the passage of current and so the process can be continued. As the initial amount of aluminum is reduced to a pure metal and taken out of solution, more aluminum oxide can be dissolved in the fused salt. Pure aluminum continues to form as long as the current flows and the anode (which is being consumed) is pushed farther into the liquid.

Craig re-created Hall's experiment in an electric pot furnace capable of reaching 1,000 degrees C. The ingredients are melted in a number 0000 graphite crucible of the "plumbago" type (volume 118 milliliters). This crucible in turn sits in a nickel crucible that protects against leaks in case the graphite cracks while it is being heated. The nickel crucible rests on a strip of soft iron three millimeters thick and 19 millimeters wide. The strip is bent so that it extends under the crucible and also up through the top opening of the furnace. The strip serves as the connection to the low-potential end of the source of current. The strip, the nickel crucible, the graphite crucible and the hot liquid form part of the electrical pathway.

The electrical connection Craig makes to the high-potential end of the current source is through a graphite rod immersed in the hot liquid. The rod is eight millimeters in diameter and 30.5 centimeters long. The rod and the iron strip are connected to a source of direct current by large alligator clips. The rod is held in place by an insulated clamp. The temperature of the furnace is monitored with a chromel-alumel thermocouple that is encased in a ceramic sheath and extends to the bottom of the furnace. A chemical hood set in place over the equipment draws off the fluoride fumes generated by the experiment.

Craig fills the crucible with a mixture of 130 grams of sodium fluoraluminate (cryolite) and 40 grams of aluminum fluoride. When the mixture melts at 830 degrees C., he adds three grams of aluminum oxide in the form of a fine powder. He then inserts the graphite rod and connects the current



Norman Craig's furnace for separating aluminum metal

source. The difference in voltage between the rod and the iron strip is about five volts and the current is about 10 amperes.

A ceramic lid covers most of the top opening of the furnace to reduce the heat loss. The graphite rod extends through the remaining opening, which Craig packs with quartz wool. The furnace is operated at a temperature of about 930 degrees C. The current is kept on for four hours, during which time another two grams of aluminum oxide is added to replenish the amount in solution. As the graphite rod is consumed, it is pushed farther into the liquid. At least once during the experiment the rod is so nearly consumed that it must be replaced.

After the furnace has cooled, the graphite crucible is taken out of it and cracked open with a hammer. At the bottom of the crucible lies a nugget of aluminum roughly one centimeter in diameter. The procedure usually yields several smaller nuggets.

Craig suggests that several changes can be made to reduce the cost of the experiment. The crucible could be heated with a butane torch in a furnace of the Hall type made from a ceramic drain tile. The electrical source can be an automobile storage battery to which an adjustable power resistor is added to decrease the voltage. The nickel crucible could be replaced with an iron crucible or possibly a clay flowerpot. The conducting strip of iron could be replaced with a second graphite rod immersed in the liquid in the same way as Hall did it. The thermocouple is not necessary because the melting of the materials in the crucible reveals the temperature. You may find other changes that can be made. Be careful, however, to do the experiment protected by a chemical hood or some other means of good ventilation. Also take precautions against spilling the hot liquid. In addition you must guard carefully against electric shock if the electrical source you are employing is at high voltage.

You might also enjoy investigating a Bunsen-Grove battery, but be extremely careful with the acids and work in a well-ventilated area. Craig constructed one of these batteries in a 400-milliliter beaker. A sheet of zinc served as one of the electrodes. The sheet curved along the wall of the beaker but did not form a complete circle. Craig measured an output of four amperes from the battery. If you repeat Hall's experiment with such a battery, be sure to have enough zinc on hand. About one pound of zinc is consumed for each ounce of aluminum that is produced.

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

COMPUTER RECREATIONS

MATHEMATICS, MAGIC AND MYSTERY. Martin Gardner. Dover Publications, Inc., 1956.

THE CONJURER'S CALCULATOR (MAGIC WITH A POCKET COMPUTER). John C. Sherwood. Micky Hades International, Box 476, Calgary, Alberta, 1976.

THE MAGIC BOOK: THE COMPLETE BEGINNER'S GUIDE TO ANYTIME, ANYWHERE, SLEIGHT-OF-HAND MAGIC. Putnam Publishing, 1977.

THE STRATEGIC NUCLEAR FORCES OF BRITAIN AND FRANCE

FRENCH NUCLEAR DIPLOMACY. Wilfrid L. Kohl. Princeton University Press, 1971.

GREAT BRITAIN'S STRATEGIC NUCLEAR DETERRENT. John F. Rybicki in *Armed Forces Journal International,* Vol. 118, No. 5, Whole No. 5676, pages 32–36; January, 1981.

EUROPE'S NUCLEAR SUPERPOWERS. George M. Seignious II and Jonathan Paul Yates in *Foreign Policy,* No. 55, pages 40–53; Summer, 1984.

ALLIED NUCLEAR FORCES COMPLICATE NEGOTIATIONS. Eric J. Grove in *Bulletin of the Atomic Scientists,* Vol. 42, No. 6, pages 18–23; June/July, 1986.

QUASICRYSTALS

WHICH TETRAHEDRA FILL SPACE? Marjorie Senechal in *Mathematics Magazine,* Vol. 54, No. 5, pages 227–243; November, 1981.

METALLIC PHASE WITH LONG-RANGE ORIENTATIONAL ORDER AND NO TRANSLATIONAL SYMMETRY. D. Shechtman, I. Blech, D. Gratias and J. W. Cahn in *Physical Review Letters,* Vol. 53, No. 20, pages 1951–1953; November 12, 1984.

THE STRUCTURE AND STATISTICAL MECHANICS OF GLASS. D. R. Nelson in *Applications of Field Theory to Statistical Mechanics: Proceedings of the Sitges Conference on Statistical Mechanics, Sitges, Barcelona/Spain, June 10–15, 1984,* edited by L. Garrido. Springer-Verlag, 1985.

PENTAGONAL AND ICOSAHEDRAL ORDER IN RAPIDLY COOLED METALS. David R. Nelson and Bertrand I. Halperin in *Science,* Vol. 229, No. 4710, pages 233–238; July 19, 1985.

TRANSPLANTATION IN THE CENTRAL NERVOUS SYSTEM

NEURAL TISSUE TRANSPLANTATION RESEARCH. Edited by Robert B. Wallace and Gopal D. Das. Springer-Verlag, 1983.

NEURAL GRAFTING IN THE MAMMALIAN CNS. Edited by Anders Björklund and Ulf Stenevi. Elsevier Science Publishers, 1985.

CHOLINERGIC VENTRAL FOREBRAIN GRAFTS INTO THE NEOCORTEX IMPROVE PASSIVE AVOIDANCE MEMORY IN A RAT MODEL OF ALZHEIMER DISEASE. Alan Fine, S. B. Dunnett, A. Björklund and S. D. Iversen in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 82, No. 15, pages 5227–5230; August, 1985.

PROGRESS IN LASER FUSION

LASER COMPRESSION OF MATTER TO SUPER-HIGH DENSITIES: THERMONUCLEAR (CTR) APPLICATIONS. John Nuckolls, Lowell Wood, Albert Thiessen and George Zimmerman in *Nature,* Vol. 239, No. 5368, pages 139–142; September 15, 1972.

LASER-DRIVEN FUSION. Keith A. Brueckner and Siebe Jorna in *Reviews of Modern Physics,* Vol. 46, No. 2, pages 325–367; April, 1974.

QUANTUM ELECTRONICS. Amnon Yariv. John Wiley & Sons, Inc., 1975.

AN OVERVIEW OF INERTIAL FUSION REACTOR DESIGN. M. J. Monsler, J. Hovingh, D. L. Cook, T. G. Frank and G. A. Moses in *Nuclear Technology/Fusion,* Vol. 1, No. 3, pages 302–358; July, 1981.

THE ORIGIN OF CORN

CORN: ITS ORIGIN, EVOLUTION, AND IMPROVEMENT. Paul C. Mangelsdorf. Harvard University Press, 1974.

ZEA DIPLOPERENNIS (GRAMINEAE): A NEW TEOSINTE FROM MEXICO. Hugh H. Iltis, John F. Doebley, Rafael Guzmán M. and Batia Pazy in *Science,* Vol. 203, No. 4376, pages 186–188; January 12, 1979.

THE MYSTERY OF CORN: NEW PERSPECTIVES. Paul C. Mangelsdorf in *Proceedings of the American Philosophical Society,* Vol. 127, No. 4, pages 215–247; August, 1983.

THE ARCHAEOLOGICAL RECORD ON THE PROBLEM OF THE DOMESTICATION OF CORN. R. S. MacNeish in *Maydica,* Vol. 30, pages 171–178; 1985.

THE SOLAR CYCLE IN PRECAMBRIAN TIME

HISTORICAL EVIDENCE FOR THE EXISTENCE OF THE SOLAR CYCLE. John A. Eddy in *The Solar Output and Its Variation,* edited by Oran R. White. Colorado Associated University Press, 1977.

PRECAMBRIAN VARVES AND SUNSPOT CYCLES. G. E. Williams in *Weather and Climate Responses to Solar Variations,* edited by Billy M. McCormac. Colorado Associated University Press, 1983.

SOLAR AFFINITY OF SEDIMENTARY CYCLES IN THE LATE PRECAMBRIAN ELATINA FORMATION. G. E. Williams in *The Australian Journal of Physics,* Vol. 38, pages 1027–1043; December, 1985.

SOLAR SIGNATURE IN SEDIMENTARY CYCLES FROM THE LATE PRECAMBRIAN ELATINA FORMATION, AUSTRALIA. G. E. Williams and C. P. Sonett in *Nature,* Vol. 318, No. 6046, pages 523–527; December 12, 1985.

THERMAL NICHES OF STRIPED BASS

SYNOPSIS OF BIOLOGICAL DATA ON STRIPED BASS, *MORONE SAXATILIS* (WALBAUM). E. M. Setzler et al. NOAA Technical Report, National Marine Fisheries Service Circular 433, U.S. Department of Commerce, June, 1980.

STRIPED BASS: ENVIRONMENTAL RISKS IN FRESH AND SALT WATER. *Transactions of the American Fisheries Society,* Special Symposium Issue, edited by Charles C. Coutant, Vol. 114, No. 1; January, 1985.

THE EMERGENCE OF MAYA CIVILIZATION

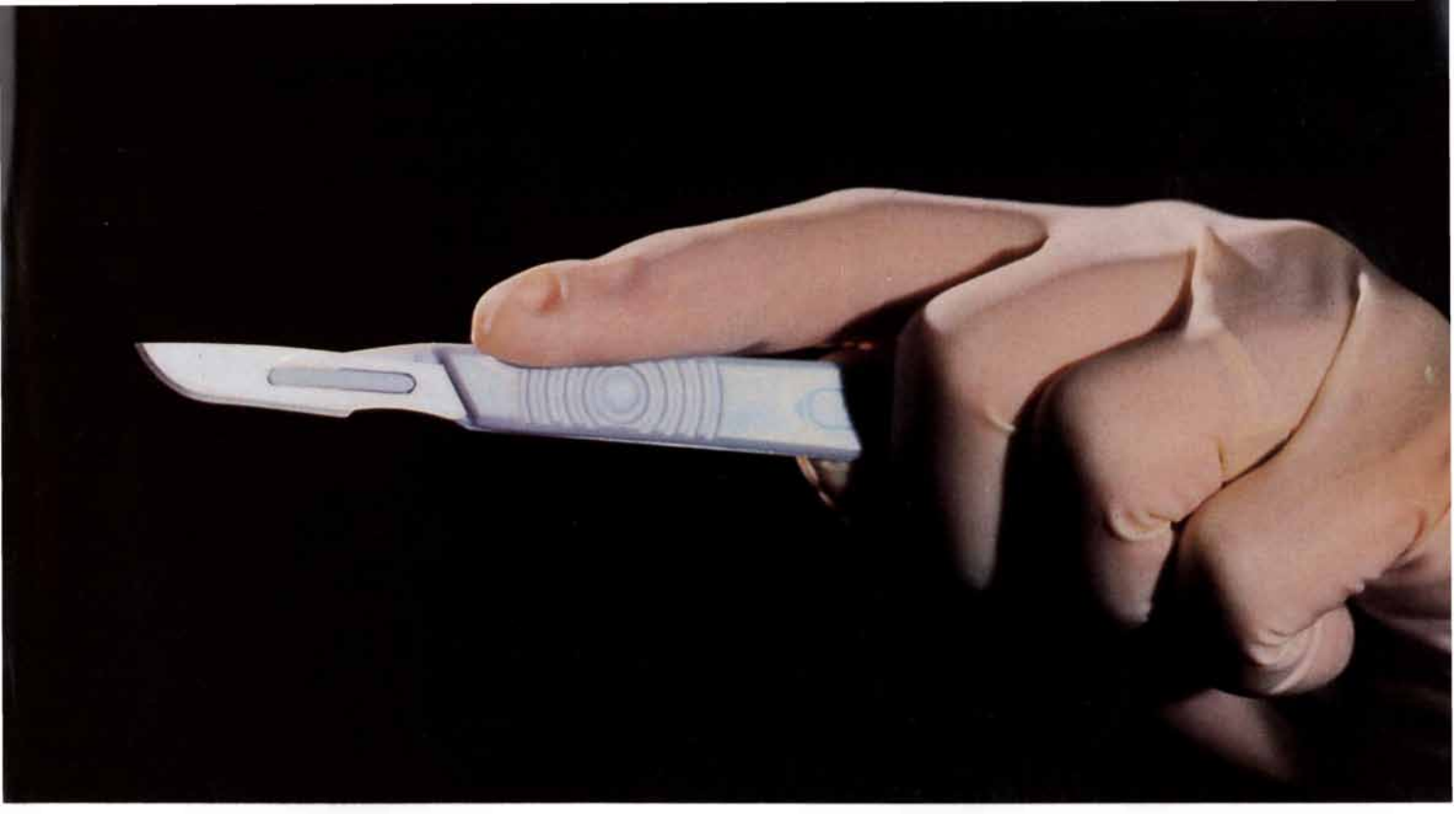
SOCIAL PROCESS IN MAYA PREHISTORY. Edited by Norman Hammond. Academic Press, 1977.

ANCIENT MAYA CIVILIZATION. Norman Hammond. Rutgers University Press, 1982.

THE AMATEUR SCIENTIST

THE IMMORTAL WOODSHED: THE STORY OF THE INVENTOR WHO BROUGHT ALUMINUM TO AMERICA. Junius David Edwards. Dodd, Mead, 1955.

CHARLES MARTIN HALL—THE YOUNG MAN, HIS MENTOR, AND HIS METAL. N. C. Craig in *Journal of Chemical Education,* Vol. 63, No. 7; July, 1986.



Until recently, this was the leading edge.

The leading edge of technology is replacing the surgeon's scalpel in the removal of kidney stones. Today, almost 90 percent of kidney stone sufferers have an alternative—the Lithotripter.

Developed by Dornier, a member of the Daimler-Benz Group, the Lithotripter disintegrates kidney stones through shock waves. Described as a "medical miracle" by the Secretary of Health and Human Services, the Lithotripter eliminates the pain and discomfort normally associated with surgical procedures. And because it is completely noninvasive, the Lithotripter significantly shortens recovery time, thus reducing cost to the patient. Thus far, over 100,000 kidney stone sufferers have been successfully treated by the Lithotripter.

Dornier developed the Lithotripter as a technological refinement of research which had originally been targeted toward studying the effect of shock waves on supersonic aircraft. Lithotripter—it's the only device of its kind in the world, working today for a healthier future.

DORNIER

A MEMBER OF THE DAIMLER-BENZ GROUP

1986 Mercury Sable. When a wagon is a great road car, this is what it looks like.

The idea that a wagon can be a great road car is indeed extraordinary. Which is why Sable is no ordinary looking wagon.

Its sophisticated shape not only permits it to move easily through the air but also helps it hold the road.

Sable's windows are fitted flush to its body sheet metal, smoothing the passage of air over the car as well as quieting the ride.

Sable has front-wheel drive, rack-and-pinion steering, multi-port fuel-injected 3.0-liter V-6 and four-wheel independent suspension.

And Sable is protected by a three-year, unlimited-mileage powertrain warranty. $100 deductible applies after 12,000 miles. Ask to see this limited warranty at your dealer.

For more information, call toll-free 1-800-822-9292

LINCOLN-MERCURY DIVISION Ford

MERCURY. The shape you want to be in.