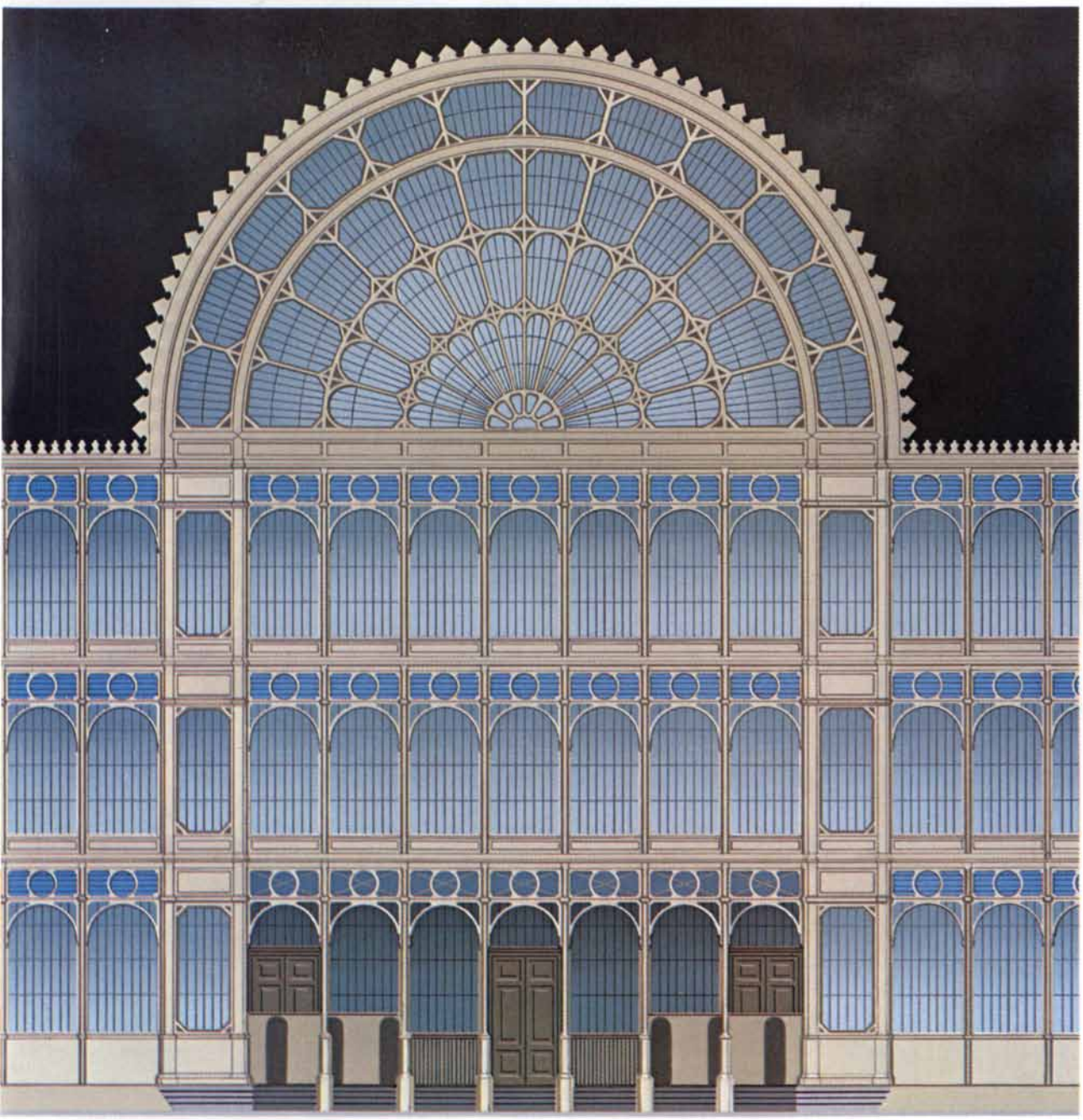
SCIENTIFIC AMERICAN



THE CRYSTAL PALACE

$2.50

October 1984

THE NEW RENAULT ALLIANCES

Imagine the challenge. You introduce Renault Alliance. The car that made European technology affordable. What do you do to step up the benefits dramatically without overstepping affordability? Introducing the new Renault Alliance Convertible–the lowest list-priced convertible in America. The best-selling Alliance—the DL. And the luxurious Alliance Limited. Each available with an impressive list of optional features: An electronically tuned AM/FM six-speaker stereo cassette system. Computerized Sys-

Alliance 2-door **$6161**†

†Base model 1.4-litre engine. List price. Tax, destination charges, optional/regional equipment extra. All options not available on all models.

tems Sentry that monitors vital fluids and brakes. An entry system utilizing infrared technology to lock and unlock your doors from a distance. Power windows. Even a choice of engines, including a new 1.7 litre. Plus all

UPWARD AFFORDABILITY

of Alliance's high standards: Sedan comfort for as many as five, a quiet, smooth ride, pedestal seats for extra rear legroom, electronic fuel injection for efficiency: **40** EST HWY, **34** EPA EST MPG.* The new front wheel drive Alliances. From Renault, Europe's leading car maker. *Built in America.* And priced to keep down the cost of moving up.

RENAULT

THE ONE TO WATCH

*Base model 1.4-litre engine. Use for comparison. Your mileage may vary with speed, trip length, weather. Actual highway figures lower.

Buy or Lease. From American Motors. Safety Belts Save Lives.

Wild Turkey Hill.
The legend began 150 years ago.

For signed Ken Davies lithograph, 18" x 19" of Wild Turkey painting, send $10 payable to "ANCO," Box 929-SN, NYC 10268

There's been a distillery on Wild Turkey Hill for almost 150 years. Sometimes the air stands motionless in the valley, where the gentle-running waters of the Kentucky River flow. But on top of the hill, constant breezes cool our Wild Turkey whiskey naturally as it ages in the barrel. Wild Turkey Hill is a place like no other. And it helps us make a Wild Turkey whiskey like no other.

WILD TURKEY®/101 PROOF/8 YEARS OLD

AUSTIN NICHOLS DISTILLING CO., LAWRENCEBURG, KENTUCKY © 1984

Established 1845 **SCIENTIFIC AMERICAN** October 1984 Volume 251 Number 4

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 84121. 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

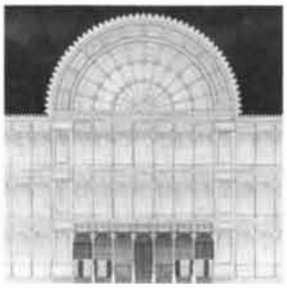
State and ZIP

Old Address

Street

City

State and ZIP



THE COVER

The painting on the cover depicts a view of the Crystal Palace. Built in London's Hyde Park in 1851 to house the Great Exhibition of the Works of Industry of All Nations (see "The Crystal Palace," by Folke T. Kihlstedt, page 132), the Crystal Palace was perhaps the greatest achievement on display. Remarkable not only for its size (it enclosed a volume of 33 million cubic feet and required 400 tons of glass) and the speed with which it was built (it was completed in only 17 weeks) but also for the aesthetic and structural innovations introduced by its designer, Joseph Paxton, the Crystal Palace presaged the look and method of modern architecture. The walls consisted of curtains of glass hung on an iron framework; the building was made of identical modular units designed for rapid construction; the major structural members, such as trusses and cross braces, were unadorned and clearly visible. The view on the cover is of the main entrance, which was at the south end of the arched central transept.

THE ILLUSTRATIONS

Cover painting by Ted Lodigensky

Page	Source
13	SCIENTIFIC AMERICAN
20	Lee Sallows
22–27	Edward Bell
40–45	George V. Kelvin
46	Ian Worpole
48–49	U.S. Army
51–52	Stanley B. Prusiner, School of Medicine of the University of California at San Francisco
53–58	Ilil Arbel
59	Stanley B. Prusiner, School of Medicine of the University of California at San Francisco
61	Ian Worpole
62	Andrew Tomko
63	Adam M. Dziewonski
64–65	Ian Worpole
66–68	Andrew Tomko
74	National Radio Astronomy Observatory
75	The Granger Collection
80	Craig Taylor, Woods Hole Oceanographic Institution
85	Carol Donner
86–87	Arnold I. Caplan, Case Western Reserve University
90–91	Hank Iken, Walken Graphics
92–94	Carol Donner
99	Hank Iken, Walken Graphics
100–104	Gabor Kiss
105	Hank Iken, Walken Graphics
107	Tom D. Dillehay, University of Kentucky
108–111	Patricia J. Wynne
112	Tom D. Dillehay, University of Kentucky
117	Patricia J. Wynne
121	N. Iroshnikova
124–129	S. Stulov
132–133	The Illustrated London News Picture Library
134–135	SCIENTIFIC AMERICAN
136–139	Alan D. Iselin
140	Mary Evans Picture Library (*top*); Astor, Lenox and Tilden Foundations, The New York Public Library (*bottom*)
141	R. F. Bonifield, courtesy Engineering Societies Library (*left*); courtesy Sears, Roebuck and Co. (*right*)
142	Jon Brenneis (*top*); George Hall, Woodfin Camp & Associates (*bottom*)
144	Quesada/Burke
146–152	Michael Goodman

By 8 a.m., 110 of USAir's 121 planes are in the air.

One luxury most business travelers don't have is time. So USAir helps you make the most of yours.

We do it by getting 91% of our fleet off the ground by 8 a.m., so you can get an early start on your day.

We even design many of our schedules so that you can fly out and back in the same day and still have a full business day at your destination. That's especially true in the Northeast, where we offer *80% more* domestic departures than any other airline.

We're strong in the Northeast because that's where we started. But we didn't stop there. Today, our system stretches as far west as California, Arizona and Colorado and as far south as Texas, Florida and Louisiana. In all, we serve more than 90 cities in the U.S. and Canada.

Giving you convenient schedules, professional service and money-saving fares has helped make us one of the most successful major airlines in the world.

For information or reservations, contact your travel agent or corporate travel arranger. Or give us a call at USAir. We'll be up early to serve you—because we want our airline to be your airline.

USAIR

WHY RCA IS

In the vast universe of corporations, RCA stands alone. Certainly, others manufacture television sets. Still others are major forces in communications, broadcasting, defense and aerospace. But no other company does all of the things that RCA does.

That distinction not only makes RCA unique, it has enabled us to develop talents and abilities that, we believe, no other single corporation possesses.

Three years ago, new top management at RCA took a long, hard look at those talents and abilities. And at the industries in which we compete. The resulting strategy was a simple one—do what we do best. In short, focus on three businesses that offer the greatest potential for future growth—electronics, communications and entertainment.

The results have been dramatic. In home video systems RCA has built a leading market share in television receivers, video cassette recorders and cameras. And we're pioneering the ingenious charge-coupled-device (CCD) in a revolutionary broadcast camera that can practically see in the dark.

In space, our business has taken off. RCA is the leader in the design and construction of both communications and meteorological satellites. And our cameras and video equipment are playing a major role in the Space Shuttle program.

AEGIS, the Navy's seaborne weapons defense system, was developed and built by RCA. The first two cruisers in this new series are now in service and we've been awarded contracts for eleven additional ships.

RCA
ONE OF A KIND

In 1983, NBC captured more prime-time Emmy Awards than the other two networks combined. RCA Records had a smash year with hits by Kenny Rogers, Alabama and Hall & Oates. And, in conjunction with Columbia Pictures, RCA has become a leader in prerecorded music/video cassettes.

This tremendous success has had another effect—RCA's earnings in the second quarter were the highest in the Company's history.

If these facts surprise you, write to us at: This Is RCA, P.O. Box 91404, Indianapolis, Ind. 46291 and we'll send you a few more surprises. You see, we want you to rethink RCA. As a corporation, as an investment or as a place to work. Because when you do, you'll realize that RCA really is one of a kind.

ONE OF A KIND.

LETTERS

Sirs:

Ellen L. Bassuk's moving report ["The Homelessness Problem," SCIENTIFIC AMERICAN, July] on the travails of the mentally disabled homeless poor merits praise for its careful depiction of the contribution of deinstitutionalization to the crisis of homelessness and for its rejection of shelter as a solution. But it overstates, in our view, the case for disability as the greatest of the hardships of the homeless.

It has not been established, as Dr. Bassuk initially claims, that "a large majority" of those who are currently homeless "suffer from mental illness, ranging from schizophrenia to severe personality disorders." Rather, the bulk of the evidence we have collected supports the more modest version of her assertion, that "a significant fraction of the people who frequent shelters have diagnosable mental disturbances."

In the first place, two of the three studies to which she alludes in support of her Boston findings were conducted under circumstances that made the finding of a high proportion of mental distress a foregone conclusion. The only New York City study of any reliability that found overwhelming evidence of mental disorder took place in the psychiatric emergency room of a municipal hospital. The Philadelphia study was done in a shelter specifically set up to serve disturbed homeless individuals.

Dr. Bassuk recognizes that most such studies, her own included, "have been largely descriptive and have been plagued by methodological problems." But she ought also to have noted that most studies have been done on populations that cannot, on prima facie grounds alone, be taken as representative of the homeless poor as a whole.

Moreover, the existing psychiatric profiles of the homeless depict only single men and women. While this would have been adequate two decades ago, great changes have since taken place in the ranks of the homeless. The U.S. Department of Housing and Urban Development's national shelter survey found that 21 percent of the shelter population today is made up of family members. In places such as New York City the proportion is even higher: homeless family members in emergency accommodations currently outnumber the single homeless in shelters by a ratio of 1.7 to one. It would be a mistake to assume that clinical findings about the single homeless—even supposing they are reliable—apply to homeless families....

An enlightened public policy toward homelessness must begin with the recognition that what is distinctive about the homeless poor is not their disability (although it may be marked) but the peculiar form their poverty takes: having no place to live.

KIM HOPPER

ELLEN BAXTER

Community Service Society
New York, N.Y.

Sirs:

Kim Hopper and Ellen Baxter raise important questions about the relation between mental illness and homelessness. Their comments highlight the conflicts that inevitably result when comprehensive research about a subject is lacking. As I stated, I agree that "the precise extent to which mental illnesses are prevalent among the homeless remains a matter of controversy."

In spite of serious methodological problems, however, there seems to be mounting evidence that a significant fraction of people who frequent shelters are psychiatrically disabled. As described in my article, it includes several New York City studies as well as a finding described in a footnote in an article by Hopper and Baxter in the *American Journal of Orthopsychiatry*. Psychiatric screening at the Keener facility in New York City, they reported, indicated that "84 percent of the men were determined to be mentally ill to some degree," although they commented that the sample might not have been representative.

To respond to Hopper and Baxter's criticism of the Philadelphia study: Although its authors did not define criteria for admission to the emergency shelter or describe those individuals seeking help who were referred elsewhere, the authors stated that the shelter "was considered a food and lodging program, not a treatment site."

As Hopper and Baxter point out, no systematic clinical data have been collected describing the psychiatric or medical needs of homeless families, even with their appearance in large numbers in recent years. The research findings described in my article are intended to refer to homeless individuals only.

Whatever the precise figures are, however, the fact remains that many homeless people are disabled and require humane care. The point is not whether psychiatric impairment is "the greatest of the hardships of the homeless." The point is that shelters are an inadequate long-term solution to the problems of the homeless whether they are disabled or not. Disability only indicates the need for services that go beyond permanent housing. As discussed in my article, the debate on homelessness would undoubtedly be enlightened by more rigorous research. Only with reports from researchers and service providers around the country can we develop a comprehensive picture of the problem that we can translate into rational public policy.

ELLEN L. BASSUK

Harvard Medical School
Boston, Mass.

Sirs:

First you report the behavior control of vertebrates by internal parasites ["Parasites That Change the Behavior of Their Host," by Janice Moore; SCIENTIFIC AMERICAN, May], reminiscent of the classic alien-infestation theme of science fiction. Now the "Ice Nine" of Kurt Vonnegut's *Cat's Cradle* has been surpassed ["Ice X," "Science and the Citizen," SCIENTIFIC AMERICAN, June].

As Vonnegut's protagonist discovers, teaching water new tricks can be risky. Drs. Polian and Grimsditch may want to review the conclusion to *Cat's Cradle* before proceeding.

STEPHEN DENZER

JOHN A. BURNS

Honolulu, Hawaii

Sirs:

Jearl Walker's discussion of the physics of his grandmother's peerless homemade ice cream ["The Amateur Scientist," SCIENTIFIC AMERICAN, April] roused memories of *my* grandmother's. But our ancestors are in some disagreement. The dasher on our family's freezer came in two parts, an inner part with "ears" and an outer one with a wood scraper. The inner part turned clockwise with the outer container, and the scraper went counterclockwise, thereby stirring the mixture more thoroughly than the freezer Walker describes.

When our ice cream was finally finished, it was de rigueur to remove the entire dasher assembly, both because the frozen cream was too hard to remove later with the dasher still inside and because I was allowed to lick it—the immediate reward for my work. Finally, we put a cork in the top of the container to keep out salt water and packed the apparatus in newspapers, better insulation than the towel Walker mentions.

The fact that the temperature of the salt-ice bath, about −15 degrees Celsius, is close to 0 degrees Fahrenheit is no accident: in the early 18th century Fahrenheit calibrated his thermometer from that temperature because the greatest cold he could achieve with "scientific" control was that of the salt-ice mixture.

WILLIAM B. HUNTER

Houston, Tex.

Robotic Control Language

IBM

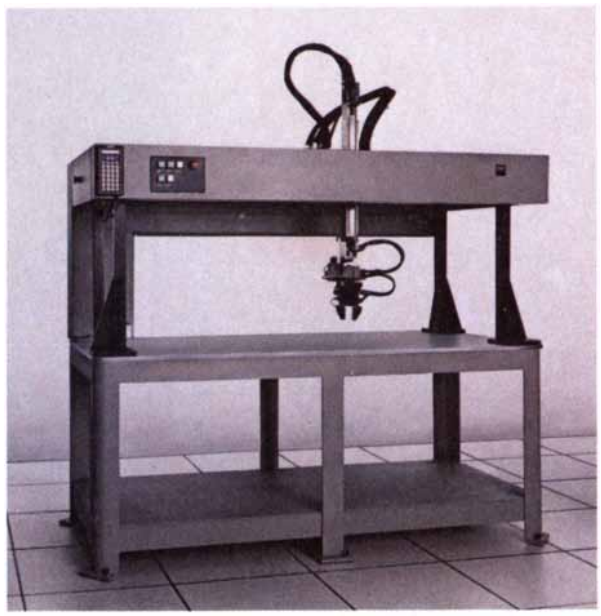
To: Jeanine
From: Bill
Subject: IBM Technology

I've been reviewing some of our past and present technological achievements, and it occurred to me that the scientific, engineering, and academic communities might like to know more about them. Will you select a topic from the following list? Thanks.

Vacuum tube digital multiplier
IBM 603/604 calculators
Selective Sequence Electronic Calculator (SSEC)
Tape drive vacuum column
Naval Ordnance Research Calculator (NORC)
Input/output channel
IBM 608 transistor calculator
FORTRAN
RAMAC and disks
First automated transistor production
Chain and train printers
Input/Output Control System (IOCS)
STRETCH computer
"Selectric" typewriter
SABRE airline reservation system
Removable disk pack
Virtual machine concept
Hypertape
System/360 compatible family
Operating System/360
Solid Logic Technology
System/360 Model 67/Time-Sharing System
One-transistor memory cell
Cache memory
Relational data base
First all-monolithic main memory
Thin-film recording head
Floppy disk
Tape group code recording
Systems Network Architecture
Federal cryptographic standard
Laser/electrophotographic printer
First 64K-bit chip mass production
First E-beam direct-write chip production
Thermal Conduction Module
288K-bit memory chip
Robotic control language

Bill—
Our robotic control language is an important factor in programmable automation. It's a terrific story—let's go with it.
Jeanine

Figure 1. *IBM's robotic control language, AML (A Manufacturing Language), is used in all robotic systems that IBM markets. Shown here is the high-end product, the IBM 7565 Manufacturing System, which represents leading-edge technology in intelligent robotics for complex assembly. This system is controlled by an IBM Series/1 computer that monitors manipulator position and sensory feedback 50 times per second, making real-time adjustments if necessary. IBM's mid-range products use AML/Entry, a simplified version of AML, enabling them to be programmed with IBM Personal Computers.*



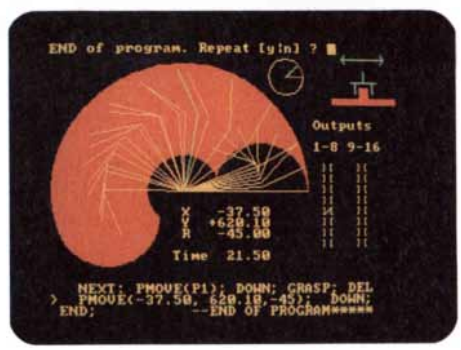


Figure 2. *This display from the AML/Entry Application Simulator is used to check the logic and efficiency of an AML/Entry program, display the robotic arm manipulation (from above), and provide timing estimates.*

Manufacturers everywhere face increasing pressure to produce products of higher quality at lower costs. Today, programmable robotic systems, with their ability to adapt to diverse manufacturing environments, are tools that help meet this challenge.

More than a decade ago, when IBM researchers began investigating flexible automation, they noted that control of manipulation was only one aspect of successful robotic applications; related tasks included terminal and storage input and output, communications, and computation. To integrate all these tasks, they designed a new general-purpose computer language and extended it with functions needed to control a robotic system.

The language that evolved, AML (A Manufacturing Language), is the most advanced robotic control language commercially available today.

In addition to robotic commands such as MOVE, SPEED, and MONITOR—which control motions through sensory feedback—AML has a wide range of communications and data processing capabilities.

AML is able to support different levels of user sophistication because of its structure: a powerful base language designed for functional enhancement through subroutines and application packages. For example, the GRASP subroutine found in many assembly applications is written in AML but is used exactly as if it were a primitive command. Experienced programmers can combine the existing base commands to construct higher-level routines.

AML is an interactive language. It provides the user with the ability to stop a program, check the logical and physical status of the system, change the program, and continue execution. This is critical for efficient development of robotic applications, which must deal with the variability of the real world.

AML has proved to be well adapted for implementing a wide variety of operator inter-

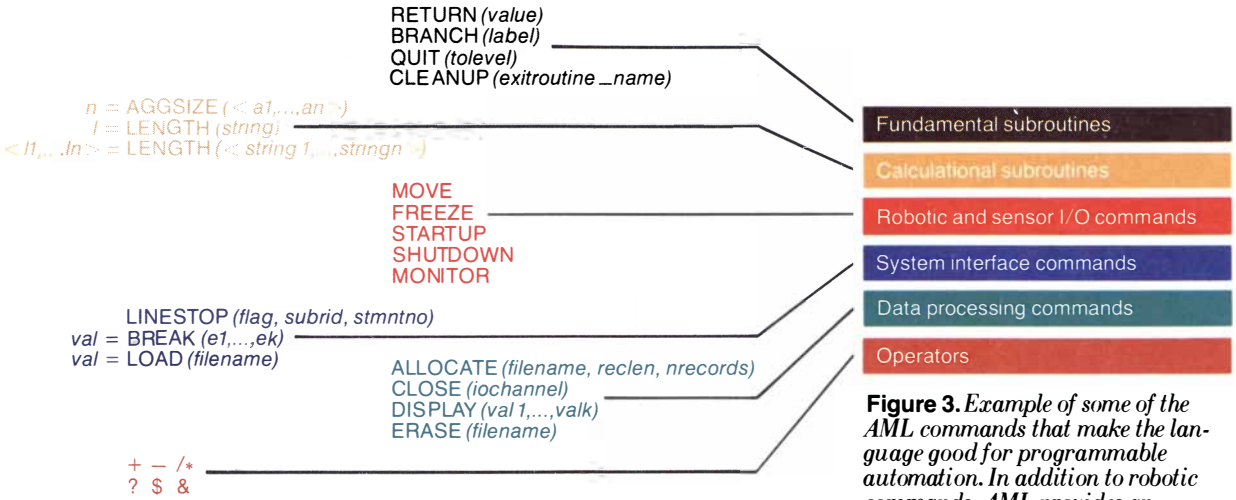


Figure 3. *Example of some of the AML commands that make the language good for programmable automation. In addition to robotic commands, AML provides an abundance of communications and data processing commands.*

faces, including a menu-driven display screen and the common "guiding through the motions" method. In the latter, the operator moves the manipulator through the steps of a task by using a hand-held, push-button pendant. After the operator completes the steps, the system automatically writes an efficient program in AML.

IBM uses its own robotic technology. For instance, IBM is working on computer-integrated manufacturing of typewriters: more than 250 robotic units will put together most of the typewriter subassemblies in an automated plant the size of two football fields. In other sites throughout the world, IBM robotic systems are used in such applications as testing circuits, producing cables, and assembling printer type chains.

Many IBM scientists, engineers. and programmers contributed to the development of the innovative robotic control language. AML. Their contributions are only part of IBM's continuing commitment to research, development, and engineering.

IBM

For a free technical article about IBM's robotic control language, please write: IBM Thomas J. Watson Research Center Dept. 403E/P3, P.O. Box 218 Yorktown Heights, NY 10598

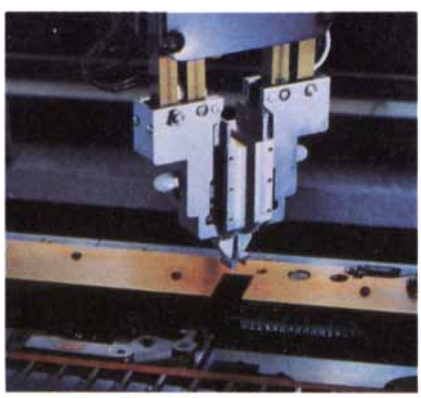


Figure 4. *In the automatic assembly of type chains for an IBM high-speed printer—one of the many examples of IBM robotic systems at work within the company—an AML program is used to consult a data base to determine the correct sequence of type slugs. This application makes extensive use of sensing and programmed error recovery to ensure high reliability.*

```
Step_1:
  CMOVE (<feeder_app(fdr),        --Move to
          feeder_orient, .5>);      grasping
                                    position

  IF DCMOVE (<<0,0,–.75>>,
             ANY_FORCE (2*OZS),
             <.5>) THEN
    BEGIN                         --Hit something
                                      on way in
    DCMOVE (<<0,0,2>>);           --Back out
    OP_CHECK ('jammed');          --Notify
    END;                              operator
Step_2:
  cc = GRASP (0.1,                --Attempt to
              <–.04, .04>,          grasp slug
              PINCH_FORCE (1*LBS));
```

Figure 5. *This is an excerpt of AML code from the program for the application shown in Figure 4. It directs the gripper to open 0.5 inches while approaching a feeder for the next slug of type. It then moves the gripper to the grasping position and grasps the slug with a gripping force of one pound. If an unexpected force is encountered while approaching the feeder, appropriate error-recovery actions are taken.*

GM

New *Calais* from Oldsmobile.

Building a car of this quality meant revolutionizing the way cars are built.

Of all the things that changed when the 1985 front-wheel-drive Calais was created, what changed most was the thinking behind it. It was clear, the only way to build a car to these exacting standards was to build it a whole new way.

It starts with computer design to engineer in quality from the word go.

On the assembly line, work teams take over. Every member receives 40 to 100 hours of specialized training before a single bolt is tightened. Over 200 robots ensure that welds are computer precise. And the critical parts of the body are electro-chemically coated to fight rust.

Another innovation: a computer-built, full-size model of the new Calais to help assure precise fit. Like a giant jigsaw puzzle, it makes sure parts fit the way they're supposed to. If the size of a fender is even 1/1000th of an inch off, electronic sensors will catch it.

One thing, though, hasn't changed where the new Olds Calais is concerned. It's that special Oldsmobile feeling of pride you enjoy every time you take a ride.

The new Olds Calais. Test drive one. You'll agree everything about it is, well, downright revolutionary.

There is a special feel in an Oldsmobile

Let's get it together...buckle up.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

OCTOBER, 1934: "Early in August Dr. William Beebe and his associate, Otis Barton, broke all deep-sea diving records in their bathysphere by descending 3,028 feet below the ocean surface off Bermuda. This famous spherical diving chamber is equipped this year with windows of fused quartz, which provided perfect protection during the record dive and no doubt will prove as satisfactory at the still greater depths to which Dr. Beebe and Mr. Barton expect to descend shortly. In the Bermuda seas many millions of creatures that inhabit the blackest of the ocean half a mile deep will be studied by Dr. Beebe, who now will be able to photograph all stages of sea life."

"Latest records show 31,000 motor-vehicle fatalities in 1933, or 5 percent above those in 1932. We cannot hope ever to control entirely the human emotions that cause accidents due to carelessness; it seems next to impossible to protect the motorist against himself by enacting regulatory laws that are so frequently broken. But there is one cause of accidents that can and should be carefully considered by every community, and the remedy applied. A study of the situation shows that where streets are provided with the best light, accidents have decreased as compared with streets where the lighting is poorer."

"Mercury, the liquid metal which we knew as the fascinating quicksilver in childhood days, is responsible for a new light source that is expected to exert considerable influence on artificial lighting practice of the future. This new light source is known as the high-pressure mercury lamp. The principle of producing light by bottling metallic vapors and then passing an electric charge through them was the basis of the sodium-lamp invention in 1917. In the new lamp mercury in an inner tube is vaporized by an electric arc and the pressure kept relatively high through the maintenance of a high operating temperature. The resultant light is distinctly bluish-white in color instead of the bluish-green of low-pressure design."

"The popular idea that snakes move with extreme rapidity was put to test and found wanting in snake speed trials made and reported before a meeting of the American Association for the Advancement of Science by Dr. Walter Mosauer of the University of California at Los Angeles. A man can walk faster than the fleetest snake, upon which Dr. Mosauer held a stop watch, can glide, for the maximum velocity did not exceed three and one-third miles per hour, approximately equivalent to 67 seconds for the 100-meter dash. Of seven typical North American snakes tested, the red racer was speediest, while the California boa, with a rate of only a quarter-mile per hour, was slowest of all."

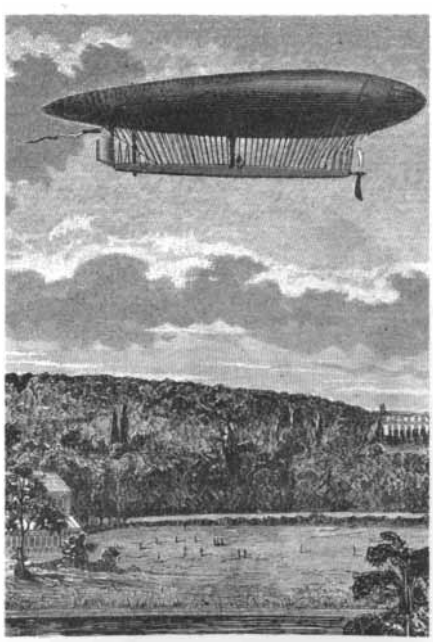
"The fact that a 75,000,000-dollar tree-planting project in the middle west has been authorized, and that preparations for realization are now going forward, lends added proof to the short-sightedness and selfishness of man in the development of the United States. Where once stood mighty forests of virgin timber will now be found vast stretches of denuded territory, swept by burning winds in the summer and stripped of fertile top-soil by the action of rains that run off unimpeded. This is the work of man, who ruthlessly cut down forests for the lumber, without thought for the future or for the effects which might arise from logging operations on an uncontrolled scale. Now the harvest has been reaped, and at this late date something is to be done about a situation that has become so acute as to constitute a national menace."

OCTOBER, 1884: "The project for a metropolitan railway submitted by the State for the examination of the General Council of Bridges and Roadways, as well as for that of the Municipal Council of Paris, has been definitely adopted by the Government, and declared of general interest. According to the proposed scheme, the railway will be subterranean for the greater part of its length. The city will gratuitously concede the subsoil of the streets, and it is owing to this that the cost per kilometer will be reduced to the expense of constructing the long tunnel and the two tracks. The passenger will descend to a depth of about 8 meters, and will travel under the earth, just as if he were in the Saint Gothard Tunnel."

"The telegraph appears to have made more progress in the United States than in any other country. The number of American telegraph offices in 1882 was 12,917, and the number of telegrams forwarded during the year was 40,581,-177. The number of telegraph offices in Great Britain and Ireland was 5,747, the number of telegrams forwarded being 32,965,029. Germany had 10,803 offices, the number of telegrams forwarded being 18,362,173."

"Experiments in telephoning to long distances have been made by the engineers of the International Bell Telephone Company between St. Petersburg and Bologoe (about 3,700 kilometers = 2,465 miles). Conversation could be kept up notwithstanding a rather high induction. It is but right to add that the experiments were made during the night, when the telegraph lines were not at work."

"M. Hervé Mangon has communicated to the French Academy of Sciences a report in which he states that a navigable balloon has at length been perfected by a captain of engineers named Charles Renard. The difficulty was to obtain a motive force in the car of the balloon, the apparatus of which should not be too ponderous for the sustaining power of the balloon itself. Captain Renard discarded the idea of a steam engine, and found the dynamic agent which he sought in electricity, with an apparatus of accumulators, by the force stored in which an engine of ten-horse power could be propelled during several hours. Under these conditions an ascent was made at a recent date. The balloon rose from Meudon and proceeded to Villebon, when, to the astonishment of those watching its progress, it described a semicircle and returned, notwithstanding the apparent opposition of a slight breeze, to the place whence it came. The trial was repeated, with similar results, the aeronaut subsequently declaring that the points where the balloon should halt, and return to its place of departure, had been fixed upon with precision beforehand."



The navigable balloon near Paris

Cracking

Summary:

Microelectronics is a major force behind the information age. It permits faster information processing through packing more and more components on a chip. GTE scientists have developed methods of producing advanced VLSI chips with 1.2μm feature size. The next stage is 0.8μm, and further reductions are on the way.

The ability of today's microelectronics to process information lags behind industry's need to transmit it.

GTE is working on this problem both by increasing data-handling capability, and by reducing data-processing time.

This is being achieved with VLSI (Very Large-Scale Integration) system densities approaching a million components per quarter-inch square, with reaction times in subnanoseconds, and with computer-aided design.

GTE scientists are developing advanced compiling systems for the full hierarchical design of more and more complex integrated circuits. They have also developed gate-level and functional-level circuit simulators that were five to ten times faster than previous systems.

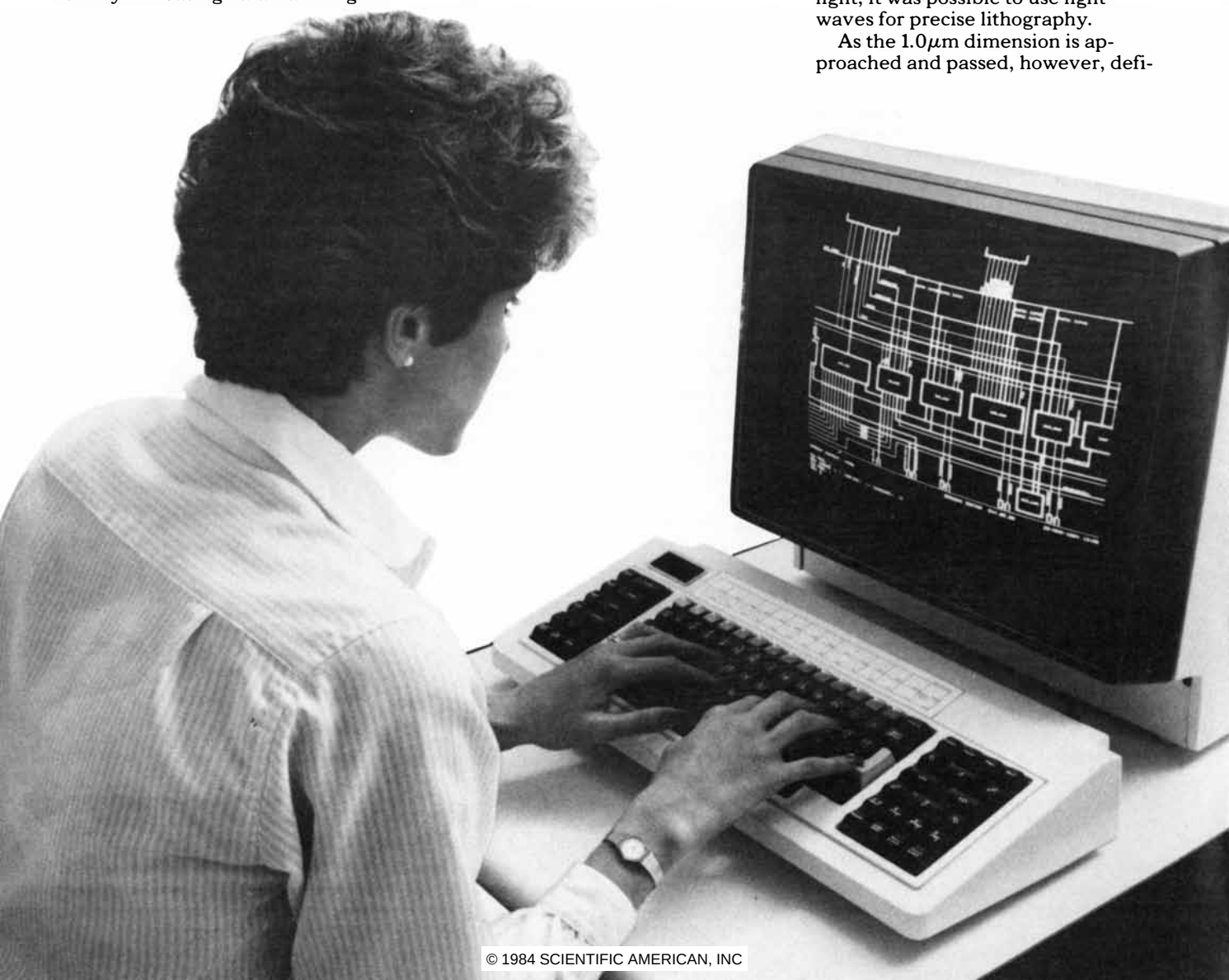
Making the chip.

The computer is also put to work in the chip manufacturing process. It directs the lithography, level by level, as well as other processes such as selective etching of deposited materials by ionized gases.

Currently, we are completing pilot-plant studies of a 1.2μm process and will transfer it to production facilities.

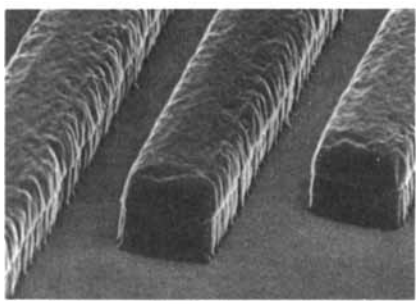
But feature dimensions continue to shrink. When they were comfortably above the wave length of visible light, it was possible to use light waves for precise lithography.

As the 1.0μm dimension is approached and passed, however, defi-



the 1 μm barrier.

nition begins to blur, and other techniques are needed. Among these is electron-beam lithography, with sub-micron resolution.



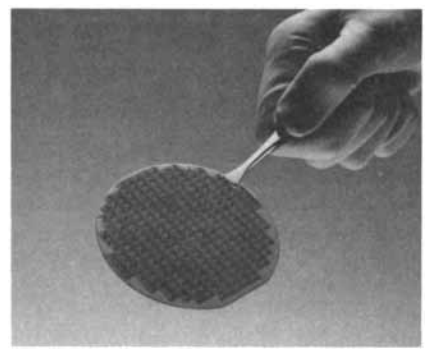
We are now working in the 0.8μm dimension. And we have identified experimental devices of 0.5μm and below as our next targets. (This dimension range begins to approach the distance an electron travels in solids before it scatters. By the time it has a collision, it has performed its work. Imagine the speed and precision this signifies.)

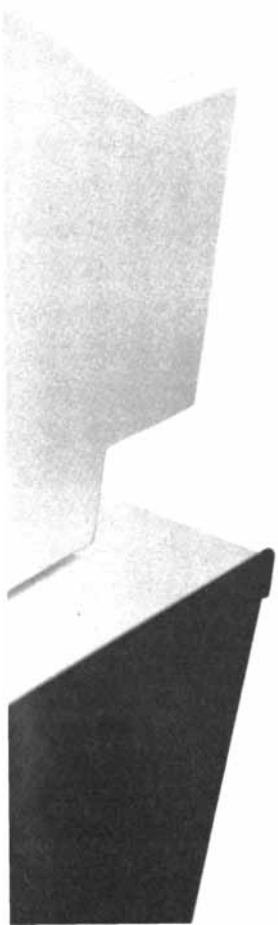
VLSI tomorrow.

Shrinking feature size to submicron dimensions is only part of the story. Other ongoing work in our laboratories includes replacing silicon as the chip matrix with faster-acting gallium arsenide, and building up the chip with epitaxial-film deposition, layer by layer.

Such VLSI research is helping accelerate the evolution of more sophisticated microprocessors, "burst" switching, PABX-on-a-chip, high-speed digital technology, and more.

In the box at the right is a partial list of pertinent papers by GTE people on VLSI and related subjects. For any of these, you are invited to write GTE Marketing Services Center, Department TPIA, 70 Empire Drive, West Seneca, NY 14224.





Pertinent Papers.

End Point Detection for Reactive Ion Etching of Aluminum, J. Electrochem. Soc., 1984.

Highly Selective Dry Etching of Polysilicon Using Chlorinated Gas Mixtures for VLSI Applications, Electrochemical Society Meeting, May 6-11, 1984.

Negative Resistance Switching in Near-Perfect Crystalline Silicon Film Resistors, 30th American Vacuum Society Symposium, November, 1983.

Vertical, Dual-Gate CMOS NAND in Two Laser-Recrystallized Silicon Layers over Oxidized Silicon Substrate, Materials Research Society Spring Meeting, February, 1984.

Reactive Sputter Etching of Single Crystalline Silicon, Proceedings of 3rd Annual Symposium on Plasma Processing and Extended Abstracts 83-1, 163 Electrochemical Society Meeting, May, 1983.

Reactive-Ion Etching of Single Crystalline Silicon with Cl_2 + $SiCl_4$, Proceedings of Fourth Conference on Plasma Processing, Electrochemical Society Meeting, May, 1983.

GTE

THE AUTHORS

HANS A. BETHE, RICHARD L. GARWIN, KURT GOTTFRIED and HENRY W. KENDALL ("Space-based Ballistic Missile Defense") are respectively professor emeritus of physics at Cornell University, IBM Fellow at the Thomas J. Watson Research Center of the International Business Machines Corporation, professor of physics at Cornell and professor of physics at the Massachusetts Institute of Technology. Bethe, who has taught at Cornell since 1935, was chief of the theoretical-physics division at the Los Alamos Scientific Laboratory from 1943 to 1946. He has served as a consultant to many private and Government organizations concerned with nuclear arms and energy, including the President's Science Advisory Committee. In 1967 he was awarded the Nobel prize in physics. Garwin is a graduate of the Case Institute of Technology and the University of Chicago, where he received his Ph.D. in physics in 1949. He has taught at Chicago, Cornell and Columbia University, has been a consultant to the Department of Defense and Los Alamos and has served on the President's Science Advisory Committee. Gottfried, born in Vienna, emigrated to Canada in 1939 and studied at McGill University, where he got a bachelor's degree in 1951 and a master's in 1953. He earned his doctorate at M.I.T., then taught at Harvard University before moving to Cornell. Kendall was educated at Amherst College and M.I.T., where he got his Ph.D. in 1955. He remained at M.I.T. through 1956 as a National Science Foundation Fellow, then took a position in the High Energy Laboratory at Stanford University. In 1961 he joined the M.I.T. faculty.

STANLEY B. PRUSINER ("Prions") is professor of neurology at the School of Medicine of the University of California at San Francisco, where he is also a lecturer in biochemistry and biophysics; concurrently he teaches virology at the University of California at Berkeley. He studied at the University of Pennsylvania, where he was granted an A.B. in 1964 and an M.D. in 1968. In 1972, after four years of biochemistry research at the National Heart and Lung Institute, he went to U.C.S.F. as a resident in neurology. He joined the faculty there in 1974. Prusiner was the recipient of a Research Career Development Award from the National Institutes of Health and the U.S. Public Health Service for 1975–76.

DON L. ANDERSON and ADAM M. DZIEWONSKI ("Seismic Tomography") are geophysicists who for the past year have taught together at the California Institute of Technology. Anderson is a professor there and director of Caltech's Seismological Laboratory; Dziewonski, who is chairman of the department of geological sciences at Harvard University, spent the academic year at Caltech as a Fairchild Distinguished Scholar. Anderson earned his B.S. degree from the Rensselaer Polytechnic Institute in 1955, then worked as an exploration seismologist for the Chevron Oil Company. In 1956 he joined an Air Force geophysical research team and led six expeditions to Greenland to study the elastic properties of sea ice; this research later proved relevant to his study of the earth's mantle. He went to Caltech for his master's (1959) and his doctorate (1962) and subsequently joined the faculty. Dziewonski was educated in Poland; he earned an M.S. at the University of Warsaw in 1960 and a Doctor of Technical Sciences degree from the Academy of Mines and Metallurgy in Cracow in 1965. He came to the U.S. in 1965 as a postdoctoral fellow at the Southwest Center for Advanced Studies (now the University of Texas at Dallas). In 1969 he became a member of the faculty there, and in 1972 he accepted a position at Harvard.

ARNOLD I. CAPLAN ("Cartilage") is professor of biology and codirector of the Developmental Biology Center and the Cellular and Molecular Biology Training Program at Case Western Reserve University. He is a graduate of the Illinois Institute of Technology, where he received a B.S. in chemistry in 1963, and of the Johns Hopkins University School of Medicine, which awarded him a Ph.D. in physiological chemistry in 1966. His doctoral research concerned the structure and function of the inner and outer mitochondrial membranes, a subject he pursued in postdoctoral research at Brandeis University. There he became interested in the biochemistry of muscle, cartilage and bone development, the focus of much of his work since then. Caplan joined the faculty of Case Western Reserve in 1969.

MARGHERITA HACK ("Epsilon Aurigae") is director of the Astronomical Observatory of Trieste. She was granted a Ph.D. in physics from the University of Florence in 1945; at about the same time she scored athletic triumphs in the Italian high-jump and long-jump championships. In 1947 she joined the Astrophysical Observatory of Arcetri in Florence and in 1954 she moved to the Astronomical Observatory of Brera at Milan-Merate. She took up her current position in 1964. Hack has collaborated on two volumes about

stellar spectroscopy and has written textbooks on general and radio astronomy; she is also director of the Italian popular magazine *L'Astronomia*.

TOM D. DILLEHAY ("A Late Ice-Age Settlement in Southern Chile") teaches in the department of anthropology at the University of Kentucky as an associate professor and is director of the Kentucky Anthropological Research Facility. His doctorate, which he got in 1976, is from the University of Texas at Austin. Beginning in 1972 he lived in Chile and Peru, doing archaeological research on the political and economic systems of pre-Hispanic societies of the Andes. While he was in South America he held several academic positions: professor at the Pontificia Universidad Católica de Chile in Santiago-Temuco (1975), Fulbright professor at the Universidad Nacional Mayor de San Marcos in Lima (1976) and director of the department of history and anthropology at the Universidad Austral de Chile in Valdivia (1977–79). In 1979 Dillehay returned to the U.S. and took his present job, but he travels to South America every year to do field work.

A. I. TSELIKOV ("The Continuous Processing of Metals in the U.S.S.R.") is a research metallurgist, a designer of steel plants and rolling mills and an inventor, with 140 innovations to his name. He currently is director of the All-Union Research and Design Institute of Metallurgical Machinery in Moscow. He was graduated from the Moscow Technical College in 1928 and has worked as a designer at the Bureau of Metallurgical and Thermal Constructions and at the Izhevsky Iron and Steel Works. Since 1935 he has taught at a number of colleges and universities, first as a lecturer and from 1943 on as a professor. Tselikov was elected a member of the U.S.S.R. Academy of Sciences in 1953 and has received the Lenin and State prizes of the Soviet Union.

FOLKE T. KIHLSTEDT ("The Crystal Palace") is associate professor in the department of art at Franklin and Marshall College. He got his bachelor's degree at Dartmouth College (1962), his master's at the University of Pennsylvania (1967) and his doctorate from Northwestern University (1973). His dissertation examined structural innovations in American exposition architecture from 1901 through 1939. His first teaching post was in the department of art history at the University of Cincinnati; with an interruption for his Ph.D. work he taught there from 1967 until 1974, when he moved to Franklin and Marshall. Kihlstedt is currently working on a book about the relation between architecture and the automobile, to be called *The Wheels of Modernism*.

Announcing

THE NATIONAL HISTORICAL SOCIETY

CIVIL WAR CHESS SET

Richly detailed portrait sculptures of great American heroes
—in solid pewter, solid brass and fine enamels.
A heirloom chess set to be enjoyed for generations.
Created by the world-famous craftsmen of The Franklin Mint.

THE NATIONAL HISTORICAL SOCIETY is dedicated to bringing the excitement and power of American history—as well as its significance—to people in every part of the land.

It is in keeping with this purpose that the Society is about to issue its own Civil War Chess Set. A dramatic tribute to the heroes of both North and South—and a work all the more intriguing because the playing pieces include richly detailed three-dimensional *portrait sculptures* of the great Generals of Union and Confederacy, captured for the ages in solid pewter, solid brass and fine enamels.

This extraordinary new chess set will be crafted to the highest standards of quality and historical authenticity. The National Historical Society has appointed The Franklin Mint to create the sculptures, each of which will be a new and original design. Some figures will be shown standing, some seated, some kneeling, some mounted on horseback. And each figure will be painstakingly crafted of solid pewter, hand-finished, then set atop a solid brass pedestal base embellished with a circular band of richly colored enamel—*blue* for the soldiers of the North, *gray* for those of the South.

Every sculpture, moreover, will be so rich with authentic detail that only the artists and master craftsmen of The Franklin Mint, steeped as they are in the tradition of *precision coinage,* could have achieved it. Indeed, every nuance of facial expression, uniform and weaponry—right down to the buttons, braiding, sabers and carbines—will be depicted with meticulous accuracy.

Thus, The National Historical Society Civil War Chess Set is also a magnificent *collection.* A triumphant achievement of portrait sculpture—and the ultimate in micro-detailed miniaturization.

ALL FIGURES SHOWN ACTUAL SIZE.

Major General
William Tecumseh Sherman
BISHOP

General in Chief
Ulysses S. Grant
KING

General in Chief
Robert E. Lee
KING

Major General
J.E.B. Stuart
KNIGHT

Available only by direct subscription. Issue Price: $17.50 per sculptured chess piece.
Limit: One complete set per subscriber. Please enter your subscription by October 31, 1984.

This handsome pewter-finished chessboard and fitted presentation case will be provided as part of the set.

A dramatic showpiece for your home or office

The chessmen themselves are scaled so that each one will suit the function assigned to it in the game of chess. And the handsomely crafted, pewter-finished playing board has been sized with equal care. Specially fitted, to also serve as the cover for the case which will house all 32 playing pieces, the board completes a presentation so attractive that the chess set will be played and displayed with pride and satisfaction. A Certificate of Authenticity, and specially written reference materials, will also be provided.

Exhibited on a table or cabinet in your living room, family room, den or office, this is a possession certain to evoke both admiration and respect from all who see it. A unique tribute to unique Americans. A work of heirloom quality, that will bring you endless pleasure through the years. And a chess set eminently worthy of being passed on from generation to generation.

The subscription rolls are now open. The work may be obtained *only* by direct subscription, with a limit of one complete set per subscriber.

The chessmen will be issued to you at the attractive price of $17.50 each, with the specially designed playing board and protective case provided at no additional charge. As a subscriber, you will receive two sculptured pieces every *other* month. You will, however, be billed for only *one* chessman at a time—a total of just $17.50 per month. In addition, you will have the option to complete your set earlier, if you wish—but you will be under no obligation to do so.

Here, then, is a work that will bring lasting pleasure to chess enthusiasts, history buffs, collectors of military miniatures—to anyone who appreciates our nation's heritage. Indeed, it is an unmistakably American chess set, that will make a dramatic addition to any room. And an exciting showpiece that will be displayed, enjoyed and treasured by each succeeding generation.

To acquire The National Historical Society Civil War Chess Set, no advance payment is required. But please note that the accompanying Subscription Application is dated and should be returned postmarked by October 31, 1984.

COMPUTER RECREATIONS

A computational garden sprouting anagrams, pangrams and few weeds

by Yank D. Weed

As the alias above suggests, I cultivate the occasional anagram and in so doing must reject dozens of weeds before one flower appears. An anagram is just a word or group of words obtained by rearranging the letters in another word or group of words. Some practitioners of this difficult but occasionally rewarding art maintain that the composition of ANAGRAMS amounts to an ARS MAGNA. But others, not so gifted with the necessary combinational instinct, shy away from the practice after a few halfhearted attempts.

Anagrammatic wordplay dates back at least to the 17th century. It was a literary pastime in the court of Louis XIII, who even appointed his own royal anagrammatist. The art continues to flourish in this century, and two of my predecessors, Martin Gardner and Douglas R. Hofstadter, have written on the subject [see "Bibliography," page 154].

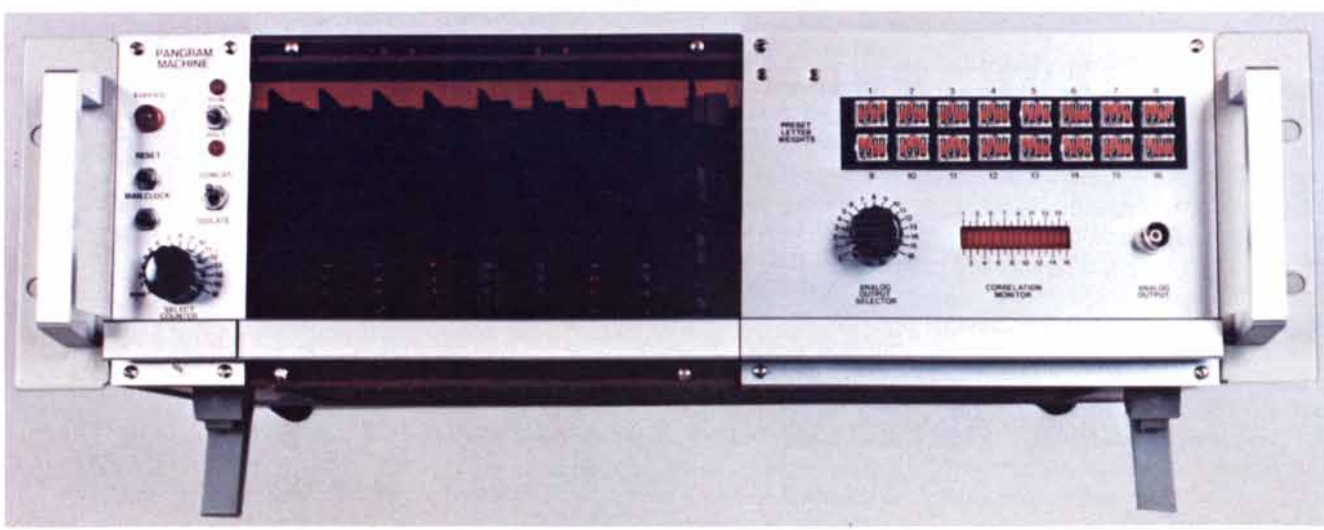
Those who rely on native skill alone, however, may find their talents challenged by the arrival of new, automatic forms of wordplay: Jon L. Bentley of AT&T Bell Laboratories has summarized the state of the art in single-word anagram-finding programs, and James A. Woods of the National Aeronautics and Space Administration's Ames Research Center has developed a program that generates anagrams of entire sentences. Finally, Lee Sallows, a British engineer at the Catholic University of Nijmegen in the Netherlands, has constructed a machine that hunts for pangrams, sentences describing themselves in terms of the number of letters they contain [*see photograph below*].

In preparing this topic, I experimented with my own name for a while. Using only the family name and first two initials, what new names could I make for myself? "Wayne Kedd" had a fine, decisive ring to it. "Eddy Kanew" suggested someone paddling through the Canadian North; "A. K. Dewdney" obviously showed very little imagination. Could I form suggestive phrases or even sentences from my name? "Dandy week" was a pleasant surprise but I was shocked by "Dewy naked." Throughout this exercise there remained a suspicion that I was missing a real flower or two. How could I be sure I had all valid word combinations? Not counting blanks, the nine letters of my name have more than 300,000 arrangements or permutations!

The simplest kind of anagram involves single words. Given a word, find another word that uses the same letters but in a different order. Many English words are anagrams of other English words. For example, *stop, tops, post, spot, pots, opts* are all mutual anagrams. When a human being searches for anagrams of this simple kind, new arrangements of the letters are perhaps suggested by subgroups within the word. For example, I found the anagrams of *stop* in exactly the order listed above, and it may be that when I looked at *stop,* I spotted *top* and simply moved the *s* to the end of the word. The remaining words all seem to have been obtained by moving one or two letters at a time in this way.

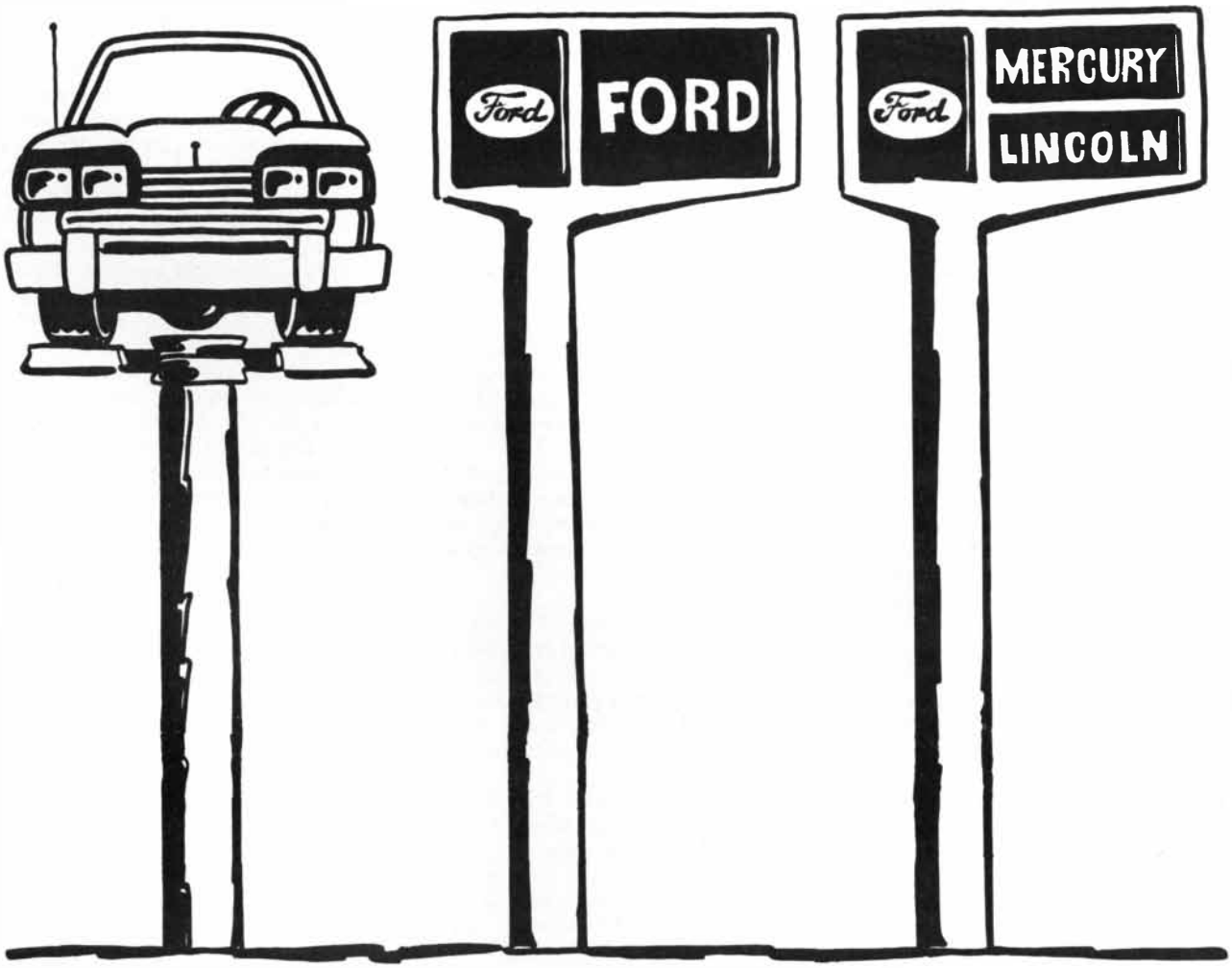
A computer program, however, is not equipped with a visual cortex and an associative memory. How does one write a program to first scan a word and then generate anagrams of it? A naive program might develop all permutations of the input and then reject all nonwords. Since there is no known purely computational test to distinguish words from nonwords, the program must have access to a dictionary stored in its memory. Even so, generating all permutations is surely wasteful because an enormous number of them must be compared individually with all the entries in an equally enormous dictionary.

Bentley, who writes "Programming Pearls," a regular feature in *Communications of the ACM* (Association for Computing Machinery), devoted a column last year to what he called "Aha! Algorithms." One of these, a refinement of known techniques, was an algorithm for generating anagrams. It qualifies as an Aha! (denoting the onset of insight) because of a very clever use of "signatures." The algorithm first computes signatures by copying each word in the computer's dictionary and rearranging the letters of the copy in alphabetical order. Next it sorts the words of the dic-



Lee Sallows' pangram machine

Car repairs guaranteed for life.



Only at these signs!

The Lifetime Service Guarantee.
Only Ford and Lincoln-Mercury dealers offer it.

Only Ford, Lincoln, and Mercury owners can get it.

The Lifetime Service Guarantee is so extensive, it covers virtually every repair on your car or light truck, for as long as you own it. Period.

And it doesn't matter whether you bought it new, or used. Or who you bought it from.

No other dealer, no other repair service — foreign or domestic — offers this kind of security.

Free Parts. Free Labor.
Now, when you pay for a covered repair once, you never have to pay for the same repair again. Ever.

The dealer who did the work will fix it free. Free parts. Free labor. For life.

But here's the best part: the Lifetime Service Guarantee is itself free.

A limited warranty on thousands of parts in normal use.

Things not covered are routine maintenance parts, belts, hoses, sheet metal and upholstery.

For Life.
How can we guarantee repairs for life? Because we're confident of the quality of our workmanship and the parts we use.

The Lifetime Service Guarantee. See your participating Ford or Lincoln-Mercury dealer.

LIFETIME SERVICE GUARANTEE

WE FIX CARS FOR KEEPS.

tionary in the order of the signatures. When this has been done, all words with the same signature will be stored together [*see illustration below*]. It is now a simple matter, when one is given a word as input, to generate its signature, look up all the words having that signature and print them out. The looking up is done by binary search, one of the oldest and fastest techniques for retrieving information from ordered data [*see illustration on page 24*]. At the end of his article Bentley lists the algorithm in program form.

As one thinks about it the realization comes that the programs described by Bentley solve not only individual-word anagram problems but also all possible anagram problems: once the dictionary has been sorted and the signature table has been generated, all single-word anagram play has been reduced to simple table lookups. Can anagrams ever be quite as much fun? The answer depends on one's philosophy of creativity. Do we create in the hope that no other person (or machine) can match our performance or do we create simply for the joy of personal discovery?

For wordplayers who create for the first reason, there are always multiple-word anagrams to fall back on:

MULTIPLE-WORD ANAGRAMS
PLAGUE RAW MORTAL MINDS

Apparently, however, even multiple-word anagrams pose no great problems for a correctly programmed computer. But we shall see how mortal minds still have a role to play in this case.

In his spare time Woods, a computer scientist, has developed a program for multiple-word anagrams. He was bitten by the anagram bug in 1983, when he decided to enter a biweekly anagram contest sponsored by *BAM* (Bay Area Music) *Magazine*. His entries, however, were to be computer-generated, and with an early version of his anagram program running he converted BACK ON THE CHAIN GANG into AHA, COGNAC KNIGHT BANE, which won an honorable mention. After a number of improvements to the program, including an Aha! idea of his own, Woods was ready to tackle some big names. For example, the name of Donald E. Knuth, a well-known computer scientist with a strong recreational bent of his own, was transformed in a number of ways:

(DONALD ERVIN KNUTH)
HUNT DRINK AND LOVE
INVENT HODAD KNURL
HALT UNKIND VENDOR

Readers may remember the famous numerologist (and inquiring skeptic) Dr. Matrix, who once stalked these pages with his daughter Iva:

(IRVING JOSHUA MATRIX)
HA—OUR JIVING MARXIST
HIS VAT, OUR MIXING JAR
I SAVOR RUM. I JIG. THANX.

It is generally considered permissible to add punctuation to multiple-word anagrams. Whether words such as THANX are acceptable or not depends entirely on one's dictionary.

As a final example of output from Woods's anagram program perhaps it is only fair that I include the following refractions of my own name (Alexander Keewatin Dewdney):

Al wandered—weekend anxiety
dexedrine wakened late yawn
Dean, a twinkle-eyed exwarden
dead wine and watery Kleenex
Ted Kennedy exiled; a war anew
Andean needed wax triweekly

The dictionary used by Woods is largely a customized affair. Since his program cannot live without it, the dictionary is stored in the computer as a disk file. It began its existence as a Unix System 5 Standard Dictionary (distributed by AT&T Bell Laboratories) of roughly 30,000 words, but Woods tripled its size by scanning various disk files and adding new words whenever they were encountered. Of course, the kind of dictionary discussed here is simply a gigantic list of words—with no definitions included.

As mentioned above, a multiple-word anagram program must be somewhat more agile than its single-word colleague. In addition more human intervention is called for at the output end because the words tend to tumble out in a random order. For example, one of the anagrams above may well have been printed out as MARXIST JIVING HA OUR; it was then up to Woods to find an arrangement of the words that made sense. It must be rather like sowing seeds to type in the words to be anagrammatized and then watch as dozens of potential flowers spring up; here is where the word-gardener comes into play. Some anagrams make almost no sense at all no matter how the words are arranged. These are weeds, to be pulled up immediately. Others might be rescued by some punctuation or by the invention of a little tale to go with them. But some anagrams can be arranged into perfect, comprehensible phrases or sentences such as HUNT DRINK AND LOVE. This is a flower to be preserved.

The Woods anagram program is a perfect example of the value of a good heuristic (an inexact but frequently useful procedure for getting an answer quickly). In outline, the program treats the set of input words merely as a string of characters. It cycles through the dictionary, testing each word there against the string. Are all the letters of the word in the string? If they are, put the word in a temporary list, subtract its letters from the string and start the dictionary search over again. Eventually either the letters in the input string are all used up or the program is left with an unmatchable set of letters. On page 26 the action of Woods's algorithm can be traced for the single-word input *compute*. The set of all possible successes and failures of the algorithm is laid out in a tree. Each node of the tree contains a word extracted from the residual list of the preceding node. Beside the extracted word (in parentheses) is the new, reduced residual list.

Because Woods's algorithm is written recursively, it automatically backtracks as soon as no more matches remain to be tried in the current residual list. If that list happens to be empty, however, it first prints out all the words on its temporary list—the sequence of words leading from the root of the tree to the current node. This sequence of words will be an anagram of the input words. Backtracking, moving forward, backtracking again, the algorithm eventually traces out every branch of the tree, in effect. In the case of the input phrase *compute*, nearly half of the branches yield (unweeded) anagrams. These would appear to the user as a printed sequence, which begins with

COP MUTE
CUP ME TO
CUP MOTE

Woods's algorithm is actually more sophisticated than I have so far led the reader to believe. For example, if it eventually tried every dictionary word contained in the input string *compute*, the tree would have nearly twice as many nodes sprouting from the string. The heuristic used by Woods to cut down the algorithm's word could be called a "rarest first" rule: from each residual string (as well as the input string) choose only those words containing the rarest letter in the string. A rare letter is one that tends to appear infrequently in words of

SIGNATURE	WORD
⋮	⋮
aecrs	acres
aecrs	cares
aecrs	races
aecrs	scare
aecrt	cater
aecrt	crate
aecrt	react
aecrt	trace
aecrv	carve
aecrv	crave
⋮	⋮

Part of a word-signature dictionary

The Square Rigger Overnight/Attaché

a Lands' End exclusive for the person who's all business 24 hours a day.

By day, this attaché is a well thought-out business accessory. A compartmented soft-luggage case that holds all the important papers, file folders, wedges of computer print-outs, you might want to bring with you. It also has pockets for the tools of your trade—pens, pencils, calculators, extra glasses. Even a clip on which to fasten elusive keys.

By night, however, this bag attends to the creature comforts you *can* take with you on those unexpected, here today, back tomorrow business trips. The overnight compartment is commodious enough to hold a change of clothes, a full kit (shaving or cosmetic), shoes, robe, slippers, accessories. Yet, its canvas divider enforces separation of the clean from the unclean, with no disrespect to either.

An unbeatable combination.
An unbelievable price.

Seriously, haven't you been waiting for this? We have. Which is why, flushed with the success of our Lands' End Square Rigger Attaché ($37.50), we created this bag—a triumph of mind over matter and quality over price. Would you believe all this for just $65.00? Well, believe it. $65.00 is the price.

As to quality, the material is 100% cotton duck—the kind they make deck chairs from. A Zepel® coating discourages soil and stain.

Padded handles, welt seams, and a two-word guarantee.

For those who are into details, the Square Rigger Overnight/Attaché does have padded handles, and seams that are four thicknesses strong. The zippers are our Lands' End oversize plastic ones, unbreakable and self-lubricating, with those big loop pulls you don't have to hunt to find. The bag also has an outside pouch for the magazines you snatch up on your rush to make your flights.

LANDS' END
DIRECT MERCHANTS

of fine wool and cotton sweaters, Oxford button-down shirts, traditional dress clothing, snow wear, deck wear, original Lands' End soft luggage and a multitude of other quality goods from around the world.

And like everything we offer you at Lands' End—from clothing to shoes to accessories—you're protected by a guarantee so ironclad, we state it in two blunt words:

GUARANTEED. PERIOD.

Send for our free catalog. Better still, call our toll-free number 24 hours a day —800-356-4444.

☐ Please send free catalog.

Lands' End Dept. Q-18
Dodgeville, WI 53595

Name_____

Address_____

City_____

State_____ Zip_____

Or call Toll-free:
800-356-4444

the dictionary. In the *compute* example the rarest letter in the input string is *p*. Consequently all the initial words selected from this string contain *p*. If the principle is applied at each node, failures tend to happen a lot sooner, the tree has fewer branches and the algorithm has less work to do. The rarest-first heuristic never misses an anagram since the rarest letter must eventually be used in a completely successful sequence of matches, and so why not match it first?

Besides using a variant of Bentley's signature method Woods's algorithm contains many other time- and space-saving ideas. Woods will be pleased to send a paper describing his algorithm to anyone who writes to him at the NASA Ames Research Center, Moffett Field, Calif. 94035.

Those who want to write down their own versions of the single- or multiple-word anagram programs described here will certainly have to get a dictionary in order to indulge in computer wordplay. One can also hook up to the Unix network and ask for Woods's dictionary through his computer mail address, which is *ames!jaw*.

Once a multiple-word anagram program is running with its own dictionary, the user may enjoy weeding his or her own garden of ready-made anagrams. Here there is still some scope for human creativity (neglecting for a moment the creative art of writing a good program), since the anagrams will sometimes come spilling out on the printer or screen faster than one can read them. The word *compute* alone resulted in 10 anagrams most of which I would immediately weed out as unpromising. Of course, one person's weed is someone else's flower, and while I might see some relation between *compute* and *up! comet,* others might prefer *mute cop* or *cut poem*.

For those who want a more curious, convoluted and (apparently) compulsive pastime, there are pangrams. Pangrams are sentences that contain each letter of the alphabet, such as the well-known sentence used to test typewriters, "The quick brown fox jumps over the lazy dog," or even those that contain each letter a definite number of times. The following example contains each consonant once and each vowel twice:

Why jog exquisite bulk, fond crazy vamp,
Daft buxom jonquil, zephyr's gawky vice?
Guy fed by work, quiz Jove's xanthic lamp—
Zow! Qualms by deja vu gyp fox-kin thrice.

Written by the 19th-century logological poet Edwin Fitzpatrick, these sentences are only a short leap away from the most advanced pangrams of all, the self-documenting pangrams that have intrigued Sallows for the past two years. For Sallows, "pangram" has meant precisely this kind of sentence:

This first pangram has five a's, one b, one c, two d's, twenty-nine e's, six f's, four g's, eight h's, twelve i's, one j, one k, three l's, two m's, nineteen n's, twelve o's, two p's, one q, eight r's, twenty-six s's, twenty t's, three u's, five v's, nine w's, three x's, four y's and one z.

What this sentence asserts about itself is true. For example, it has five *a*'s, four on the first line and one on the last line.

The change from Fitzpatrick's pangram to Sallows' is from a sentence *X* about which we can say "*X* has so many *a*'s, so many *b*'s, ... and so many *z*'s" to a sentence *X* that has this very form—and that is also true. Although logically this is a leap from other-reference to self-reference, a leap with which many readers are now familiar, it represents a much bigger leap in actual content: gone is the organic quality of "Jove's xanthic lamp," to be replaced by the analytic flavor of "... one *x*, two *y*'s and one *z*." If Fitzpatrick's pangram is a flower, then Sallows' pangram is perhaps more like a crystal.

Indeed, Sallows refers to his constructions as "crystalline." His interest in pangrams, although it is several years old, did not blossom until he spotted the following pangram in a Dutch newspaper, *Nieuwe Rotterdamse Courant,* in March, 1983:

Dit pangram bevat vijf a's, twee b's, twee c's, drie d's, zesenveertig e's, vijff's, vier g's, twee h's, vijftien i's, vier j's, een k, twee l's, twee m's, zeventien n's, een o, twee p's, een q, zeven r's, vierentwintig s's, zestien t's, een u, elf v's, acht w's, een x, een y en zes z's.

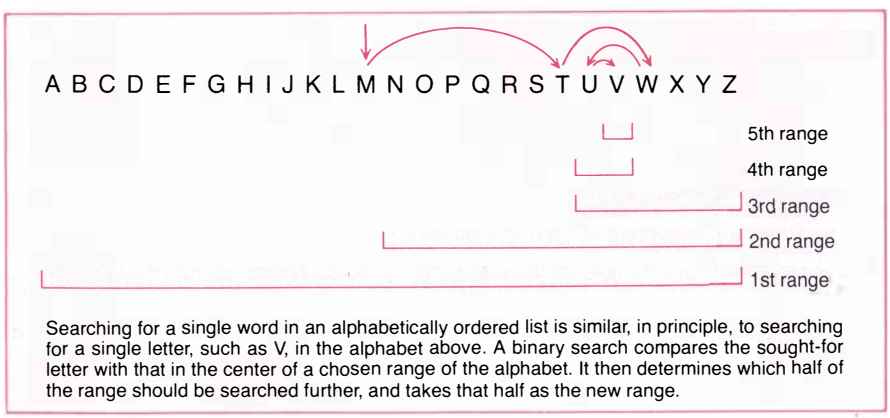
Stunned by the prismatic beauty of this elegant specimen, Sallows was at first envious and then dismayed to find that the author of the article, a well-known wordplay expert by the name of Rudy Kousbroek, had thrown out a challenge addressed to him personally: "Lee Sallows will doubtless find little difficulty in producing a magic English translation of this sentence."

Having already considered the possibility of computer-generated pangrams, Sallows set to work writing a succession of programs in Lisp. Analysis revealed that the search boiled down to simply filling in the question marks in the following pseudopangram:

This pangram contains five a's, one b, two c's, two d's, ? e's, ? f's, ? g's, ? h's, ? i's, one j, one k, ? l's, two m's, ? n's, ? o's, two p's, one q, ? r's, ? s's, ? t's, ? u's, ? v's, ? w's, ? x's, ? y's and one z.

Setting up predefined ranges for the numbers in question, Sallows' final Lisp program methodically began to work its way through all possibilities, testing 100 new combinations per second. The program ran as a low-priority "batch job" every night and each morning Sallows would rush eagerly to his terminal, call up the job file and swiftly scan it for the magic word EUREKA. This would indicate that the program had discovered a pangram. But morning after morning no EUREKA appeared and Sallows began seriously to consider just how long he ought to expect the job to take. A few quick calculations with the predefined number ranges soon revealed a time of 31.7 million years. He writes of this moment: "I was so unprepared for the blow contained in this revelation that initially I could hardly take it in.... Now that the truth had dawned I began cursing my naiveté in ever embarking on such a fool's errand."

At this point lesser mortals might have given up entirely, but the experience seems only to have steeled Sallows' resolve. Others had urged him to develop a cleverer program, yet Sallows, an electronics engineer and not a computer scientist, felt uncomfortable in the realm of algorithmic analysis. Only one approach seemed to make sense: a special-purpose computer dedicated to the search for pangrams. In other words, a pangram machine!



Searching for a single word in an alphabetically ordered list is similar, in principle, to searching for a single letter, such as V, in the alphabet above. A binary search compares the sought-for letter with that in the center of a chosen range of the alphabet. It then determines which half of the range should be searched further, and takes that half as the new range.

A binary search

Over a period of three months Sallows "devoted every spare second to constructing this rocket for exploring the far reaches of logological space." When it was finished, it contained 100 integrated circuits distributed over 13 printed-circuit cards. Its front panel displayed 67 indicator lamps for reading the current combination under test and a special EUREKA light that would turn on only when a pangram had been found. The extremely fast circuits of Sallows' machine could explore one million combinations per second instead of his program's 100. By reducing the predefined ranges for the number words filling in the question marks of his pseudopangram, Sallows developed a new estimate of the time necessary for trying all possible combinations: 32.6 days.

The pangram machine was launched on October 3, 1983. For the first few days Sallows would wake up in the morning dominated by a single question: Has it halted? He writes: "It took nerves of iron to go patiently through the morning's ablutions before sedately descending to the living room, where the machine was installed on my writing bureau. Then with great deliberation I would open the door, go in and *look*." But morning after morning the indicator lamps were found blinking their way merrily through millions of combinations while the EUREKA lamp stayed ominously dark. On the sudden inspiration of fresh insight, Sallows finally stopped the machine and made further modifications. By November 19 the newly improved MARK II pangram machine was up and running, and one evening two days later he "was sitting in front of the machine... when suddenly the EUREKA lamp came on and my stomach turned a somersault. Tense with excitement, I carefully decoded the light displays into the set of number words represented. A painstaking check completely verified the following perfect pangram":

This pangram contains four a's, one b, two c's, one d, thirty e's, six f's, five g's, seven h's, eleven i's, one j, one k, two l's, two m's, eighteen n's, fifteen o's, two p's, one q, five r's, twenty-seven s's, eighteen t's, two u's, seven v's, eight w's, two x's, three y's & one z.

In the days that followed more pangrams appeared, and Sallows' collection now amounts to hundreds, including a series of 30 pangrams with 30 different verbs such as "contains," "lists," "includes" and so on. Exulting in his triumph, Sallows appears ready to make great claims for the kind of special-purpose analog-digital devices exemplified by his pangram machine: "This apparatus has succeeded in quickly producing... solutions to an essentially mathematical problem in the face of which digital computers (I exclude supercomputers and parallel processors) are wholly ineffectual."

It seems to me there are several people in the computing community, particularly the academicians, who would take issue with Sallows' statement. As if to goad on those very people, Sallows has made the following wager: "I bet 10 guilders nobody can come up with a self-enumerating solution (or proof of its nonexistence) to the sentence beginning 'This computer-generated pangram contains... and...' *within the next 10 years*." Naturally, if someone accepts and meets this challenge, the results will be reported in a future column of this department.

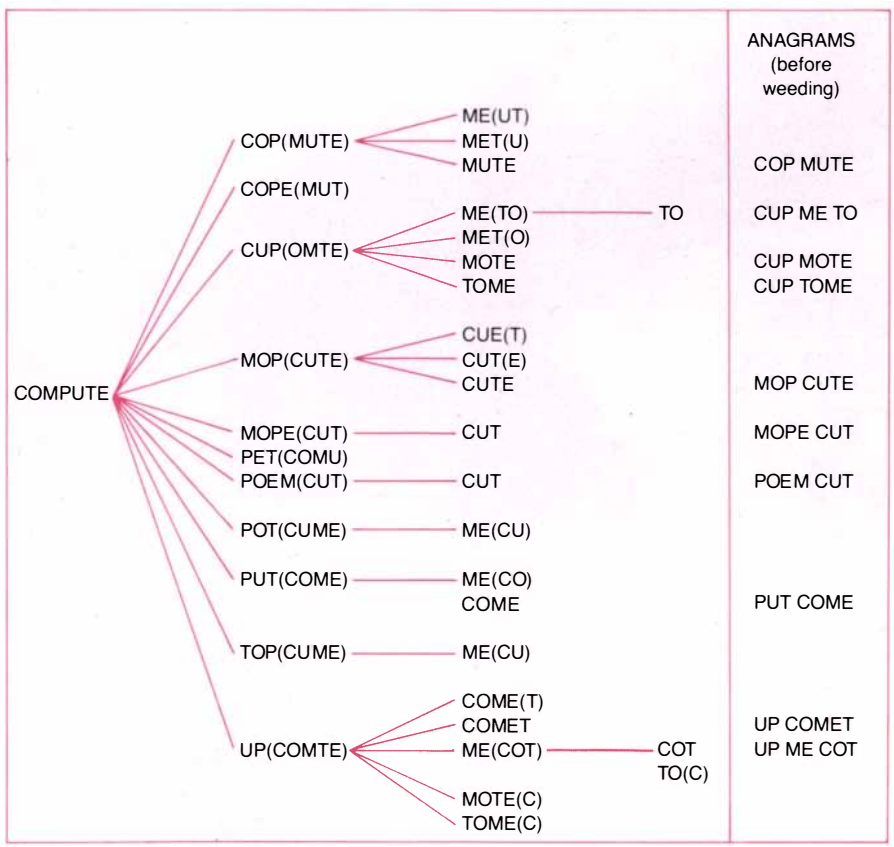
So far I have said almost nothing about the pangram machine or how it works. Although a detailed description of its electronic design is beyond our scope, the machine can be discussed in terms of the algorithm it embodies. Essentially a brute-force search through a large number of possibilities, the algorithm nonetheless uses some intelligent structures.

Constructing a pangram, as noted above, is really just a matter of filling in some 16 blanks representing the only letters whose frequency of occurrence is likely to change from one would-be pangram to the next. Each number word, say *twenty-seven,* can be represented by the following "profile," which lists the number of occurrences of each letter in the word:

e f g h i l n o r s t u v w x y
3 0 0 0 0 0 2 0 0 1 2 0 1 1 0 1

A particular combination of number words can be represented by a matrix such as the one in the illustration on the opposite page. The rows of the matrix represent the letters *e* through *y* in the list above. Each row contains the profile of a particular number word indicating how many times the row letter is said to appear in the current combination. Below the matrix is an additional row listing the frequency of additional letters in the pangram text (such as "This pangram contains..." or "This pangram lists...") and, below that, a sequence of column sums. In the illustration all but one of the column sums match their corresponding row numbers: the current combination is a near miss.

In essence Sallows' pangram machine "counts" its way through all possible combinations of number words, which are restricted to certain ranges. For example, one set of ranges used successfully by Sallows involved 23 to 32 *e*'s, one to 10 *f*'s, one to 10 *g*'s, one to 10 *h*'s, six to 15 *i*'s and so on. His chief concern in setting the ranges was to make them narrow enough to keep the search time small but not so narrow that all pangrams are accidentally weeded out. The



A Woods tree the leaves of which contain anagrams of the word **compute**

	E	F	G	H	I	L	N	O	R	S	T	U	V	W	X	Y	CURRENT COMBINATION
E	3						2			1	2		1	1		1	27
F					1					1					1		6
G	2			1					1		1						3
H	1	1			1								1				5
I	3					1	1						1				11
L								1			1			1			2
N	1						1				2			1		1	20
O	2	1					1	1	1		1	1					14
R					1					1					1		6
S	2		1	1	1		1				3			1		1	28
T	2				1		3				2			1		1	29
U	2			1					1		1						3
V					1					1					1		6
W	1						1				1						10
X		1						1	1			1					4
Y	1	1			1								1				5
	7	2	2	2	4	1	10	11	2	24	7	1	2	5	1	1	
	27	6	3	5	11	2	20	14	6	28	21	3	6	10	4	5	

A pangram matrix for a near miss, which contains not 29 T's, as stated, but 21

actual counting involves stepping each of the 16 row number words through their respective ranges rather like a crazy odometer. Thus the first set of numbers to be tried might be 23, 1, 1, 1, 6,..., the next 24, 1, 1, 1, 6,... and continuing to 32, 1, 1, 1, 6,.... The combination after this would involve resetting the first counter back to 23 and advancing the next one, as in 23, 2, 1, 1, 6,....

A more detailed account of Sallows' pangram adventure, including a description of his machine and its operation, is available from this paradigital pangram pioneer at Buurmansweg 30, 6525 RW Nijmegen, The Netherlands.

Besides the recreational value inherent in the various forms of automatic wordplay described here, there is sometimes an additional and deeper benefit in computer recreations, namely the development of intellectual skills. In his "Aha! Algorithms" column Bentley stressed the value of his anagram program as an example of how a little insight can go a long way toward making a program more efficient. Indeed, the same thing is true of Woods's program, and even of Sallows' machine. Whether a clever program running on a standard computer will ever outperform Sallows' machine is a bit beside this point of view. Sallows' effort represents simultaneously a funny kind of obsession and a tremendous drive to achieve results. Perhaps it is even true that his results were achievable in no other way under the circumstances.

The most interesting reaction to the column on programs that play checkers [July] came from Marion F. Tinsley, the current world checkers champion. Tinsley, a topologist at Florida A&M University, wrote: "On the basis of games that I have seen played by the Jensen-Truscott program, it would rank about 200th in the U.S." Indeed, the American Checkers Federation issued the following challenge to checkers programmers: Tinsley will play a 20-game match with a computer program for a stake of $5,000 a side, winner take all. During the five-year period of the challenge no one picked up the gauntlet. William B. Grandjean, secretary of the American Checkers Federation, told me recently that the challenge will be reissued at the slightest sign of interest from the computing community.

Readers wanting more information about organized checkers are encouraged to write to Grandjean at 3475 Belmont Avenue, Baton Rouge, La. 70808.

SCIENCE/SCOPE

NASA's Project Galileo Probe, which will explore the planet Jupiter later this decade, must arrive at a precise angle if it is to carry out its measurements of the chemical composition and physical state of the Jovian atmosphere. The Hughes Aircraft Company-built probe will arrive at 107,000 miles per hour. If the probe hits at too shallow an angle, it will skip off into space; too steep, it will be reduced to ashes. Even at the proper angle, the probe will encounter extremes never before faced by spacecraft. In less than two minutes, much of the forward heat shield will be eroded by temperatures of thousands of degrees. With atmospheric entry forces reaching 360 times the gravitational pull of Earth, the 742-pound probe will take on a weight equal to an empty DC-10 jetliner. Project Galileo is scheduled to be launched from the space shuttle in May 1986 and to arrive at Jupiter in August 1988.

Improvements to a "super cooler" used with infrared sensors in space will extend the life and boost the efficiency of the device. The cooler, vital to defense applications and geological surveys, is a Vuilleumier cycle cryogenic refrigerator. It is designed to chill sensors near absolute zero to increase their sensitivity to thermal radiation. These coolers are ideal for use in space because the low internal forces required by this kind of cooling cycle cause little wear on bearings and seals. Hughes is working under a U.S. Air Force contract to extend the unattended operating life of the cooler beyond five years. The cooler will use less power, so smaller and fewer batteries will be needed to power the device during eclipse periods—a savings of hundreds of pounds.

Significant improvements in infrared simulation technology loom with the development of a device that converts complex visible scenes into infrared images. The device would at first be used for testing missile seekers and other military systems. It converts visible images into infrared by means of a modified silicon liquid-crystal light valve. The Hughes device is being developed to be fully compatible with standard video rates and computer-image generation systems.

In the last 20 years, over $611 million in savings have been negotiated by Hughes and the Department of Defense as a result of engineering proposals for cutting costs of military systems. Since the inception of the Value Engineering program, Hughes has had 675 proposals accepted in 50 programs. The changes stemmed from advanced technology that was not available at the time the original contracts were signed. They resulted in substantial improvements in quality, reliability, producibility, and life-cycle costs. Savings amounted to 3% of Hughes sales during the period, with the U.S. government's share amounting to nearly $500 million. The Value Engineering program is designed to encourage employees to look at the functions of a product and develop alternatives that cost less, perform better, and improve reliability.

Hughes Missile Systems Group, located in Canoga Park, California, an attractive suburb of Los Angeles, is seeking engineers and scientists for such developmental and engineering programs as AMRAAM multimode guidance, Phoenix, and IR Maverick. Openings are in radar and electro-optical systems design, systems software and hardware/software integration, analog and digital circuits design, hybrid process engineering, systems performance, and microwave and power supply/transmitter design. Qualified applicants are assured prompt replies. Please send resume to Hughes Engineering Employment Manager, Dept. S2, Fallbrook at Roscoe, Canoga Park, CA 91304. Equal opportunity employer. U.S. citizenship required.

For more information write to: P.O. Box 11205, Marina del Rey, CA 90295

HUGHES
AIRCRAFT COMPANY

BOOKS

Large-scale irrigation, gauge theory, synthetic psychology, blue cornflakes

by Philip Morrison

LONG-DISTANCE WATER TRANSFER: A CHINESE CASE STUDY AND INTERNATIONAL EXPERIENCES, edited by Asit K. Biswas, Zuo Dakang, James E. Nickum and Liu Changming. Published for the United Nations University by Tycooly International Publishing Limited, distributed in the U.S. by UNIPUB, 205 East 42 Street, New York, N.Y. 10017 ($65; paperbound, $37.50). The goal is ancient and unchanging: more grain, more food on the table. The means too is no different from what it has been for thousands of years, a secure supply of irrigation water. The scale is big but not grandiose; the best-studied proposal would begin with a canal, resembling a newfound river with the flow of the broad Susquehanna, that would carry water from the Chang Jiang (Yangtze River) to the rain-short North China plain.

This unusual book gathers the views of Chinese and foreign experts on the wide issues raised by the project. About a fourth of the text treats recent international experience with large-scale water transfer; two dozen papers or more examine the concrete but tentative Chinese proposal in many aspects, from provincial rivalry to the local species of fingerlings that usually pass safely through spinning turbines. Actual engineering problems are only outlined: the canal route is well mapped, but no cross section of the canal is shown, nor even a rough budget struck to support the cost estimate. Impact and consequences hold central attention, discussed under the rubric *Shiwu zong shi yi fen wei er:* There are two sides to everything.

It is the Chang Jiang that offers the planners their big chance. That grand river carries more water down to the sea than any other river of temperate latitudes. Tropical Amazon and Congo, draining their rain forests, do deliver more each year; our Mississippi-Missouri ranks only seventh worldwide. Southward from the broad Chang Jiang there lies the warmer, wetter world of pond and paddy, where rice fills every bowl. But beyond the other bank it is the North China plain that stretches wide, from Nanjing north to Beijing. There Chinese civilization first flowered across the fertile lands that are the silty deposit of the restless and high-handed Huang He (Yellow River). That shifting stream now flows nearly parallel to its much greater counterpart, but 300 miles to its north. To this day the intensely cultivated northern plain, as big as three or four of our farm states, where little natural vegetation is to be found among the fields and orchards, is home to a fifth of China's billion people, those whom not rice grains but wheat flour sustains.

It has never been easy to take abundance from those plains. The staple winter wheat is planted in the fall, and after a frosty winter it ripens to harvest just before the rains return. Spring drought all too often takes toll of the greening crop; even worse, sudden summer rains can bring disastrous floods, and the main watercourses have shifted catastrophically half a dozen times during the long written record. Irrigation is widespread there now, from tube wells by the hundreds of thousands, as well as from reservoirs and riverine diversion. The Huang He with the Huai and the Hai are now largely controlled streams. Yet the entire varying and uncertain runoff across the northern plains adds up to an order of magnitude less water than the Chang Jiang predictably drains unused to the sea each year.

In one well-studied version the new Susquehanna of Henan would start from an existing reservoir linked to the middle reaches of the Chang Jiang, to head north toward Beijing along the higher contours just west of the plains. The trunk canal, 15 or 20 feet deep and 1,500 feet wide, would flow smoothly under gravity 600 miles northward, lessening as it flows, according to the backward habit of such riverlike artifacts. It would divert a 30th part of the flow of the great river, to add some 30 cubic kilometers a year to the drier plains, more than half of all that is now available to those tens of millions of northern fields. With the new resources, grain could be grown in places now too dry to yield wheat, yet safe from flood, and the spring drought loss in fields now irrigated could be averted as well.

The cost? In earth-moving, a dozen Panama Canals or so, a few times what has been emplaced during the past generation to dyke and channel the dangerous elevated course of the Yellow River. In yuan, 10 billions, an investment of 10 millions of working years. The engineering challenges include carrying the canal somehow across more than 150 rivers along its route, including the Yellow River itself.

There are a dozen other issues, from saltwater backup in the great estuary once its flow is lessened, to underground storage in old abandoned river channels, to the spread of town pollution, to the climate change induced by a new river. None of the environmental effects seems enough to bar the project. The real question is whether the new water will grow more grain in the long run than the same effort spent locally on improving farm practice field by field. A wise decision is not yet obvious.

What does experience say? The best-known large-scale irrigation project of our time is that of the Nile, a scheme based on the Soviet-engineered Aswan High Dam, which yields an annual water supply of almost 90 cubic kilometers. Mahmoud Abu-Zeid of the Ministry of Irrigation in Cairo is unequivocal: negative effects are "far outweighed by the benefits achieved." The summer cresting of the Nile granted only one crop a year; the new perennial irrigation enables double- or triple-cropping on most irrigated lands. No grain production figures are presented. The loss of Nile silt is inexpensively balanced by the fertilizer made from the dam's hydropower, at least for the nutritive nitrogen content. The loss in sardine catch at the river mouth from the absence of Nile silt is outweighed by the fisheries of the reservoir itself. The spread of the snail vector of schistosomiasis is controllable. There is real promise of considerable new land watered in part by the Nile flow and in part by the drainage from established fields. But all water brought to the land bears some dissolved salts. Too much water sinks in, raises the water table to free long-buried subsurface salts as well into the crop roots. The greatest problem in the Delta is waterlogging from overirrigation and consequent salinity of the soil, "increasing at an alarming rate over the past 10 years."

That is nothing new. The Indus too, where irrigation is also thousands of years old, reports the same critical state of affairs. There water is supplied by 100,000 pumped tube wells and 200,000 Persian wheels to supplement the large reservoir system. In both places the cure seems reachable with hard work: controlled water use and drainage systems in place are as important as supply. The deferment of carefully worked-out plans for on-the-farm control and drainage, explicit in most irrigation project reports, including that of Aswan—"critical drainage problems were not expect-

ed for 10 to 15 years"—is as naive as it is typical. There are plenty of things to do, from properly metering and charging for water to careful leveling of the land, to lining the small ditches with clay and gypsum and brick, to large-scale installation of tile and plastic pipes, even to planting brush to reduce evaporative losses. Above all, it is clear that imposing and well-engineered cross-country arterial canals are not sufficient for success in the long run; it is rather the capillary system out on the farms, among villagers by the million, that makes or breaks the grandest schemes. The users must be organized systematically to manage their own water; a pilot program at work along the Indus is currently well received by the farmers, its credibility "far exceeding any other government programme." Can that be extended to the more than 80,000 watercourses down the long Indus, each serving an average of 40 farmers? Can that same spirit act to control the waterlogging of the North China plain? Unless the answer is a clear if slowly emerging affirmative, all those cubic kilometers from the great river might as well enter the sea, and the North China plain will yield less grain than in the past.

One paper recounts the Texas Water Plan to replace ground-water depletion on the dry High Plains of West Texas by diversion from the major rivers on that state's eastern borders. The plan, first put forward in 1968, has become "a negative example of planning" in its perfunctory treatment of environmental risks. The new proposal, still fully on the Chinese scale, is balanced but long-range: it looks forward as far as 2020, to new population and new needs.

Vivid maxims can be cited on both sides. Is delay a case of "giving up eating for fear of choking," or is the whole scheme in fact a self-deluded effort to "support deficiency with deficiency"? There is no explicit conclusion; a reader of good will is likely to come away in the mood of cautious hope. A considerable effort ought to continue in planning and research, with needed field studies; a larger and many-sided campaign should be set afoot to improve the local use of water. If that goes well, in not too many years the people who must in the end tend the million farm outlets to be fed by the great stream will make their hopes manifest by results. Then the engineers and their legions can proceed with surety and speed, and this "magnificent project to transform nature" will have ripened from temptation into a national opportunity.

This readable book by the experts is one in an important series from the United Nations University. Its editor-in-chief is an Oxford-based consultant; it is international in production, phototypeset in Nairobi and printed in Shannon; it was published simultaneously both in English and in Chinese. Those who produced this diverse, direct and valuable book should have spent a little more effort on the design, legibility and scale of its many tables, maps and charts.

AN ELEMENTARY PRIMER FOR GAUGE THEORY, by K. Moriyasu. World Scientific Publishing Co., distributed in the U.S. by Heyden & Son, Inc., 247 South 41 Street, Philadelphia, Pa. 19104 ($28; paperbound, $14). Within the past decade gauge field theory has achieved a strength and success not paralleled since the glorious decade 50 years back when stars, atoms, metals and molecules first disclosed their mysteries under the probe of the newborn quantum mechanics. This new reign is not in fact a revolution like that of 1925; rather, it is that same regime's powerful sovereignty newly extended over the once wild frontier of particle physics. The present theory rests firmly on the concepts of quantum field theory as well as on that theory's relativistic classical basis; its substance is a much more subtle use of symmetry. Once again success has followed an appeal to new groups of transformations. The new symmetries, to be sure, are still more sophisticated and abstract, and even more surprising they are now honored more in the breach than in the invariance.

This brief book is an informal, physically based introduction to gauge theory as far as the tested unification of electromagnetic and weak forces, and beyond that to quarks and their colorful dynamics. It only lightly enters the past few speculative years of grand unification and stops short of higher space-time dimensionality. The text winkles out the foundation and meaning of the new equations, mainly by direct comparison with older and simpler theories, setting the discussion within their history.

Let only the somewhat mathematical enter here, those "with only a background in quantum mechanics." Satisfied readers need to be familiar with Schrödinger wave functions, with four-vectors and with partial differentiation, but not with quantum field theory itself. For them, mostly scientists and students reasonably at home with theoretical physics outside of particle physics, the book is a compact 150-page Baedeker to a knowing tour of the wide new landscape, bestarred with fresh views. The treatment runs to 10 chapters, whose order is as much chronological as logical, for the theory grew more or less as understanding came to the theorists.

The root idea is that of the geometric relations enforced on the inner nature of systems as they move from place to place in space-time. It was Hermann Weyl in 1919 who introduced the word gauge in this context (*Eichung,* he wrote) in his stillborn but powerful effort to generalize on Einstein's new general relativity. It was plain how Einstein had managed to build a geometric theory in which every observer locally saw a space-time obedient simply to special relativity. But the relative orientations of the local space-time axes employed by the observers changed smoothly over finite distances within the space-time, curved overall as gravitating matter determined. That change was rotational in nature; physical intervals in space-time could change in direction but not in length. Weyl sought to drop that restriction; in his theory quantities could change their scale—their gauge—as well. The connection between such quantities from place to place was for him given by a form that entered exactly as the potential enters in classical electromagnetic theory, a new part of the energy-momentum change during particle motion. It was soon realized, however, that real particles rigorously maintain charge and intrinsic mass as they move.

Weyl's theory was prematurely all but abandoned. Within the decade quantum theory showed how the phase of the wave function of a charged particle, although not its magnitude, can and does change with position within an electromagnetic field. These changes need to be balanced neatly by the phase effects of the electric potential if charge is to be constant. Of all Weyl's ideas only the word gauge and that single deep result about charge remained. In about 1960 Y. Aharonov and David Bohm showed that such phase changes were not merely present for theoretical consistency but were experimentally measurable. Experiment with interfering beams in elegant electron-optical setups soon confirmed their claim: relative phase depends on the potentials in space, even when no electron ever enters the field-containing region.

Some years before 1960 C. N. Yang and Robert Mills had generalized the familiar well-accepted residue of gauge theory in a daring way, long neglected in its turn. They were concerned with finding an analogue to an electromagnetic potential that could play its phase-affecting role within a description of the strong nuclear forces. They found it neatly enough. But it was no longer a simple quantity like the potential; rather, the connection required between the nature of a nucleon at two points demanded not just a simple phase change but operators of more complicated nature, multidimensional and generally order-dependent. For the internal degrees of freedom that relate neutron to proton in nuclear interactions bear a close analogy to rotations within three-dimensional space. Gone is the simple single angle, interpretable as a two-dimensional rotation about one fixed axis, that enters in describing electric charge.

In one of the most instructive of his chapters Dr. Moriyasu generalizes the

familiar physical derivation in electromagnetic theory, in which he carefully follows the wave function of a test charge as it is carried around a small chosen path in the field. Flux changes are computed in the usual first-order way, and the field relations can be found. Exactly the same trick can be carried through using the much more complicated phase changes of a particle in a Yang-Mills field. New terms appear, the result of the intrinsically distinct algebra. The results are shaking: the fields no longer add linearly; the current density is not a conserved quantity, for only the overall current of source and field together is conserved, not its local values. The fields themselves must carry particle current. The fields are in part their own sources; in gravitational theory the energy in the field possesses mass and is therefore itself the source of more gravitation.

The Yang-Mills version was built on as close an analogy to the Maxwell case as any theory describing the symmetry of nucleon interactions can be. The more complicated interactions between nucleons not only enter the energy-momentum balance, as electric potentials do for charged particles, but also act to mandate changes in particle nature as well. Here the choice of absolute phase can mark a particle either as proton or neutron. That seems far from arbitrary. A full physical understanding was still missing, but "the moral of this story" is that a theory with such algebraic structure (current jargon dubs it a non-Abelian gauge theory) "may contain entirely new types of objects never considered possible in physics before."

The modern theory arrived in the 1970's. It was built of new experimental results and of partial theoretical insights drawn from particle physics and from the quantum theory of condensed matter, especially that part concerning superconductivity. Finally came the brilliant success of the Weinberg-Salam unified theory, newly confirmed. The indispensable opening was the recognition that Yang-Mills might make sense if some way were provided to favor one particular choice of direction within the abstract internal space that fixes the nature of the nucleon: in short, some way to break the symmetry. It was no use simply to ignore the symmetry; the trick had to be compatible with the observed powerful equality among internucleon forces on which the Yang-Mills idea was based. The answer was and remains the introduction of yet another particle field. It could be a simple one, free of electric charge, free of spin, fixing solely the choice of nucleon type. Moreover, that choice, once made, had to hold at every space-time point in the same way.

The solution was given by Peter Higgs in the mid-1960's. The "Higgs particles"—unobserved, so far hypothetical,

known only by construction—simply determine the milieu in which all the other fields are embedded. They are only a background; so to speak, they provide a modern ether, at once quantum and relativistic. They are assigned by fiat a plausible and suitably complicated intrinsic potential energy, just right to ensure their symmetry-breaking function. They are nonetheless constrained to couple with the gauge field in the prescribed symmetrical way. The semiclassical view taken here is much the same that one can hold of the weakly bound correlated electron pairs in the theory of superconductors. Of course the "ether" there is the crystal lattice, a genuinely physical background, and the Higgs particles are the correlated electrons in fact present, all actively interacting on a scale much smaller than that of the macroscopic superconducting phenomenon itself. The implication that another level of real particles and fields lies at some tiny spatial scale within the vacuum of our present measures is plain, and now much elaborated.

The remaining chapters continue the tale a little sketchily into the domain of color gauge fields, presumed to be exactly symmetrical locally under an entirely new group. It can explain the binding of its sources, the quarks, with a very unusual force, whose surprising increase as quark separation increases finds here a plausible explanation in terms of virtual particle clouds. The text ends with a topological account of the magnetic monopole, although it cannot in fact demonstrate more than a methodological connection with the monopolelike solutions found for the much more complicated gauge fields of grand unified theory. A final few pages on group theory explain the terms so freely floating around in today's papers in particle physics.

This is a first-rate entry to the subject for those who heed the differential signs above the door. It is a deal deeper than the wordy analogies and parables that must suffice for the usual popular accounts. There is room for further witty pedagogy still; the persuasive but somewhat abstract geometry could be aided by a more determined dynamical invocation of the energy-momentum implications coupling terms always have.

VEHICLES: EXPERIMENTS IN SYNTHETIC PSYCHOLOGY, by Valentino Braitenberg. The MIT Press ($14.95). From Galileo to Einstein and Bohr, the thought experiment has provided a powerful tool of scientific exposition. Its clarity and freedom from formalism give it an appeal other modes lack.

The small and cheerful book at hand, by a well-known researcher on the brain from Tübingen, has exploited the virtues of the style with unprecedented consistency, originality and aptness. His thought experiments are not analytic efforts to extract what principles lie behind an imagined observation but instead synthetic constructions. They are little toys of the mind, devised out of simple if fictional components, entirely functionally described. The hardware realization is unimportant, shadowy. Of course one thinks of small boxes with motors and wheels, or perhaps propellers for waterborne cases. It does not matter; the description here is psychological. How do the vehicles behave? What are the evolutionary and other relations between behaviors? What intellectual functions are at work? Can emotions arise?

The argumentative gain is remarkable; as the author puts it, analysis is uphill work but synthesis downhill. These devices seem transparently easy to understand at the level of function; surely they will act as we foresee, for we have imagined all the component systems ourselves. Yet so cunningly are they contrived here that over the series of a dozen-odd progressively more elaborated mechanisms, the plausible behaviors these little pseudo-organisms—call them vehicles—exhibit span the world of sentience, from mere crude getting around protozoalike to what we can call foresight and egotism and optimism. All these are elicited from credible devices that differ only in what appear to be immaterial technicalities from the real, buzzing, chip-laden boxes seen on every hand.

There are 14 types of vehicle here, one to each brief chapter. It should be plain that the complete sequence is more persuasive than any sample; just as paleontology commands assent when gaps in a sequence are tellingly filled, so does the author's argument benefit from his careful evolutionary progression. But let us describe two representative vehicle types along the way.

Vehicle type 3 enjoys the too riveting title *love*. From the start we picture vehicles as little mobile structures fitted with motors and a variety of sensors; by vehicle type 2 there is one motor for each side, a sensor for each side, and the simple relation that the stronger the sensed stimulus is, the faster the motor it connects to turns. The earlier evolutionary choices dealt with the interconnections; it is not hard to agree that straight connections, sensor to motor on the same side, will make a device that homes on the source of its sensory excitation, light or oxygen concentration or whatever, and indeed is apt to collide with it.

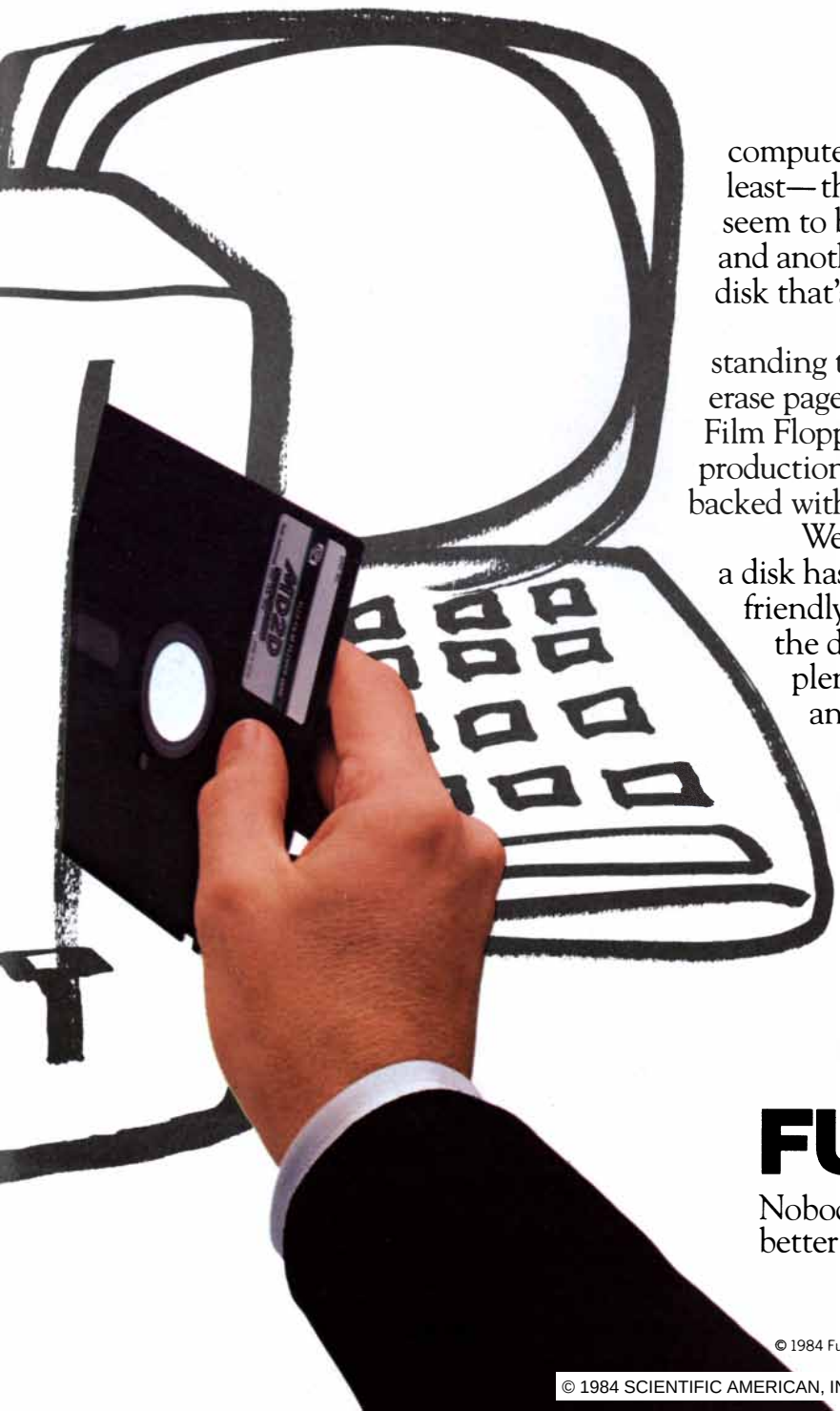
Type 3 is not so crude. A mitigating circuit enters, reversing the sensor's effect at some sensed intensity. Up to that value the sensory input speeds the motor; after a certain speed the increase of input slows the motor, finally to a stop. Type 3 vehicles show different behaviors, depending again on the connections. If the connections are same-sided, the vehicle will come to any source and stay there, "in quiet admiration from the time it spots the source to all future time." With crossed connections, it explores, still liking sources, but moving away from a source once it has gone too close if it spots another farther away.

Consider giving such a vehicle four distinct pairs of sensors, say for light, temperature, oxygen and dissolved concentration of some material. Now try a variety of connections for the four modalities, crossed and uncrossed, excitatory and inhibitory. The vehicle is really interesting. It might turn away from hot places but seek out and even collide with light bulbs; it will stay where there is oxygen and move away if oxygen supply grows low. Generalize for yourself: Many modalities, each of different influence on the motors, both in direction and in degree. Complex behavior is on the way; is it too much to speak of values, even without any analogue of inner knowledge? If you like more neutral words, such as tropisms, use them too. Our author will nonetheless feel he has made his point.

Vehicle type 7 has evolved an internal model of the world. That model depends on what its forebears gained: a large set of threshold devices that intervene between sensor and motor. These devices simply fire or not as a threshold for signal strength is crossed. But they are freely interconnected one-many, many-one and so on; they can send inhibitory as well as excitatory signals, and they excite their motors on the basis of a mix of signals from sensors with those from other threshold devices. These can, in short, compute; their chapter is named *logic*. Type 7 has logic plus a network of "special wire" that lowers its resistance on use, slowly returning it to its original low conductivity. That network connects all the threshold devices of a fairly complicated example of a logic-owning vehicle. The vehicle has gained the category of *concepts*. For if, say, aggressive vehicles in its environment are often red, the red and aggression detectors will be linked by a low-resistance wire. It will turn away from red neighbors. It might well flee the scene of many chemical leaks into the environment from a wrecked vehicle: the olfactory concept of the "smell of death."

Here the author tips his strong hand. "Let philosophers watch a breed of type 7 vehicles and . . . speculate." The creatures can generalize, recognizing, say, that any color marks the aggressive vehicles among a set of peaceable gray ones. The particular signals red, green, blue and so on have given way to a more general signal called "color." After all, though, there was a wire that recognized gray as a mix of many colors. It is no great wonder that the "not gray" should be reinforced so well. "All right," says a philosopher-discussant in the book,

A year's worth of reports, plans, schedules, charts, graphs, files, facts and figures and it could all be lost in the blink of an eye.



The most important part of your computer may be the part you've considered least—the floppy disk. After all, there doesn't seem to be much difference between one disk and another. But now Fuji introduces a floppy disk that's worth a second look.

We designed our disk with the understanding that one microscopic imperfection can erase pages of crucial data. That's why every Fuji Film Floppy Disk is rigidly inspected after each production process. And that's why each one is backed with a lifetime warranty.

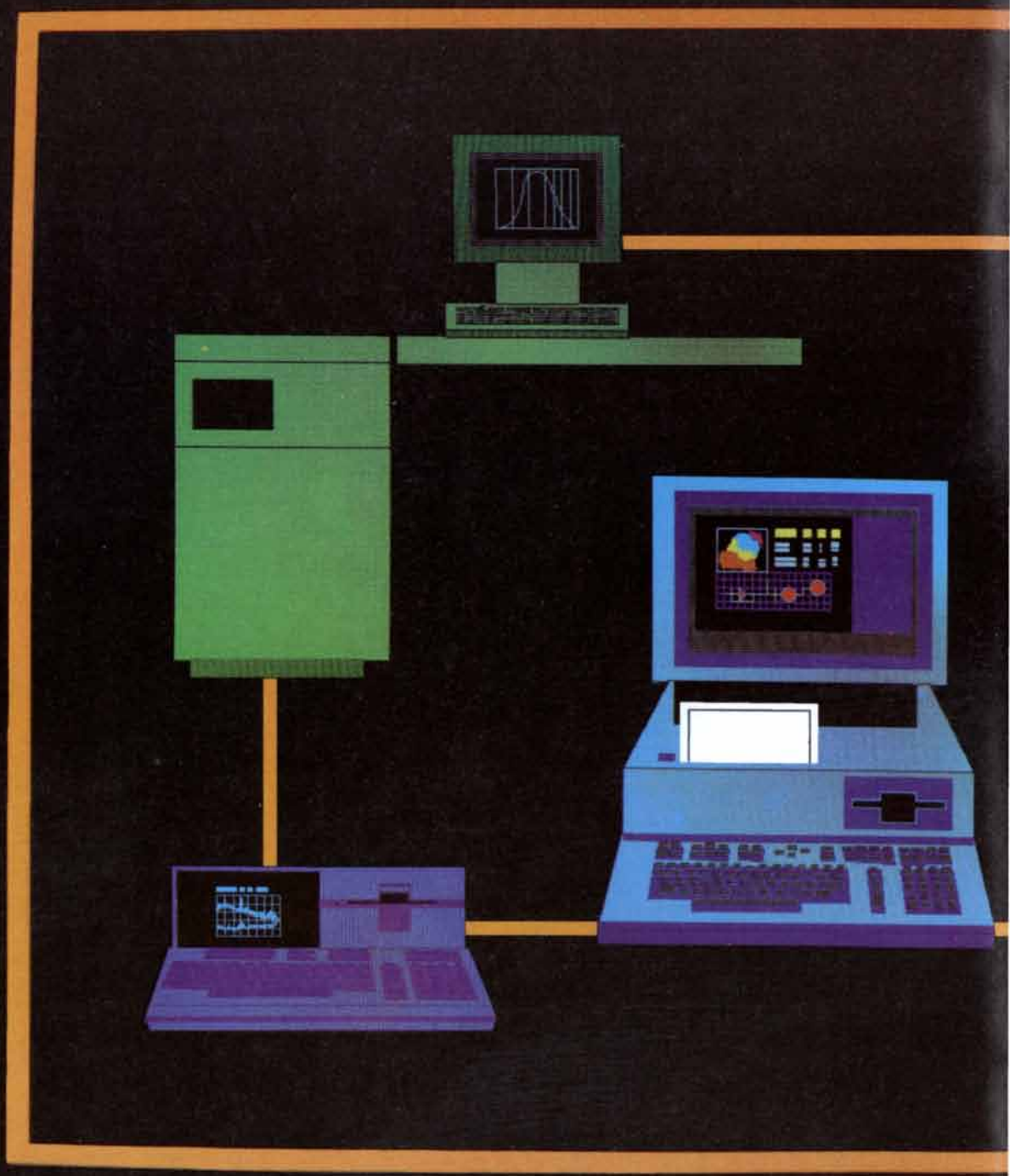
We've even considered how carefully a disk has to be handled, so we designed user-friendly packaging that makes it easier to get the disk out of the box. And we provided plenty of labeling space, so you won't have any trouble telling which disk is which.

So think twice before buying a floppy disk. And then buy the one you won't have any second thoughts about. Fuji Film Floppy Disks.

FUJI.

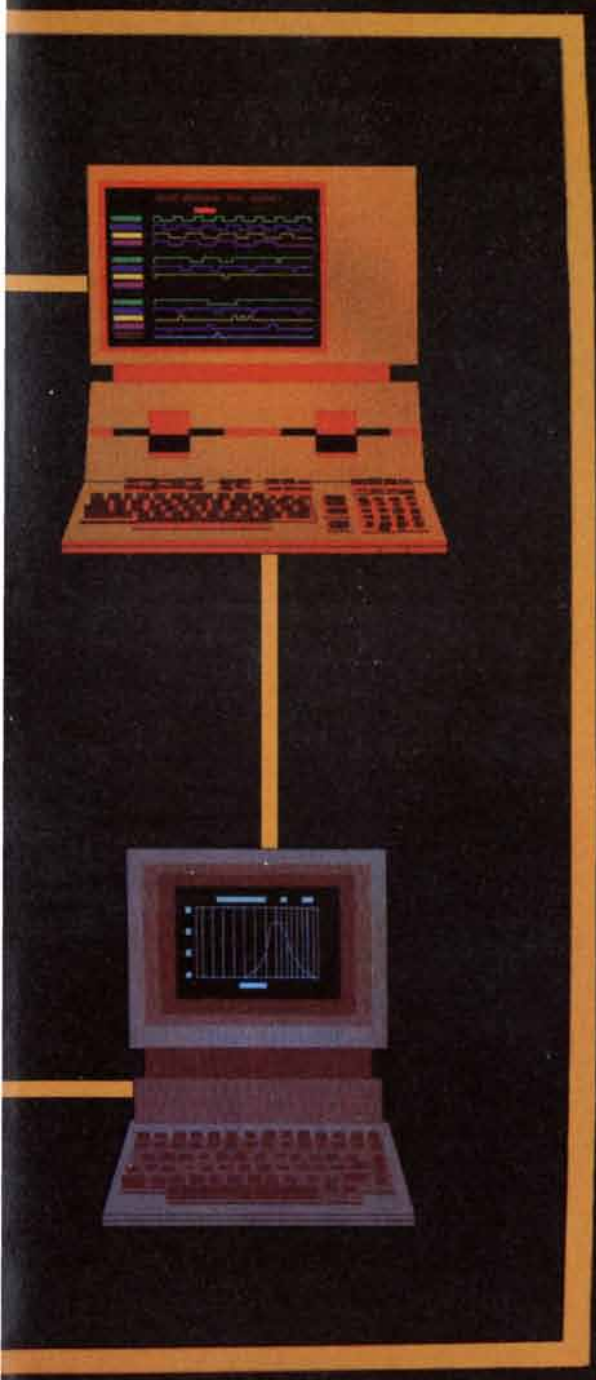
Nobody gives you better performance.

The Engineering Pr



With HP's broad range of workstations, you can choose the one that best fits your application and budget. From personal computers for engineers to desktop mainframes, with beginning prices ranging from $3,950 to $28,250.

oductivity Network.



Presenting your broadest choice of engineering workstations anywhere. From a low-cost personal computer to a powerful 32-bit system.

No doubt about it; when you put a computer on every engineer's desk, it helps them be more productive.

In fact, at Hewlett-Packard, the workstation approach has worked so well for our own engineers that we've taken it to an entirely new dimension.

It's called the Engineering Productivity Network. As you can see, it includes the broadest range of workstations available anywhere. From the recently-introduced 32-bit HP 9000 desktop mainframe to our personal computer designed just for engineers.

With such a wide range to choose from, you can match the computing power to the individual task. And give every engineer the tool it takes to do the best possible job.

To help you tailor every workstation to a task even more closely, we have over 200 specific application software packages. And since our workstations use HP-UX—our enhanced version of the UNIX* operating system—we'll be able to bring you a lot more in the future.

Of course, when you tie all these individually-optimized workstations together into a network, you can look for an even higher level of performance. Whether you choose our HP-UX networking, Ethernet local area network, or Shared Resource Manager, your engineers can share data and peripherals, communicate much better, and get more done than ever before.

The HP Engineering Productivity Network. You can start with one workstation, or with dozens. But no matter where you start, you're actually making your whole department more productive. One step at a time.

To find out more, just call the nearest HP office listed in the white pages of your telephone book, and ask for a Technical Computer Representative. Or write to Pat Welch, Dept. 41189, Hewlett-Packard, 19447 Pruneridge Avenue, Cupertino, CA 95014. In Europe, write Henk van Lammeren, Hewlett-Packard, Dept. 41189, P.O. Box 529, 1180 AM, Amstelveen, the Netherlands.

We'll be happy to show you how far you can really go with a good idea.

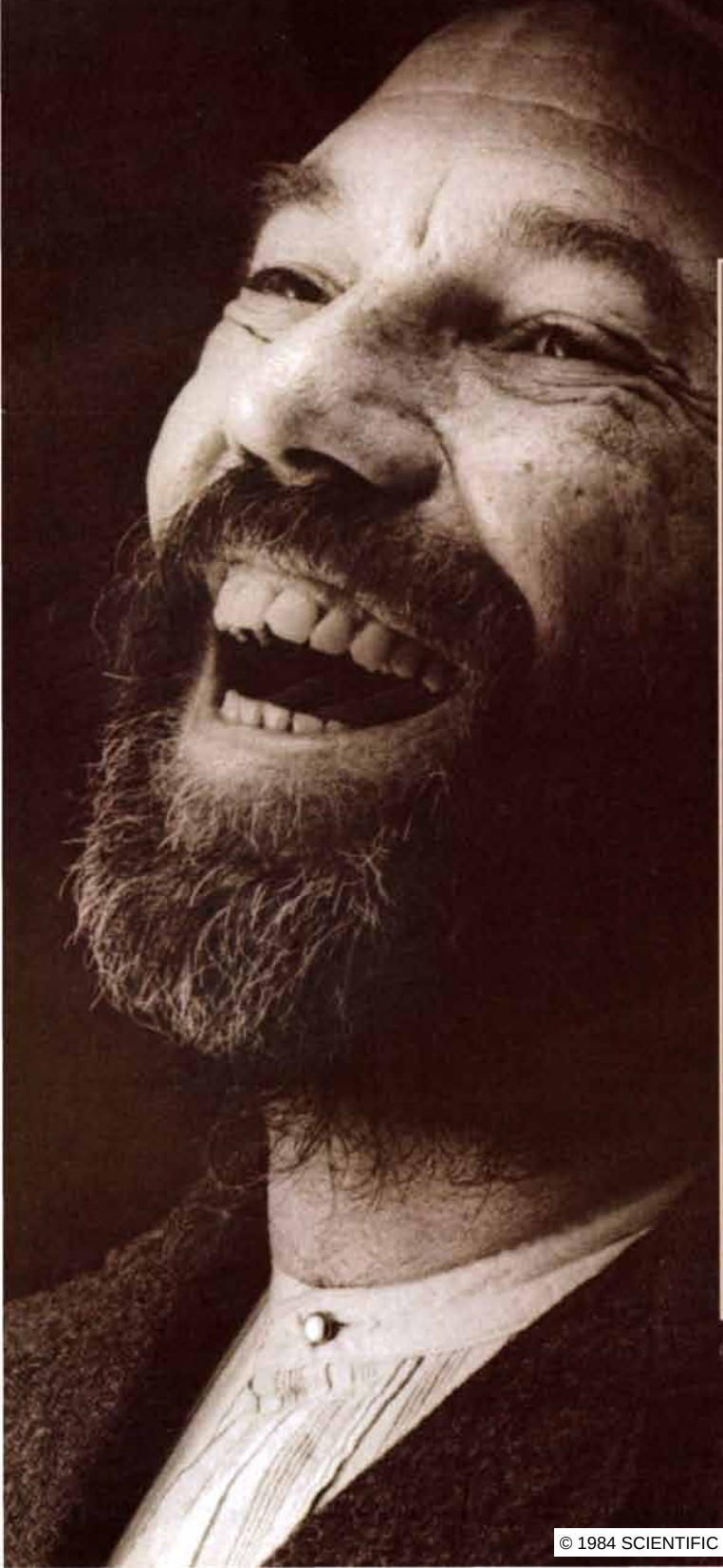
*UNIX is a trademark of Bell Laboratories.

HEWLETT PACKARD

BD-02318

IN 1912, MARIJAN KOKICH DISCOVERED AMERICA.

He came with nothing more than hopes and dreams and a taste for moussaka.



In 1980, his family discovered a great American wine with a European heritage.

Chateau Ste. Michelle Cabernet Sauvignon has varietal characteristics that rival the best wines of Europe, yet it's as American as the Kokichs' moussaka today. The next time you celebrate your family's heritage, open a bottle of Chateau Ste. Michelle. Discover the premium wines of Washington State. The wines of the Ellis Island-Liberty Centennial.

FOUNDING SPONSOR
LIBERTY
1886-1986

"but nobody in his right mind ever suspected anything more mysterious behind the 'faculty of generalization.' Fine, says I, as long as you admit it."

After the sustained cheerful parable here sampled, we are offered an interesting collection of small personal essays on facts of animal brains and behavior that have arisen out of Professor Braitenberg's own research. It would be no surprise if these had in part inspired unexpected properties of the synthetic vehicles of his fantasy.

The first essay treats of the crossed projection that is so typical of the vertebrate representation of visual space. The best-known theory we have is the work of the 19th-century Spanish neurophysiologist Santiago Ramón y Cajal. He saw the crossed projection as the correction of the image inversion due to the camera lens of the eye. The crossed chiasma reestablishes continuity of the two sides of an image otherwise broken by the optical inversion. The other systems simply followed the lead of the visual system for economy's sake. But in fact the two halves of the nerve net that project the portions of visual image higher into the frog midbrain do not connect; they remain distinct. It would not seem to matter much how the image portions are oriented, since their juxtaposition is done only through the wiring.

For 20 years the author has argued quite differently. The only sense organ that has an uncrossed connection with the cerebral hemispheres is the sense of smell. But the motor connections are crossed. That means a certain smell will more strongly affect the motor system on the side opposite to the nostril that first picks up the odor. The case is not strong; we do not know whether avoidance or attraction was more important, or how the motor organs worked among primitive vertebrates.

Like the notions embodied in vehicles of the earliest few types, these are naive ideas, mechanical explanations more in vogue before World War I than since. Today's views, manifestly less naive as computers will suggest to anyone, go much further. By type 14 we have entered fully into a world of goal-seeking. The trick of the type 14 vehicle is to build internal maps, to use them as predictors, to evaluate the several predictions by an inner scale of desirability (itself both evolved and learned from rare events) and to move and act as though the better prediction would come true. Often it will, and goals can be attained surprisingly well. "It seems to be a good strategy, this running after a dream.... It seems to me sufficient to take away any aura of mystery from goal-directed behavior."

This small, crisp, cogent book is full of intellectual delights. A set of pen-and-ink sketches by the artist Maciek Albrecht seeks to evoke the vehicles of the fantasy. A predisposed reader is perhaps too easily persuaded by the concreteness of the argument here; others will not be, those who seek to categorize behavior so as to define human uniqueness through some immanent dimension. Certainly the inner reality of mind is not yet modeled, although the path now seems open. The chief fault that crosses this plateau of strong analogy is quantitative. Vehicles of the real living sort have indeed many, many parts; in the fantastic combinatorics of all those components, are there not important features and systems in the functional architecture of the brain of which we still hold no clear idea? Even the parable is incomplete, and sure to grow.

THE HEIRLOOM GARDENER, by Carolyn Jabs. Sierra Club Books ($17.95; paperbound, $9.95). "The epicure of vegetable morsels may not rest in his search for the acme of all sweet corn until he has eaten Black Mexican fresh from the field." Such was the expressed judgment of the experts of the State Experiment Station when during the 1930's they compiled their volume on the sweet corn varieties of New York. No commonplace sort of maize, Black Mexican's kernels are "white at the milk stage but turn jet black as they dry." Perhaps the single-crop seed exchange, called CORNS, at Turpin, Okla., can help the unresting gardener test that claim of excellence. Meanwhile huge ears with strikingly deep blue kernels, a corn crop carefully watered over the centuries perhaps from the high, sacred Blue Lake itself, need no longer be shared only through the recollections of a few favored guests: the Taos Pueblo Native Seed Company, at P.O. Box EEE out there in NM 87571, will sell good seed of Taos Pueblo Blue to any gardener.

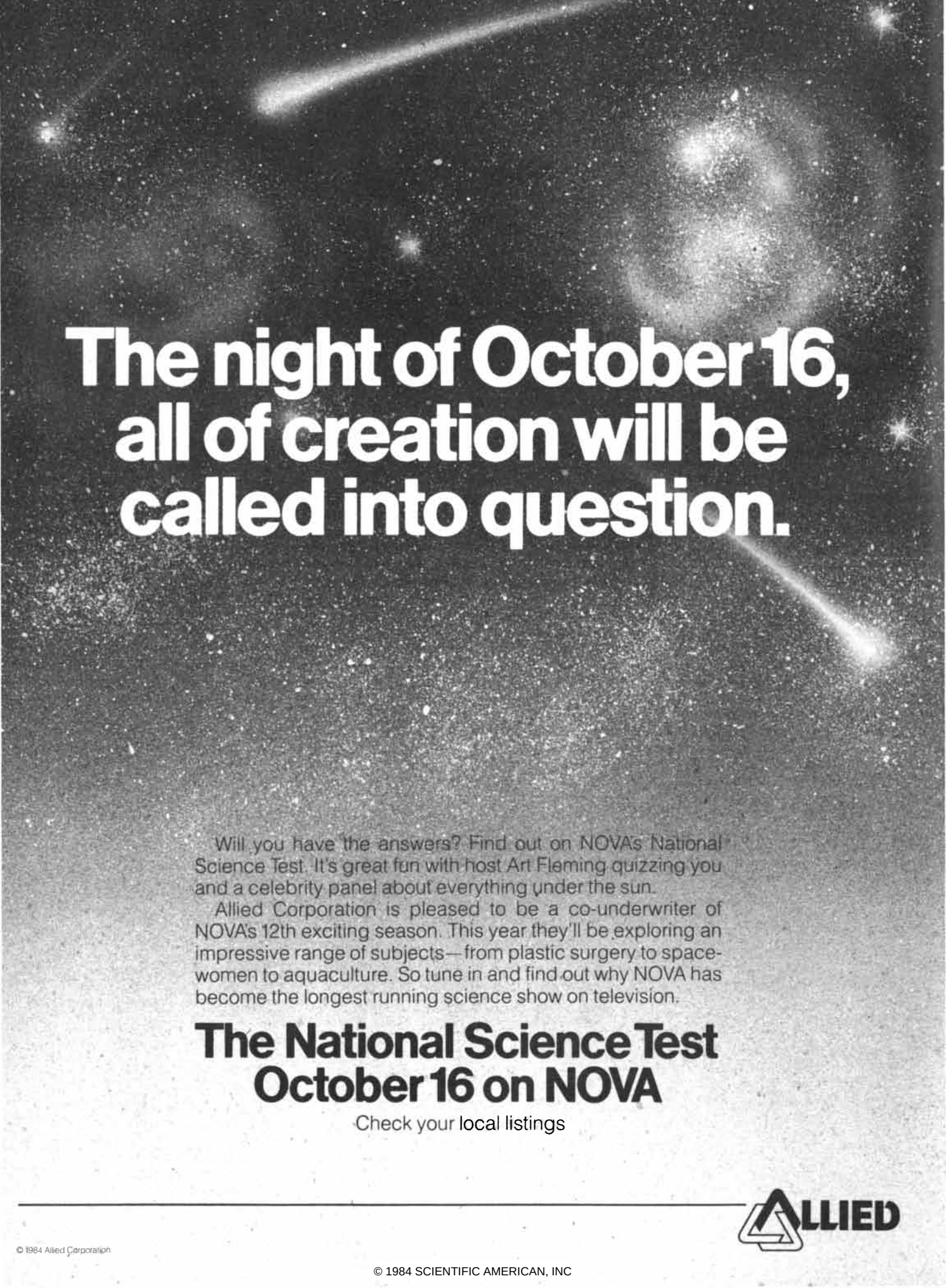
Three historical remarks underlie this well-informed book. The first is that the plants that nourish us, like the flowers that please, are human artifacts as much as they are living organisms. Heirlooms, then, include not only birch highboys and fine pewter but also old varieties of domestic plants, those taken into cultivation by plant breeders and lucky alert gardeners in every generation. Most North American cultivars were introduced from other lands, but by now the muster of centuries of naturalized immigrants is a rich collection out of our past, as much a part of needed and enjoyable diversity as the wild forms or the ancient cultivars overseas. The second truth is that, although only a few percent of us now make a living by farming, 35 million Americans are gardeners, amateurs of the chief vocation of our species (and what a lobby!). The third point is more conspicuous: economic optimization in the marketplace of so large a country is often the enemy of diversity. Big seed companies tend to rear what most commercial growers will buy: varieties with good yield, resistance to drought and disease, and an attractive but customary appearance. The past offers qualities, notably flavor, fragrance and visual variety, less prized in our profit-driven agriculture of long transport routes and out-of-season supermarket sales to hurried urban buyers.

All three points are well examined in this volume, which fits with the reforming and participatory style of the Sierra Club. Carolyn Jabs is an experienced gardener and a zealous proponent of purposeful gardening to replicate the irreplaceable legacy of the past, those heirlooms of selection and good fortune that our old orchards and gardens still hold. There have grown up cooperatives of amateur gardeners, eager to share the living legacies, small commercial enterprises aimed at providing the seeds of the past, and gardens of museums devoted to celebration of our past that wisely recognize that old plants are a key part of old times, from Fort Vancouver to Monticello. All these organizations are listed here by the page, with addresses and helpful comment.

This path to the preservation of old varieties is by their widespread recultivation. Such amateur effort is a necessary complement to the expert liquid-nitrogen storage and systematic plantings of the overtaxed National Plant Germplasm System, from the 200,000 plant forms held at Fort Collins through the entire chain of regional and clonal repositories and plant-introduction stations. That system aims largely at maintenance of introduced but unused genetic variety; it has less concern with what the amateurs do best, keep up varieties once domestically popular and prized.

The author includes valuable crop-by-crop tips on the technique of growing seed, among them the measures sometimes needed to control pollination. Wind-pollinated corn, for example, is a "challenging collectible"; to maintain your Black Mexican strain all the neighbors must cooperate or you must learn hand pollination of shielded tassels.

John Withee of Lynnfield, Mass., holds perhaps the nation's largest private collection of old seeds, more than 1,000 named varieties of beans. He grows one plant of each of 500 bean varieties each year, and around him has gathered a devoted volunteer network of growers-in-aid. One hopes this book will enlist a large number of new entrants, at every level of skill and devotion, into this kind of collecting, where antiques literally live again. The gardener's potato can be purple-skinned, as her tomato plants can bear the fine-tasting sharp Brandywine, one heirloom held in a single family for a century. As for those blue cornflakes, derived from the ancient Hopi prototype called piki bread... is Battle Creek listening?



Established 1845 SCIENTIFIC AMERICAN October 1984 Volume 251 Number 4

Space-based Ballistic-Missile Defense

President Reagan's "Star Wars" program seems unlikely ever to protect the entire nation against a nuclear attack. It would nonetheless trigger a major expansion of the arms race

by Hans A. Bethe, Richard L. Garwin, Kurt Gottfried and Henry W. Kendall

For two decades both the U.S. and the U.S.S.R. have been vulnerable to a devastating nuclear attack, inflicted by one side on the other in the form of either a first strike or a retaliatory second strike. This situation did not come about as the result of careful military planning. "Mutual assured destruction" is not a policy or a doctrine but rather a fact of life. It simply descended like a medieval plague—a seemingly inevitable consequence of the enormous destructive power of nuclear weapons, of rockets that could hurl them across almost half of the globe in 30 minutes and of the impotence of political institutions in the face of such momentous technological innovations.

This grim development holds different lessons for different people. Virtually everyone agrees that the world must eventually escape from the shadow of mutual assured destruction, since few are confident that deterrence by threat of retaliation can avert a holocaust indefinitely. Beyond this point, however, the consensus dissolves. Powerful groups in the governments of both superpowers apparently believe that unremitting competition, albeit short of war, is the only realistic future one can plan for. In the face of much evidence to the contrary they act as if the aggressive exploitation for military purposes of anything technology has to offer is critical to the security of the nation they serve. Others seek partial measures that could at least curb the arms race, arguing that this approach has usually been sidetracked by short-term (and shortsighted) military and political goals. Still others have placed varying degrees of faith in radical solutions: novel political moves, revolutionary technological advances or some combination of the two.

President Reagan's Strategic Defense Initiative belongs in this last category. In his televised speech last year calling on the nation's scientific community "to give us the means of rendering these nuclear weapons impotent and obsolete" the president expressed the hope that a technological revolution would enable the U.S. to "intercept and destroy strategic ballistic missiles before they reached our own soil or that of our allies." If such a breakthrough could be achieved, he said, "free people could live secure in the knowledge that their security did not rest upon the threat of instant U.S. retaliation."

Can this vision of the future ever become reality? Can any system for ballistic-missile defense eliminate the threat of nuclear annihilation? Would the quest for such a defense put an end to the strategic-arms race, as the president and his supporters have suggested, or is it more likely to accelerate that race? Does the president's program hold the promise of a secure and peaceful world or is it perhaps the most grandiose manifestation of the illusion that science can re-create the world that disappeared when the first nuclear bomb was exploded in 1945?

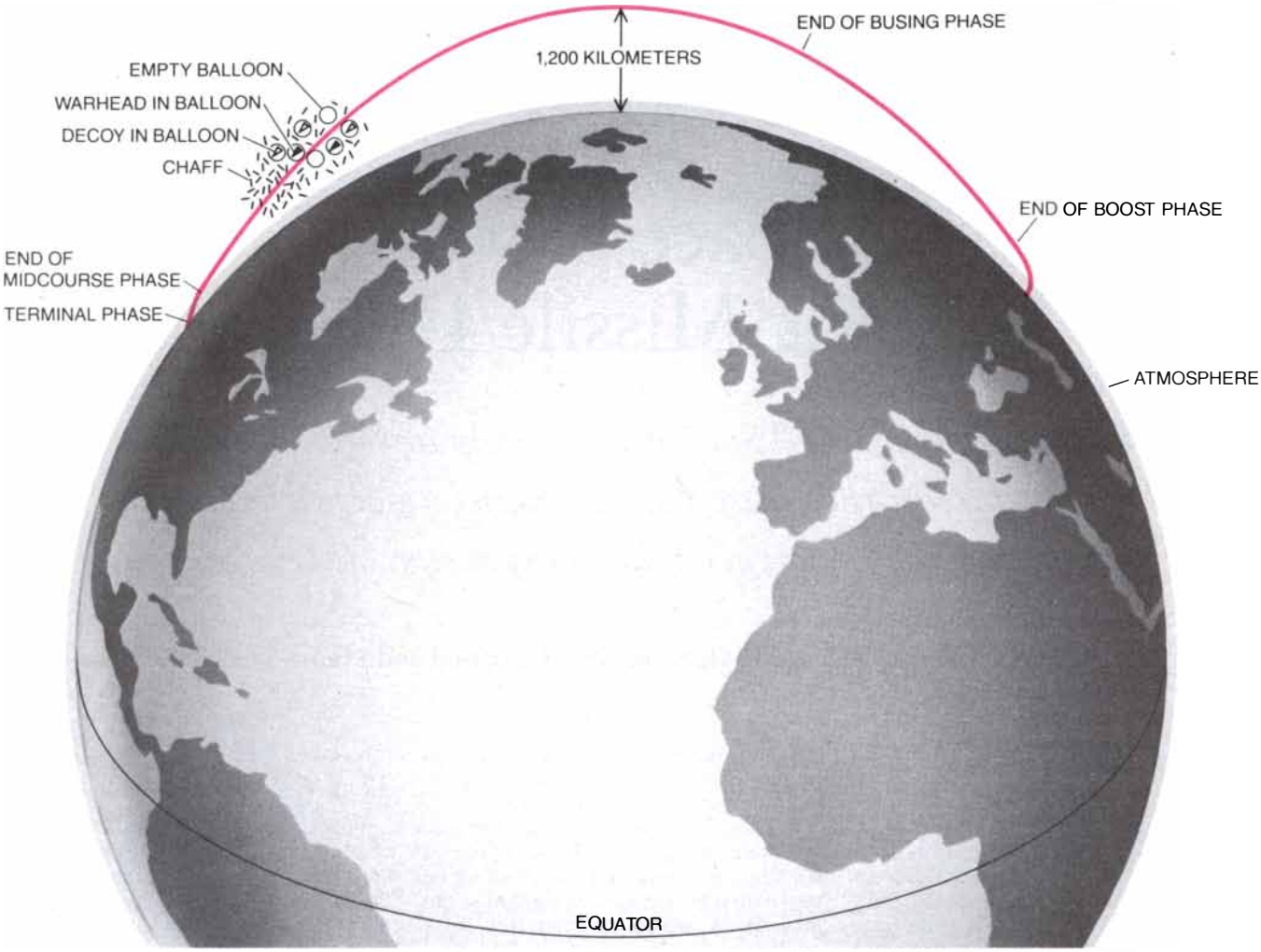
These are complex questions, with intertwined technical and political strands. They must be examined carefully before the U.S. commits itself to the quest for such a defense, because if the president's dream is to be pursued, space will become a potential field of confrontation and battle. It is partly for this reason the Strategic Defense Initiative is commonly known as the "Star Wars" program.

This article, which is based on a forthcoming book by a group of us associated with the Union for Concerned Scientists, focuses on the technical aspects of the issue of space-based ballistic-missile defense. Our discussion of the political implications of the president's Strategic Defense Initiative will draw on the work of two of our colleagues, Peter A. Clausen of the Union for Concerned Scientists and Richard Ned Lebow of Cornell University.

The search for a defense against nuclear-armed ballistic missiles began three decades ago. In the 1960's both superpowers developed anti-ballistic-missile (ABM) systems based on the use of interceptor missiles armed with nuclear warheads. In 1968 the U.S.S.R. began to operate an ABM system around Moscow based on the Galosh interceptor, and in 1974 the U.S. completed a similar system to protect Minuteman missiles near Grand Forks Air Force Base in North Dakota. (The U.S. system was dismantled in 1975.)

Although these early efforts did not provide an effective defense against a major nuclear attack, they did stimulate two developments that have been dominant features of the strategic landscape ever since: the ABM Treaty of 1972 and the subsequent deployment of multiple independently targetable reentry vehicles (MIRV's), first by the U.S. and later by the U.S.S.R.

In the late 1960's a number of scientists who had been involved in investi-



FOUR DISTINCT PHASES are evident in the flight of an intercontinental ballistic missile (ICBM). In boost phase the missile is carried above the atmosphere by a multistage booster rocket. Most modern strategic missiles carry multiple independently targetable reentry vehicles (MIRV's), which are released sequentially by a maneuverable "bus" during the busing, or postboost, phase. If the country under attack had a ballistic-missile-defense system, the bus would also dispense a variety of "penetration aids," such as decoys, balloons enclosing MIRV's and decoys, empty balloons, radar-reflecting wires called chaff and infrared-emitting aerosols. During the midcourse phase the heavy MIRV's and the light penetration aids would follow essentially identical trajectories. In the terminal phase this "threat cloud" would reenter the atmosphere, and friction with the air would retard the penetration aids much more than the MIRV's. For ICBM's the flight would last between 25 and 30 minutes; for submarine-launched ballistic missiles (SLBM's) it could be as short as eight to 10 minutes.

MISSILE	GROSS WEIGHT (KILOGRAMS)	END OF BOOST PHASE		END OF BUSING		USUAL PAYLOAD
		TIME (SECONDS)	ALTITUDE (KILOMETERS)	TIME (SECONDS)	ALTITUDE (KILOMETERS)	
SS-18	220,000	300	400	?	?	10 MIRV'S ON ONE BUS
MX	89,000	180	200	650	1,100	10 MIRV'S ON ONE BUS
MX WITH FAST-BURNING BOOSTER	87,000	50	90	60	110	SEVERAL MICROBUSES WITH MIRV'S AND PENETRATION AIDS
MIDGETMAN	19,000	220	340	—	—	SINGLE WARHEAD
MIDGETMAN WITH FAST-BURNING BOOSTER	22,000	50	80	—	—	SINGLE WARHEAD WITH PENETRATION AIDS

CHARACTERISTICS OF FIRST TWO PHASES in the flight of an ICBM are given for five missiles: the SS-18, a very large, multiple-warhead ICBM already deployed by the U.S.S.R.; the MX, a large, multiple-warhead ICBM currently under development by the U.S.; the Midgetman, a smaller, single-warhead ICBM now in the early planning stages in the U.S., and two hypothetical missiles comparable to the MX and the Midgetman that have been specifically designed to counter a boost-phase ballistic-missile-defense system. In this case the assumption is that both missiles would be equipped not only with suitable penetration aids but also with fast-burning boosters, thereby reducing the time available for the defense to detect their infrared emission. The SS-18 is constrained under the terms of the SALT II Treaty to carry no more than 10 MIRV's; it is actually capable of carrying 30 or more smaller warheads. A single-warhead missile such as Midgetman need have no bus and hence there would be no distinction in its case between the postboost phase and the midcourse phase. The table is adapted from a report prepared by Ashton B. Carter for the Congressional Office of Technology Assessment.

gating the possibility of ballistic-missile defense in their capacity as high-level advisers to the U.S. Government took the unusual step of airing their criticism of the proposed ABM systems both in congressional testimony and in the press [see "Anti-Ballistic-Missile Systems," by Richard L. Garwin and Hans A. Bethe; SCIENTIFIC AMERICAN, March, 1968]. Many scientists participated in the ensuing debate, and eventually a consensus emerged in the scientific community regarding the flaws in the proposed systems.

The scientists' case rested on a technical assessment and a strategic prognosis. On the technical side they pointed out that the systems then under consideration were inherently vulnerable to deception by various countermeasures and to preemptive attack on their exposed components, particularly their radars. On the strategic side the scientists argued that the U.S.S.R. could add enough missiles to its attacking force to ensure penetration of any such defense. These arguments eventually carried the day, and they are still germane. They were the basis for the ABM Treaty, which was signed by President Nixon and General Secretary Brezhnev in Moscow in May, 1972. The ABM Treaty formally recognized that not only the deployment but also the development of such defensive systems would have to be strictly controlled if the race in offensive missiles was to be contained.

MIRV's were originally conceived as the ideal countermeasure to ballistic-missile defense, and in a logical world they would have been abandoned with the signing of the ABM Treaty. Nevertheless, the U.S. did not try to negotiate a ban on MIRV's. Instead it led the way to their deployment in spite of repeated warnings by scientific advisers and the Arms Control and Disarmament Agency to senior Government officials that MIRV's would undermine the strategic balance and ultimately be to the advantage of the U.S.S.R. because of its larger ICBM's. The massive increase in the number of nuclear warheads in both strategic arsenals during the 1970's is largely attributable to the introduction of MIRV's. The result, almost everyone now agrees, is a more precarious strategic balance.

The president's Strategic Defense Initiative is much more ambitious than the ABM proposals of the 1960's. To protect an entire society a nationwide defense of "soft" targets such as cities would be necessary; in contrast, the last previous U.S. ABM plan—the Safeguard system proposed by the Nixon Administration in 1969—was intended to provide only a "point" defense of "hard" targets such as missile silos and command bunkers. The latter mission could be accomplished by a quite permeable terminal-defense system that intercepted warheads very close to their targets, since a formidable retaliatory capability would remain even if most of the missile silos were destroyed. A large metropolitan area, on the other hand, could be devastated by a handful of weapons detonated at high altitude; if necessary, the warheads could be designed to explode on interception.

To be useful a nationwide defense would have to intercept and eliminate virtually all the 10,000 or so nuclear warheads that each side is currently capable of committing to a major strategic attack. For a city attack it could not wait until the atmosphere allowed the defense to discriminate between warheads and decoys. Such a high rate of attrition would be conceivable only if there were several layers of defense, each of which could reliably intercept a large percentage of the attacking force. In particular, the first defensive layer would have to destroy most of the attacking warheads soon after they left their silos or submerged submarines, while the booster rockets were still firing. Accordingly boost-phase interception would be an indispensable part of any defense of the nation as a whole.

Booster rockets rising through the atmosphere thousands of miles from U.S. territory could be attacked only from space. That is why the Strategic Defense Initiative is regarded primarily as a space-weapons program. If the president's plan is actually pursued, it will mark a turning point in the arms race perhaps as significant as the introduction of ICBM's.

Several quite different outcomes of the introduction of space weapons have been envisioned. One view (apparently widely held in the Reagan Administration) has been expressed most succinctly by Robert S. Cooper, director of the Defense Advanced Research Projects Agency. Testifying last year before the Armed Services Committee of the House of Representatives, Cooper declared: "The policy for the first time recognizes the need to control space as a military environment." Indeed, given the intrinsic vulnerability of space-based systems, the domination of space by the U.S. would be a prerequisite to a reliable ballistic-missile defense of the entire nation. For that reason, among others, the current policy also calls for the acquisition by the U.S. of antisatellite weapons [see "Antisatellite Weapons," by Richard L. Garwin, Kurt Gottfried and Donald L. Hafner; SCIENTIFIC AMERICAN, June].

The notion that the U.S. could establish and maintain supremacy in space ignores a key lesson of the post-Hiroshima era: a technological breakthrough of even the most dramatic and unexpected nature can provide only a temporary advantage. Indeed, the only outcome one can reasonably expect is that both superpowers would eventually develop space-based ballistic-missile-defense systems. The effectiveness of these systems would be uncertain and would make the strategic balance more precarious than it is today. Both sides will have expanded their offensive forces to guarantee full confidence in their ability to penetrate defenses of unknown reliability, and the incentive to cut one's own losses by striking first in a crisis will be even greater than it is now. Whether or not weapons deployed in space could ever provide a reliable defense against ballistic missiles, they would be potent antisatellite weapons. As such they could be used to promptly destroy an opponent's early-warning and communications satellites, thereby creating a need for critical decisions at a tempo ill suited to the speed of human judgment.

Our analysis of the prospects for a space-based defensive system against ballistic-missile attack will focus on the problem of boost-phase interception. It is not only an indispensable part of the currently proposed systems but also what distinguishes the current concept from all previous ABM plans. On the basis of our technical analysis and our assessment of the most likely response of the U.S.S.R. we conclude that the pursuit of the president's program would inevitably stimulate a large increase in the Russian strategic offensive forces, further reduce the chances of controlling events in a crisis and possibly provoke the nuclear attack it was designed to prevent. In addition the reliability of the proposed defense would remain a mystery until the fateful moment at which it was attacked.

Before assessing the task of any defense one must first examine the likely nature of the attack. In this case we shall concentrate on the technical and military attributes of the land-based ICBM and on how a large number of such missiles could be used in combination to mount a major strategic attack.

The flight of an ICBM begins when the silo door opens and hot gases eject the missile. The first-stage booster then ignites. After exhausting its fuel the first stage falls away as the second stage takes over; this sequence is usually repeated at least one more time. The journey from the launch point to where the main rockets stop burning is the boost phase. For the present generation of ICBM's the boost phase lasts for three to five minutes and ends at an altitude of 300 to 400 kilometers, above the atmosphere.

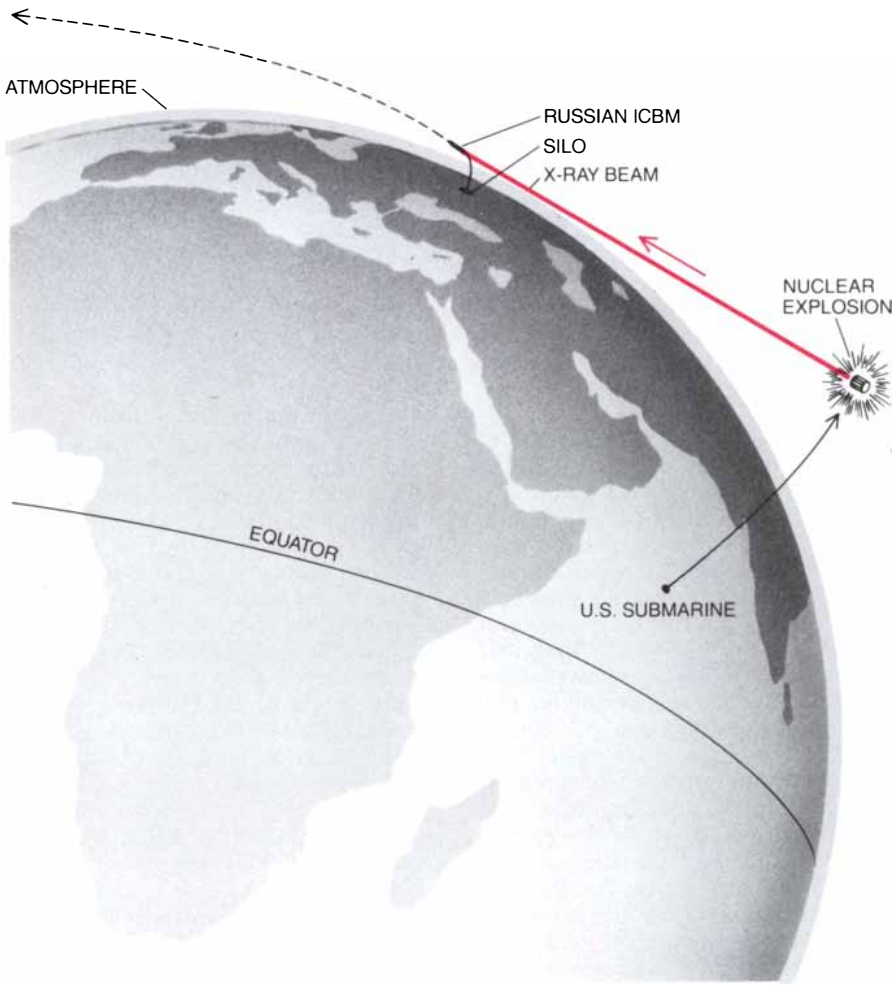
A typical ICBM in the strategic arsenal of the U.S. or the U.S.S.R. is equipped with MIRV's, which are dispensed by a maneuverable carrier vehicle called a

bus after the boost phase ends. The bus releases the MIRV's one at a time along slightly different trajectories toward their separate targets. If there were defenses, the bus could also release a variety of penetration aids, such as lightweight decoys, reentry vehicles camouflaged to resemble decoys, radar-reflecting wires called chaff and infrared-emitting aerosols. Once the bus had completed its task the missile would be in midcourse. At that point the ICBM would have proliferated into a swarm of objects, each of which, no matter how light, would move along a ballistic trajectory indistinguishable from those of its accompanying objects. Only after the swarm reentered the atmosphere would the heavy, specially shaped reentry vehicles be exposed as friction with the air tore away the screen of lightweight decoys and chaff.

This brief account reveals why boost-phase interception would be crucial: every missile that survived boost phase would become a complex "threat cloud" by the time it reached midcourse. Other factors also amplify the importance of boost-phase interception. For one thing, the booster rocket is a much larger and more fragile target than the individual reentry vehicles are. For another, its flame is an abundant source of infrared radiation, enabling the defense to get an accurate fix on the missile. It is only during boost phase that a missile reveals itself by emitting an intense signal that can be detected at a large distance. In midcourse it must first be found by illuminating it with microwaves (or possibly laser light) and then sensing the reflected radiation, or by observing its weak infrared signal, which is due mostly to reflection of the earth's infrared radiation.

Because a nationwide defense must be capable of withstanding any kind of strategic attack, the exact nature of the existing offensive forces is immaterial to the evaluation of the defense. At present a full-scale attack by the U.S.S.R. on the U.S. could involve as many as 1,400 land-based ICBM's. The attack might well begin with submarine-launched ballistic missiles (SLBM's), since their unpredictable launch points and short flight times (10 minutes or less) would lend the attack an element of surprise that would be critical if the national leadership and the ground-based bomber force were high-priority targets.

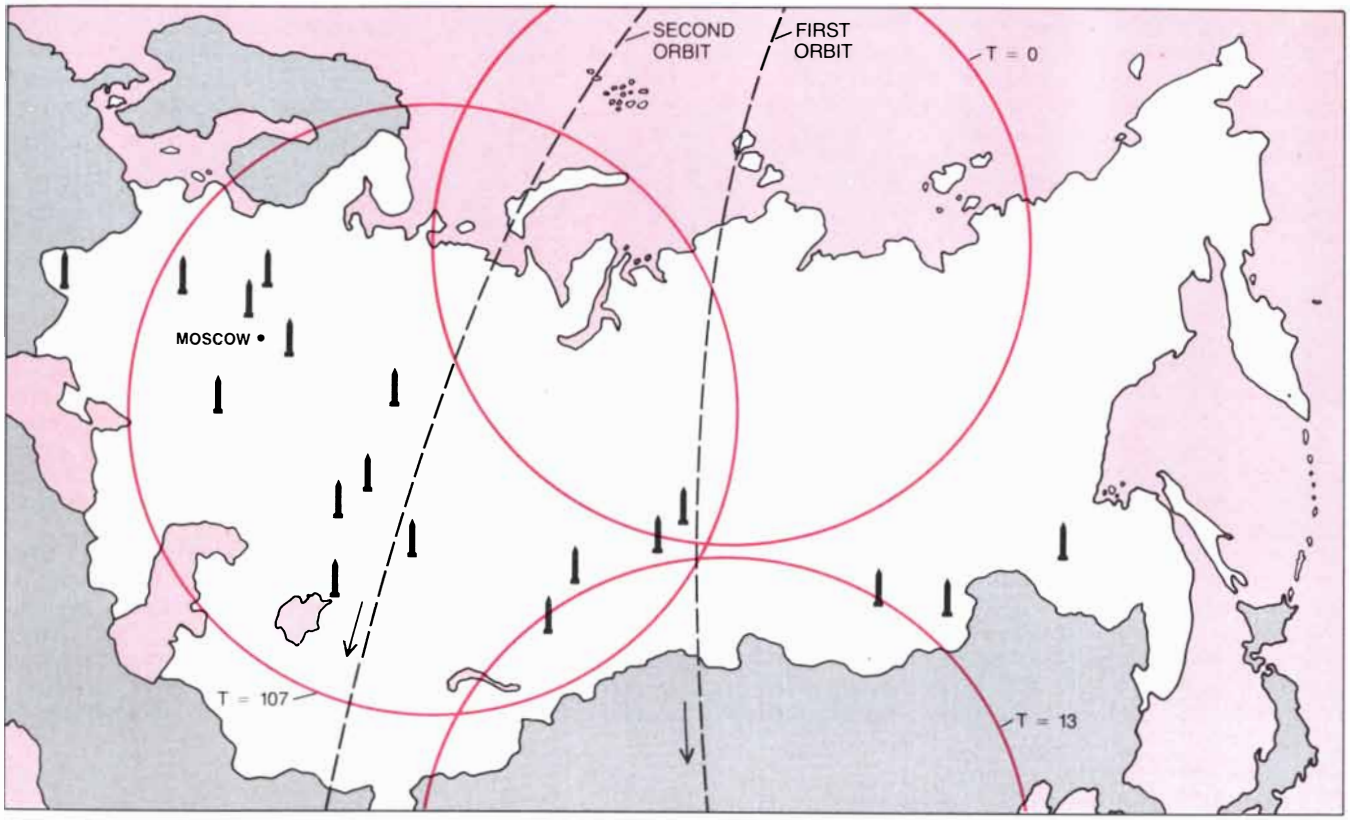
SLBM's would be harder to intercept than ICBM's, which spend 30 minutes or so on trajectories whose launch points are precisely known. Moreover, a space-based defense system would be unable to intercept ground-hugging cruise missiles, which can deliver nuclear warheads to distant targets with an accuracy that is independent of range. Both superpowers are developing sea-launched cruise missiles, and these weapons are certain to become a major part of their strategic forces once space-based ballistic-missile-defense systems appear on the horizon.



"POP-UP" DEFENSIVE SYSTEM would rely on a comparatively light interceptor launched from a submarine stationed in waters as close to the Russian ICBM fields as possible (in this case in the northern Indian Ocean). At present the leading candidate for this mission is the X-ray laser, a device consisting of a nuclear explosive surrounded by a cylindrical array of thin metallic fibers. Thermal X rays from the nuclear explosion would stimulate the emission of a highly directed beam of X-radiation from the fibers in the microsecond before the device was destroyed. In order to engage ICBM's similar to the MX rising out of the closest missile silos in the U.S.S.R. while they were still in their boost phase, the interceptor would have to travel at least 940 kilometers from the submarine to the point where the device would be detonated.

The boost-phase layer of the defense would require many components that are not weapons in themselves. They would provide early warning of an attack by sensing the boosters' exhaust plumes; ascertain the precise number of the attacking missiles and, if possible, their identities; determine the trajectories of the missiles and get a fix on them; assign, aim and fire the defensive weapons; assess whether or not interception was successful, and, if time allowed, fire additional rounds. This intricate sequence of operations would have to be automated, because the total duration of the boost phase, now a few minutes, is likely to be less than 100 seconds by the time the proposed defensive systems are ready for deployment.

If a sizable fraction of the missiles were to survive boost-phase interception, the midcourse defensive layer would have to deal with a threat cloud consisting of hundreds of thousands of objects. For example, each bus could dispense as many as 100 empty aluminized Mylar balloons weighing only 100 grams each. The bus would dispense reentry vehicles (and possibly some decoy reentry vehicles of moderate weight) enclosed in identical balloons. The balloons and the decoys would have the same optical and microwave "signature" as the camouflaged warheads, and therefore the defensive system's sensors would not be able to distinguish between them. The defense would have to disturb the threat cloud in some way in order to find the heavy reentry vehicles, perhaps by detonating a nuclear explosive in the path of the cloud. To counteract such a measure, however, the reentry vehicles could be designed to release more balloons. Alternatively, the midcourse defense could be designed to tar-



COVERAGE OF THE U.S.S.R. by an antimissile weapon with a range of 3,000 kilometers deployed in a polar orbit at an altitude of 1,000 kilometers is indicated by the three circles on this map. The circles show the extent of the weapon's effect at two times separated by 13 minutes on one circuit of the earth and at another time 94 minutes later, on the next circuit. The orbiting weapon could be either a laser or a "fighting mirror" designed to reflect the light sent to it by a mirror stationed at an altitude of 36,000 kilometers above the Equator.

get everything in the threat cloud, a prodigious task that might be beyond the supercomputers expected a decade from now. In short, the midcourse defense would be overwhelmed unless the attacking force was drastically thinned out in the boost phase.

Because the boosters would have to be attacked while they could not yet be seen from any point on the earth's surface accessible to the defense, the defensive system would have to initiate boost-phase interception from a point in space, at a range measured in thousands of kilometers. Two types of "directed energy" weapon are currently under investigation for this purpose: one type based on the use of laser beams, which travel at the speed of light (300,000 kilometers per second), and the other based on the use of particle beams, which are almost as fast. Nonexplosive projectiles that home on the booster's infrared signal have also been proposed.

There are two alternatives for basing such weapons in space. They could be in orbit all the time or they could be "popped up" at the time of the attack. There are complementary advantages and disadvantages to each approach. With enough weapons in orbit some would be "on station" whenever they were needed, and they could provide global coverage; on the other hand, they would be inefficient because of the number of weapons that would have to be actively deployed, and they would be extremely vulnerable. Pop-up weapons would be more efficient and less vulnerable, but they would suffer from formidable time constraints and would offer poor protection against a widely dispersed fleet of strategic submarines.

Pop-up interceptors of ICBM's would have to be launched from submarines, since the only accessible points close enough to the Russian ICBM silos are in the Arabian Sea and the Norwegian Sea, at a distance of more than 4,000 kilometers. An interceptor of this type would have to travel at least 940 kilometers before it could "see" an ICBM just burning out at an altitude of 200 kilometers. If the interceptor were lofted by an ideal instant-burn booster with a total weight-to-payload ratio of 14 to one, it could reach the target-sighting point in about 120 seconds. For comparison, the boost phase of the new U.S. MX missile (which has a weight-to-payload ratio of 25 to one) is between 150 and 180 seconds. In principle, therefore, it should just barely be possible by this method to intercept a Russian missile comparable to the MX, provided the interception technique employed a beam that moves at the speed of light. On the other hand, it would be impossible to intercept a large number of missiles, since many silos would be more than 4,000 kilometers away, submarines cannot launch all their missiles simultaneously and 30 seconds would leave virtually no time for the complex sequence of operations the battle-management system would have to perform.

A report prepared for the Fletcher panel, the study team set up last year by the Department of Defense under the chairmanship of James C. Fletcher of the University of Pittsburgh to evaluate the Strategic Defense Initiative for the president, bears on this question. According to the report, it is possible to build ICBM's that could complete the boost phase and disperse their MIRV's in only 60 seconds, at a sacrifice of no more than 20 percent of payload. Even with zero decision time a hypothetical instant-burn rocket that could pop up an interceptor system in time for a speed-of-light attack on such an ICBM would need an impossible weight-to-payload ratio in excess of 800 to one! Accordingly all pop-up interception schemes, no matter what kind of antimissile weapon they employ, depend on the assumption that the U.S.S.R. will not build ICBM's with a boost phase so short that no pop-up system could view the burning booster.

The time constraint faced by pop-up schemes could be avoided by putting at least some parts of the system into

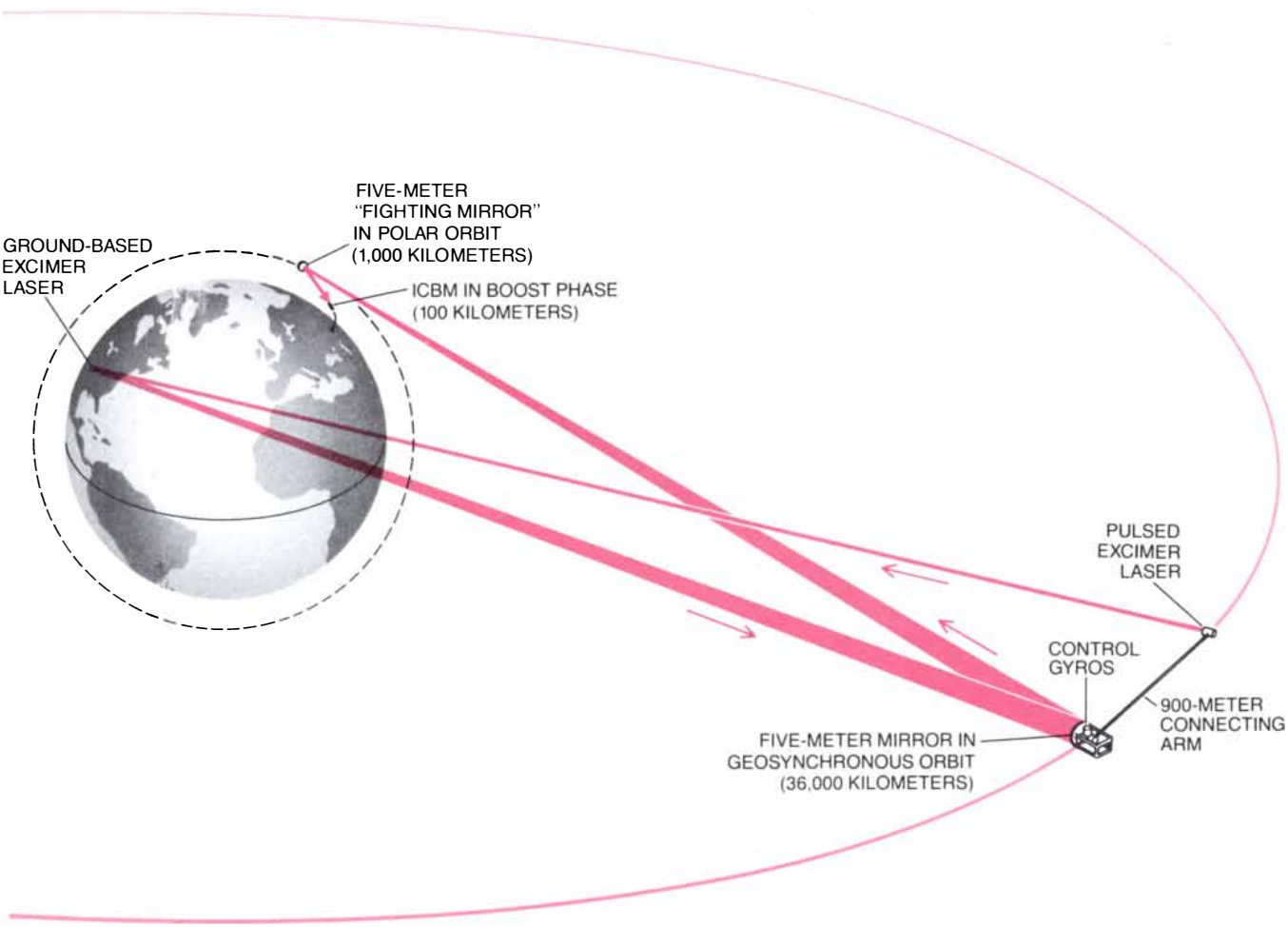
orbit. An antimissile satellite in a low orbit would have the advantage of having the weapon close to its targets, but it would suffer from the "absentee" handicap: because of its own orbital motion, combined with the earth's rotation, the ground track of such a satellite would pass close to a fixed point on the earth's surface only twice a day. Hence for every low-orbit weapon that was within range of the ICBM silos many others would be "absentees": they would be below the horizon and unable to take part in the defense. This unavoidable replication would depend on the range of the defensive weapon, the altitude and inclination of its orbit and the distribution of the enemy silos.

The absentee problem could be solved by mounting at least some components of the defensive system on a geosynchronous satellite, which remains at an altitude of some 36,000 kilometers above a fixed point on the Equator, or approximately 39,000 kilometers from the Russian ICBM fields. Whichever weapon were used, however, this enormous range would make it virtually impossible to exploit the radiation from the booster's flame to accurately fix an aim point on the target. The resolution of any optical instrument, whether it is an observing telescope or a beam-focusing mirror, is limited by the phenomenon of diffraction. The smallest spot on which a mirror can focus a beam has a diameter that depends on the wavelength of the radiation, the aperture of the instrument and the distance to the spot. For infrared radiation from the booster's flame the wavelength would typically be one micrometer, so that targeting on a spot 50 centimeters across at a range of 39,000 kilometers would require a precisely shaped mirror 100 meters across—roughly the length of a football field. (For comparison, the largest telescope mirrors in the world today are on the order of five meters in diameter.)

The feasibility of orbiting a high-quality optical instrument of this stupendous size seems remote. The wavelengths used must be shortened, or the viewing must be reduced, or both. Accordingly it has been suggested that a geosynchronous defensive system might be augmented by other optical elements deployed in low orbits.

One such scheme that has been proposed calls for an array of ground-based excimer lasers designed to work in conjunction with orbiting optical elements. The excimer laser incorporates a pulsed electron beam to excite a mixture of gases such as xenon and chlorine into a metastable molecular state, which spontaneously reverts to the molecular ground state; the latter in turn immediately dissociates into two atoms, emitting the excess energy in the form of ultraviolet radiation at a wavelength of .3 micrometer.

Each ground-based excimer laser would send its beam to a geosynchronous mirror with a diameter of five meters, and the geosynchronous mirror would in turn reflect the beam toward an appropriate "fighting mirror" in low orbit. The fighting mirror would then redi-



GROUND-BASED LASER WEAPON with orbiting optical elements is designed to intercept ICBM's in boost phase. The excimer laser produces an intense beam of ultraviolet radiation at a wavelength of .3 micrometer. The ground-based mirror would send its beam to a five-meter geosynchronous mirror, which would in turn reflect the beam toward a similar fighting and viewing mirror in a comparatively low orbit; this mirror would then reflect the beam toward the rising booster, depending on its ability to form an image of the infrared radiation from the booster's exhaust plume to get a fix on the target (*diagram at left*). In order to compensate for fluctuations in the density

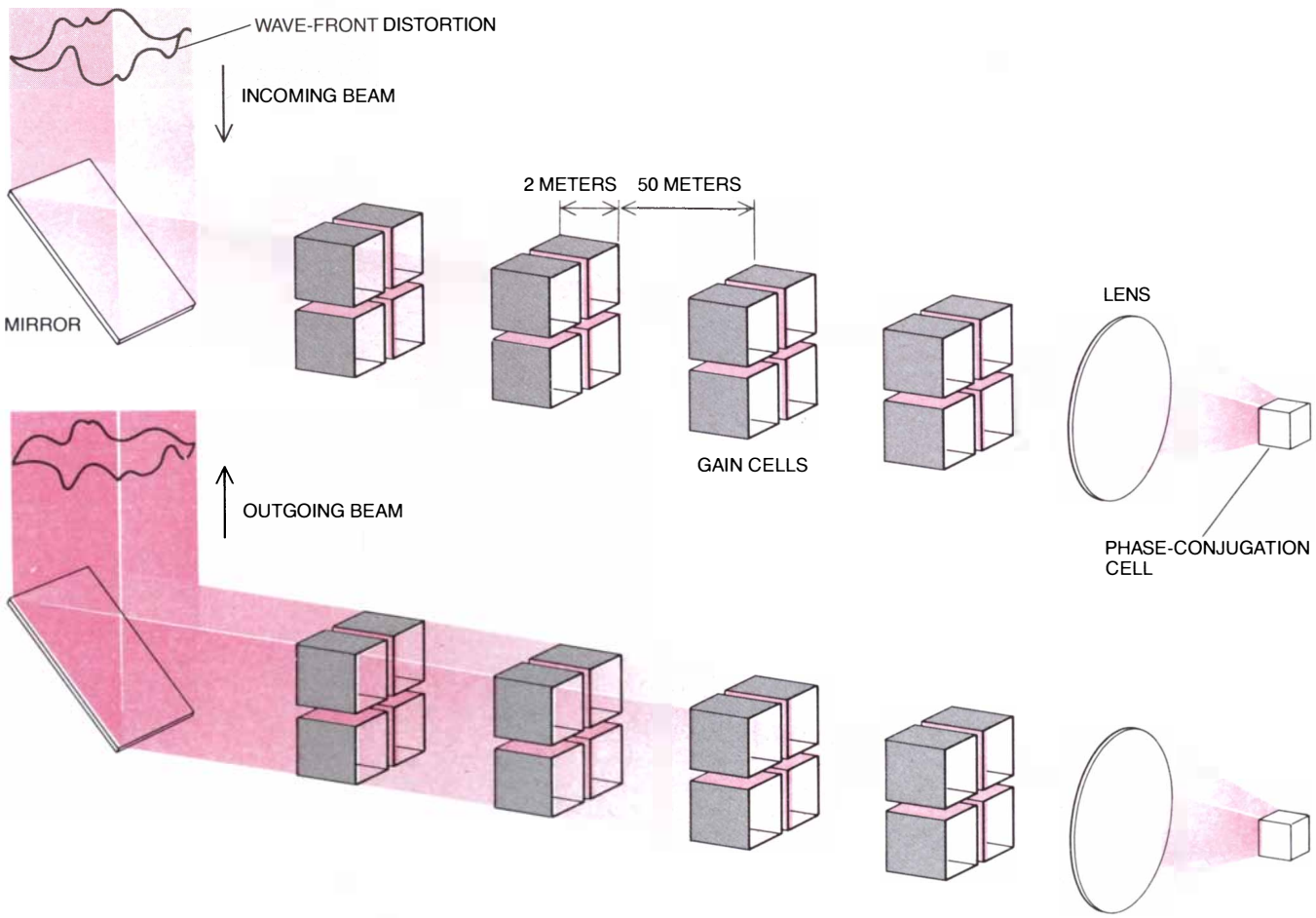
rect and concentrate the beam onto the rising booster rockets, depending on an accompanying infrared telescope to get an accurate fix on the boosters.

The main advantage of this scheme is that the intricate and heavy lasers, together with their substantial power supplies, would be on the ground rather than in orbit. The beam of any ground-based laser, however, would be greatly disturbed in an unpredictable way by ever present fluctuations in the density of the atmosphere, causing the beam to diverge and lose its effectiveness as a weapon. One of us (Garwin) has described a technique to compensate for these disturbances, making it possible, at least in principle, to intercept boosters by this scheme [*see illustration on these two pages*].

Assuming that such a system could be made to work perfectly, its power requirement can be estimated. Such an exercise is illuminating because it gives an impression of the staggering total cost of the system. Again information from the Fletcher panel provides the basis for our estimate. Apparently the "skin" of a booster can be "hardened" to withstand an energy deposition of 200 megajoules per square meter, which is roughly what is required to evaporate a layer of carbon three millimeters thick. With the aid of a geosynchronous mirror five meters in diameter and a fighting and viewing mirror of the same size, the beam of the excimer laser described above would easily be able to make a spot one meter across on the skin of a booster at a range of 3,000 kilometers from the fighting mirror; the resulting lethal dose would be about 160 megajoules.

A successful defense against an attack by the 1,400 ICBM's in the current Russian force would require a total energy deposition of 225,000 megajoules. (A factor of about 10 is necessary to compensate for atmospheric absorption, reflection losses at the mirrors and overcast skies.) If the time available for interception were 100 seconds and the lasers had an electrical efficiency of 6 percent, the power requirement would be more than the output of 300 1,000-megawatt power plants, or more than 60 percent of the current electrical generating capacity of the entire U.S. Moreover, this energy could not be extracted instantaneously from the national power grid, and it could not be stored by any known technology for instantaneous discharge. Special power plants would have to be built; even though they would need to operate only for minutes, an investment of $300 per kilowatt is a reasonable estimate, and so the outlay for the power supply alone would exceed $100 billion.

This partial cost estimate is highly optimistic. It assumes that all the boosters could be destroyed on the first shot, that the Russians would not have shortened the boost phase of their ICBM's, enlarged their total strategic-missile force or installed enough countermeasures to degrade the defense significantly by the time this particular defensive system was ready for deployment at the end of the century. Of course the cost of the entire system of lasers, mirrors, sensors



of the atmosphere the geosynchronous satellite would be equipped with a smaller excimer laser mounted on a 900-meter connecting arm ahead of the main mirror. A pulse of ultraviolet radiation from this laser would be directed at the ground-based laser, which would reverse the phase of the incoming beam and would emit a much more intense outgoing beam that would exactly precompensate for the atmospheric disturbance encountered by the incoming beam (*diagram at right*). The gain cells would be powered by pulsed electron beams synchronized with the outgoing beam. Such difficulties as mirror vulnerability must be resolved if such a device is ever to be effective.

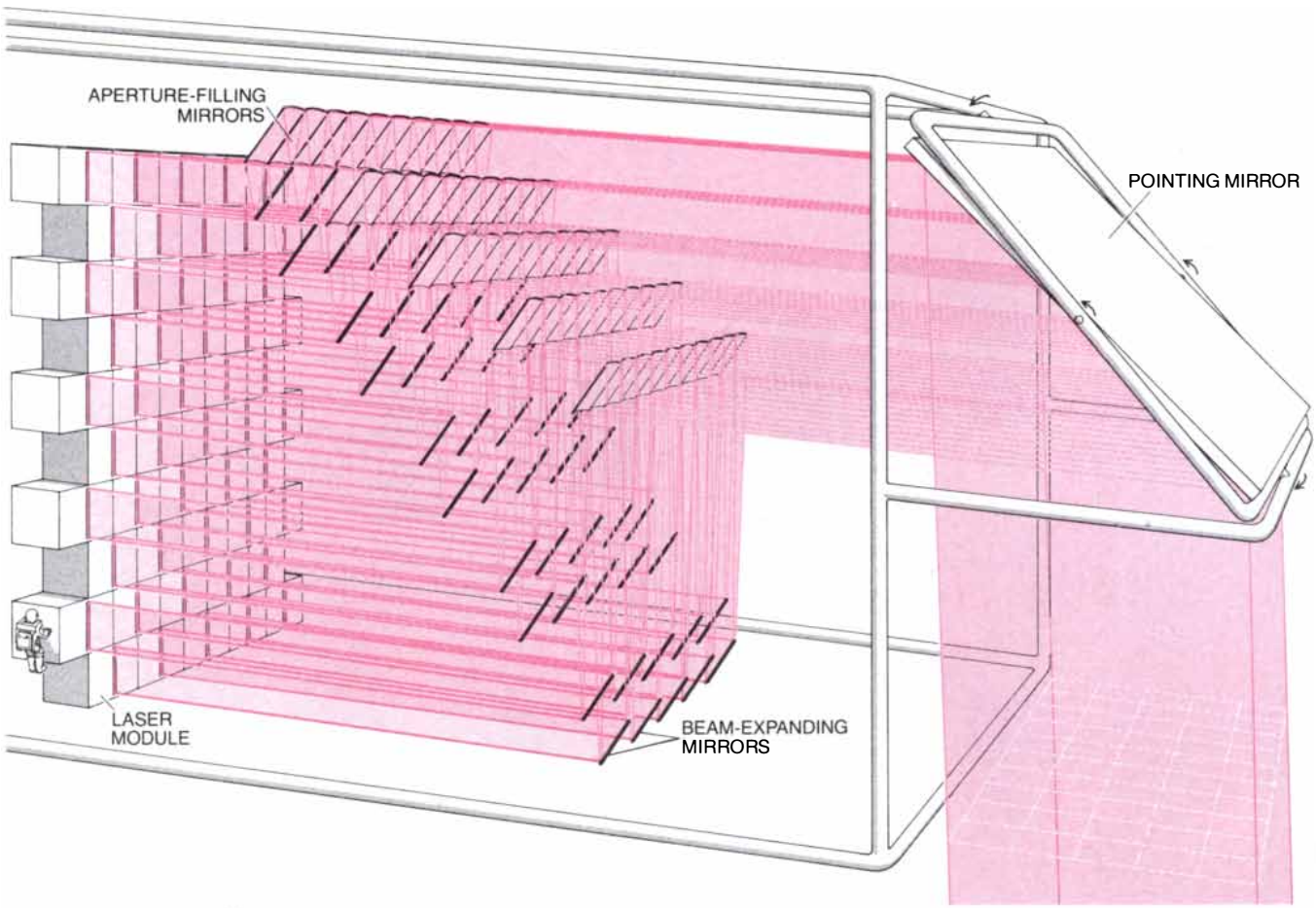
and computers would far exceed the cost of the power plant, but at this stage virtually all the required technologies are too immature to allow a fair estimate of their cost.

The exact number of mirrors in the excimer scheme depends on the intensity of the laser beams. For example, if the lasers could deliver a lethal dose of heat in just five seconds, one low-orbit fighting mirror could destroy 20 boosters in the assumed time of 100 seconds. It follows that 70 mirrors would have to be within range of the Russian silos to handle the entire attack, and each mirror would need to have a corresponding mirror in a geosynchronous orbit. If the distance at which a fighting mirror could focus a small enough spot of light was on the order of 3,000 kilometers, there would have to be about six mirrors in orbit elsewhere for every one "on station" at the time of the attack, for a total of about 400 fighting mirrors. This allowance for absenteeism is also optimistic, in that it assumes the time needed for targeting would be negligible, there would be no misses, the Russian countermeasures would be ineffective and excimer lasers far beyond the present state of the art could be built.

The second boost-phase interception scheme we shall consider is a pop-up system based on the X-ray laser, the only known device light enough to be a candidate for this role. As explained above, shortening the boost phase of the attacking missiles would negate any pop-up scheme. In this case a shortened boost phase would be doubly crippling, since the booster would stop burning within the atmosphere, where X rays cannot penetrate. Nevertheless, the X-ray laser has generated a good deal of interest, and we shall consider it here even though it would be feasible only if the Russians were to refrain from adapting their ICBM's to thwart this threat.

The X-ray laser consists of a cylindrical array of thin fibers surrounding a nuclear explosive. The thermal X rays generated by the nuclear explosion stimulate the emission of X-radiation from the atoms in the fibers. The light produced by an ordinary optical laser can be highly collimated, or directed, because it is reflected back and forth many times between the mirrors at the ends of the laser. An intense X-ray beam, however, cannot be reflected in this way, and so the proposed X-ray laser would emit a rather divergent beam; for example, at a distance of 4,000 kilometers it would make a spot about 200 meters across.

The U.S. research program on X-ray lasers is highly classified. According to a Russian technical publication, however, such a device can be expected to operate at an energy of about 1,000 electron volts. Such a "soft" X-ray pulse would be absorbed in the outermost fraction of a micrometer of a booster's skin, "blowing off" a thin surface layer. This would have two effects. First, the booster as a whole would recoil. The inertial-guidance system would presumably sense the blow, however, and it could still di-



ORBITING LASER WEAPON is shown in this highly schematic diagram based on several assumptions about the physical requirements of such a system. The weapon, which is designed to intercept ICBM's in boost phase from a comparatively low earth orbit, is scaled to generate a total of 25 megawatts of laser light at a wavelength of 2.7 micrometers from a bank of 50 chemical lasers, utilizing hydrogen fluoride as the lasing medium. The lasers, each of which occupies a cubic volume approximately two meters on a side, are arranged to produce an output beam with a square cross section 10 meters on a side. Assuming that the light from the entire bank of laser modules is in phase and that all the mirrors are optically perfect, it can be calculated that a weapon of this type could deliver a lethal dose of heat in seven seconds to a booster at a "kill radius" of some 3,000 kilometers. Some 300 such lasers would be needed in orbit to destroy the 1,400 ICBM's in the current Russian arsenal, assuming no countermeasures were taken other than "hardening" the missiles. Only the front of the weapon is shown; the fuel supply and other components would presumably be mounted behind the laser modules, to the left.

rect the warheads to their targets. Second, the skin would be subjected to an abrupt pressure wave that, in a careless design, could cause the skin to shear at its supports and damage the booster's interior. A crushable layer installed under the skin could prolong and weaken the pressure wave, however, thereby protecting both the skin and its contents.

Other interception schemes proposed for ballistic-missile defense include chemical-laser weapons, neutral-particle-beam weapons and nonexplosive homing vehicles, all of which would have to be stationed in low orbits.

The brightest laser beam attained so far is an infrared beam produced by a chemical laser that utilizes hydrogen fluoride. The U.S. Department of Defense plans to demonstrate a two-megawatt version of this laser by 1987. Assuming that 25-megawatt hydrogen-fluoride lasers and optically perfect 10-meter mirrors eventually become available, a weapon with a "kill radius" of 3,000 kilometers would be at hand. A total of 300 such lasers in low orbits could destroy 1,400 ICBM boosters in the absence of countermeasures if every component worked to its theoretical limit.

A particle-beam weapon could fire a stream of energetic charged particles, such as protons, that could penetrate deep into a missile and disrupt the semiconductors in its guidance system. A charged-particle beam, however, would be bent by the earth's magnetic field and therefore could not be aimed accurately at distant targets. Hence any plausible particle-beam weapon would have to produce a neutral beam, perhaps one consisting of hydrogen atoms (protons paired with oppositely charged electrons). This could be done, although aiming the beam would still present formidable problems. Interception would be possible only above the atmosphere at an altitude of 150 kilometers or more, since collisions with air molecules would disintegrate the atoms and the geomagnetic field would then fan out the beam. Furthermore, by using gallium arsenide semiconductors, which are about 1,000 times more resistant to radiation damage than silicon semiconductors, it would be possible to protect the missile's guidance computer from such a weapon.

Projectiles that home on the booster's flame are also under discussion. They have the advantage that impact would virtually guarantee destruction, whereas a beam weapon would have to dwell on the fast-moving booster for some time. Homing weapons, however, have two drawbacks that preclude their use as boost-phase interceptors. First, they move at less than .01 percent of the speed of light, and therefore they would have to be deployed in uneconomically large numbers. Second, a booster that burned out within the atmosphere would be immune to them, since friction with the air would blind their homing sensors.

That such a homing vehicle can indeed destroy an object in space was demonstrated by the U.S. Army in its current Homing Overlay test series. On June 10 a projectile launched from Kwajalein Atoll in the Pacific intercepted a dummy Minuteman warhead at an altitude of more than 100 miles. The interceptor relied on a homing technique similar to that of the Air Force's aircraft-launched antisatellite weapon. The debris from the collision was scattered over many tens of kilometers and was photographed by tracking telescopes [*see illustration on next two pages*]. The photographs show, among other things, the difficulty of evading a treaty that banned tests of weapons in space.

In an actual ballistic-missile-defense system such an interceptor might have a role in midcourse defense. It would have to be guided to a disguised reentry vehicle hidden in a swarm of decoys and other objects designed to confuse its infrared sensors. The potential of this technique for midcourse interception remains to be demonstrated, whereas its potential for boost-phase interception is questionable in view of the considerations mentioned above. On the other hand, a satellite is a larger and more fragile target than a reentry vehicle, and so the recent test shows the U.S. has a low-altitude antisatellite capability at least equivalent to the U.S.S.R.'s.

The importance of countermeasures in any consideration of ballistic-missile defense was emphasized recently by Richard D. DeLauer, Under Secretary of Defense for Research and Engineering. Testifying on this subject before the House Armed Services Committee, DeLauer stated that "any defensive system can be overcome with proliferation and decoys, decoys, decoys, decoys."

One extremely potent countermeasure has already been mentioned, namely that shortening the boost phase of the offensive missiles would nullify any boost-phase interception scheme based on X-ray lasers, neutral-particle beams or homing vehicles. Many other potent countermeasures that exploit existing technologies can also be envisioned. All of them rely on generic weaknesses of the defense. Among these weaknesses four stand out: (1) Unless the defensive weapons were cheaper than the offensive ones, any defense could simply be overwhelmed by a missile buildup; (2) the defense would have to attack every object that behaves like a booster; (3) any space-based defensive component would be far more vulnerable than the ICBM's it was designed to destroy; (4) since the booster, not the flame, would be the target, schemes based on infrared detection could be easily deceived.

Countermeasures can be divided into three categories: those that are threatening, in the sense of manifestly increasing the risk to the nation deploying the defensive system; those that are active, in the sense of attacking the defensive system itself, and those that are passive, in the sense of frustrating the system's weapons. These distinctions are politically and psychologically significant.

The most threatening response to a ballistic-missile-defense system is also the cheapest and surest: a massive buildup of real and fake ICBM's. The deployment of such a defensive system would violate the ABM Treaty, almost certainly resulting in the removal of all negotiated constraints on offensive missiles. Therefore many new missile silos could be constructed. Most of them could be comparatively inexpensive fakes arrayed in clusters about 1,000 kilometers across to exacerbate the satellites' absentee problem. The fake silos could house decoy ICBM's—boosters without expensive warheads or guidance packages—that would be indistinguishable from real ICBM's during boost phase. An attack could begin with a large proportion of decoys and shift to real ICBM's as the defense exhausted its weapons.

All space systems would be highly vulnerable to active countermeasures. Few targets could be more fragile than a large, exquisitely made mirror whose performance would be ruined by the slightest disturbance. If an adversary were to put a satellite into the same orbit as that of the antimissile weapon but moving in the opposite direction, the relative velocity of the two objects would be about 16 kilometers per second, which is eight times faster than that of a modern armor-piercing antitank projectile. If the satellite were to release a swarm of one-ounce pellets, each pellet could penetrate 15 centimeters of steel (and much farther if it were suitably shaped). Neither side could afford to launch antimissile satellites strong enough to withstand such projectiles. Furthermore, a large number of defensive satellites in low or geosynchronous orbits could be attacked simultaneously by "space mines": satellites parked in orbit near their potential victims and set to explode by remote control or when tampered with.

Passive countermeasures could be used to hinder targeting or to protect the booster. The actual target would be several meters above the flame, and the defensive weapon would have to determine the correct aim point by means of an algorithm stored in its computer. The aim point could not be allowed to drift by more than a fraction of a meter, because the beam weapon would have to dwell on one spot for at least several

seconds as the booster moved several tens of kilometers. Aiming could therefore be impeded if the booster flame were made to fluctuate in an unpredictable way. This effect could be achieved by causing additives in the propellant to be emitted at random from different nozzles or by surrounding the booster with a hollow cylindrical "skirt" that could hide various fractions of the flame or even move up and down during boost phase.

Booster protection could take different forms. A highly reflective coating kept clean during boost phase by a strippable foil wrapping would greatly reduce the damaging effect of an incident laser beam. A hydraulic cooling system or a movable heat-absorbing ring could protect the attacked region at the command of heat sensors. Aside from shortening the boost phase the attacking nation could also equip each booster with a thin metallic sheet that could be unfurled at a high altitude to absorb and deflect an X-ray pulse.

Finally, as DeLauer has emphasized, all the proposed space weapons face formidable systemic problems. Realistic testing of the system as a whole is obviously impossible and would have to depend largely on computer simulation. According to DeLauer, the battle-management system would face a task of prodigious complexity that is "expected to stress software-development technology"; in addition it would have to "operate reliably even in the presence of disturbances caused by nuclear-weapons effects or direct-energy attack." The Fletcher panel's report states that the "*survivability of the system components is a critical issue whose resolution requires a combination of technologies and tactics that remain to be worked out.*" Moreover, nuclear attacks need not be confined to the battle-management system. For example, airbursts from a precursor salvo of SLBM's could produce atmospheric disturbances that would cripple an entire defensive system that relied on the ground-based laser scheme.

Spokesmen for the Reagan Administration have stated that the Strategic Defense Initiative will produce a shift to a "defense-dominated" world. Unless the move toward ballistic-missile defense is coupled with deep cuts in both sides' offensive forces, however, there will be no such shift. Such a coupling would require one or both of the following conditions: a defensive technology that was so robust and cheap that countermeasures or an offensive buildup would be futile, or a political climate that would engender arms-control agreements of unprecedented scope. Unfortunately neither of these conditions is in sight.

What shape, then, is the future likely to take if attempts are made by the U.S. and the U.S.S.R. to implement a space-based system aimed at thwarting a nuclear attack? Several factors will have a significant impact. First, the new technologies will at best take many years to develop, and, as we have argued, they will remain vulnerable to known countermeasures. Second, both sides are currently engaged in "strategic modernization" programs that will further enhance their already awesome offensive forces. Third, in pursuing ballistic-missile defense both sides will greatly increase their currently modest antisatellite capabilities. Fourth, the ABM Treaty, which is already under attack, will fall by the wayside.

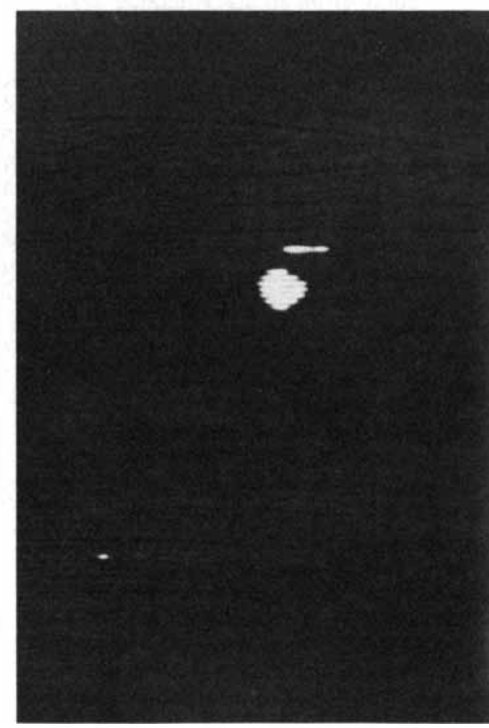
These factors, acting in concert, will accelerate the strategic-arms race and simultaneously diminish the stability of the deterrent balance in a crisis. Both superpowers have always been inordinately sensitive to real and perceived shifts in the strategic balance. A defense that could not fend off a full-scale strategic attack but might be quite effective against a weak retaliatory blow following an all-out preemptive strike would be particularly provocative. Indeed, the leaders of the U.S.S.R. have often stated that any U.S. move toward a comprehensive ballistic-missile-defense system would be viewed as an attempt to gain strategic superiority, and that no effort would be spared to prevent such an outcome. It would be foolhardy to ignore these statements.

The most likely Russian response to a U.S. decision to pursue the president's Strategic Defense Initiative should be expected to rely on traditional military "worst case" analysis; in this mode of reasoning one assigns a higher value to the other side's capabilities than an unbiased examination of the evidence would indicate, while correspondingly undervaluing one's own capabilities. In this instance the Russians will surely overestimate the effectiveness of the U.S. ballistic-missile defense and arm accordingly. Many near-term options would then be open to them. They could equip their large SS-18 ICBM's with decoys and many more warheads; they could retrofit their deployed ICBM's with protective countermeasures; they could introduce fast-burn boosters; they could deploy more of their current-model ICBM's and sea-launched cruise missiles. The latter developments would be perceived as unwarranted threats by U.S. military planners, who would be quite aware of the fragility of the nascent U.S. defensive system. A compensating U.S. buildup in offensive missiles would then be inevitable. Indeed, even if both sides bought identical defensive systems from a third party, conservative military analysis would guarantee an accelerated offensive-arms race.

Once one side began to deploy space-based antimissile beam weapons the level of risk would rise sharply. Even if the other side did not overrate the system's antimissile capability, it could properly view such a system as an immediate threat to its strategic satellites. A strategy of "launch on warning" might then seem unavoidable, and attempts might also be made to position space mines alongside the antimissile weapons. The last measure might in itself trigger a conflict since the antimissile system should be able to destroy a space mine at a considerable distance if it has any capability for its primary mission. In short, in a hostile political climate even a well-intentioned attempt to create a strategic defense could provoke war, just as the mobilizations of 1914 precipitated World War I.

Even if the space-based ballistic-missile defense did not have a cataclysmic birth, the successful deployment of such a defense would create a highly unstable strategic balance. It is difficult to imagine a system more likely to induce catastrophe than one that requires critical decisions by the second, is itself untested and fragile and yet is threatening to the other side's retaliatory capability.

In the face of mounting criticism Administration spokesmen have in recent months offered less ambitious rationales for the Strategic Defense Initiative than the president's original formulation.



SUCCESSFUL INTERCEPTION of a ballistic-missile warhead was achieved on June 10 in the course of the U.S. Army's Homing Overlay Experiment. The target was a dummy warhead mounted on a Minuteman ICBM that was launched from Vandenberg Air Force Base in California. The interceptor was a nonexplosive infrared-homing vehicle that was

One theme is that the program is just a research effort and that no decision to deploy will be made for many years. Military research programs are not normally announced from the Oval Office, however, and there is no precedent for a $26-billion, five-year military-research program without any commitment to deployment. A program of this magnitude, launched under such auspices, is likely to be treated as an essential military policy by the U.S.S.R. no matter how it is described in public.

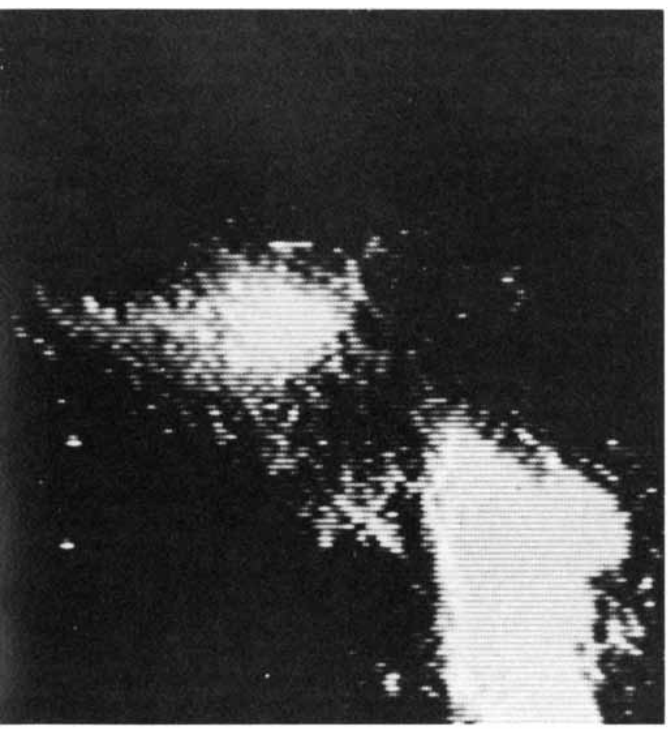
Another more modest rationale of the Strategic Defense Initiative is that it is intended to enhance nuclear deterrence. That role, however, would require only a terminal defense of hard targets, not weapons in space. Finally, it is contended that even an imperfect antimissile system would limit damage to the U.S.; the more likely consequence is exactly the opposite, since it would tend to focus the attack on cities, which could be destroyed even in the face of a highly proficient defense.

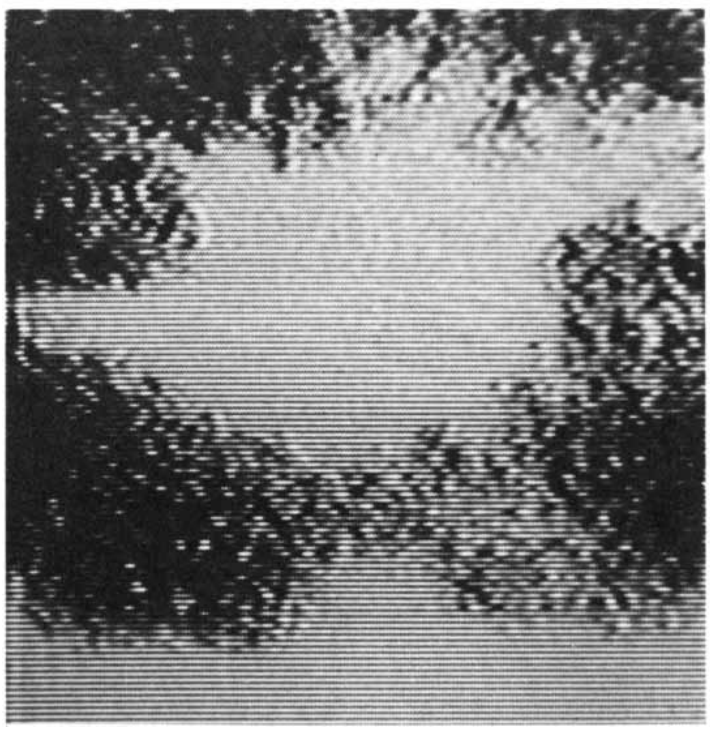
In a background report titled *Directed Energy Missile Defense in Space,* released earlier this year by the Congressional Office of Technology Assessment, the author, Ashton B. Carter of the Massachusetts Institute of Technology, a former Defense Department analyst with full access to classified data on such matters, concluded that "the prospect that emerging 'Star Wars' technologies, when further developed, will provide a perfect or near-perfect defense system . . . is so remote that it should not serve as the basis of public expectation or national policy." Based on our assessment of the technical issues, we are in complete agreement with this conclusion.

In our view the questionable performance of the proposed defense, the ease with which it could be overwhelmed or circumvented and its potential as an antisatellite system would cause grievous damage to the security of the U.S. if the Strategic Defense Initiative were to be pursued. The path toward greater security lies in quite another direction. Although research on ballistic-missile defense should continue at the traditional level of expenditure and within the constraints of the ABM Treaty, every effort should be made to negotiate a bilateral ban on the testing and use of space weapons.

It is essential that such an agreement cover all altitudes, because a ban on high-altitude antisatellite weapons alone would not be viable if directed-energy weapons were developed for ballistic-missile defense. Once such weapons were tested against dummy boosters or reentry vehicles at low altitude, they would already have the capability of attacking geosynchronous satellites without testing at high altitude. The maximum energy density of any such beam in a vacuum is inversely proportional to the square of the distance. Once it is demonstrated that such a weapon can deliver a certain energy dose in one second at a range of 4,000 kilometers, it is established that the beam can deliver the same dose at a range of 36,000 kilometers in approximately 100 seconds. Since the beam could dwell on a satellite indefinitely, such a device could be a potent weapon against satellites in geosynchronous orbits even if it failed in its ballistic-missile-defense mode.

As mentioned above, the U.S. interception of a Minuteman warhead over the Pacific shows that both sides now have a ground-based antisatellite weapon of roughly equal capability. Hence there is no longer an asymmetry in such antisatellite weapons. Only a lack of political foresight and determination blocks the path to agreement. Such a pact would not permanently close the door on a defense-dominated future. If unforeseen technological developments were to take place in a receptive international political climate in which they could be exploited to provide greater security than the current condition of deterrence by threat of retaliation provides, the renegotiation of existing treaties could be readily achieved.





launched 20 minutes later from Kwajalein Atoll in the western Pacific. This sequence of photographs was made from a video display of the interception recorded through a 24-inch tracking telescope on Kwajalein. The first frame shows the rocket plume from the homing vehicle a fraction of a second before its collision with the target above the atmosphere. The short horizontal bar above the image of the plume is a tracking marker. The smaller point of light at the lower left is a star. The second and third frames show the spreading clouds of debris from the two vehicles moments after the collision. Within seconds more than a million fragments were strewn over an area of 40 square kilometers. Just before the collision the homing vehicle had deployed a weighted steel "net" 15 feet across to increase the chances of interception; as it happened, the vehicle's infrared sensor guided it to a direct, body-to-body collision with the target. According to the authors, the test demonstrates that the U.S. now has a low-altitude antisatellite capability at least equivalent to that of the U.S.S.R.

Prions

These agents of infectious disease present a biological conundrum: Prions contain protein and reproduce in the living cell, yet no DNA or RNA has been found in them. What is the nature of their genome?

by Stanley B. Prusiner

The nucleic acids DNA and RNA would seem to be the lowest common denominator of life. Living organisms are exceedingly diverse, both as species and as individuals, but they are all alike in having a genome of nucleic acid. In bacteria as in human beings it is DNA that specifies the structure of enzymes and other proteins and thereby determines form and development; the replication of the DNA is the crucial event in reproduction. Even viruses, which cannot reproduce independently and whose status as living organisms is therefore questionable, take their identity from a molecule of DNA or RNA. The viroids, which transmit certain plant diseases, consist of nothing but RNA. The principle that genetic information invariably flows from nucleic acids to proteins has been called the central dogma of molecular biology.

It now appears that an infectious agent named a prion may stand out as a remarkable exception to the rule that every organism carries nucleic acids defining its own identity. The prion is known to be capable of initiating the production of new prions, at least in certain mammalian cells. Moreover, among the molecular components of the prion there is at least one protein, and so one would expect to find a DNA or RNA template specifying the structure of the protein. The evidence gathered so far, however, indicates the prion has no nucleic acid at all. Even if some DNA or RNA is ultimately found in the prion, there is probably not enough to encode the structure of the protein. From these facts it does not necessarily follow that the prion violates the central dogma—the latest results favor less heretical hypotheses—but there is little question its mode of reproduction is highly unusual.

Two diseases are known to be caused by prions. They are scrapie, a neurological disorder of sheep and goats, and Creutzfeldt-Jakob disease, a rare human dementia that recently came to public notice when it was identified as the cause of the death of the choreographer George Balanchine. Prions are also considered the probable agents of two other human diseases of the central nervous system: kuru, which has been observed only among highland tribes of New Guinea, and Gerstmann-Sträussler syndrome. Prions are listed among the possible causes of several other diseases, but the evidence remains circumstantial. Perhaps the most important disease in this third category is Alzheimer's disease, the commonest form of senile dementia and the fourth leading cause of death in the U.S.

Scrapie and the other diseases in which prions are implicated are classified as "slow infections." They are characterized by a prolonged incubation period of months, years or possibly decades in which the patient or the host animal is without symptoms; once the illness begins, however, it progresses steadily and generally leads to death.

The name scrapie comes from the tendency of sheep that have the disease to scrape off much of their wool. The disorder has been known for more than 200 years, and by 1935 French workers had shown convincingly that it can be transmitted from one sheep to another by inoculation. The demonstration of transmissibility implies that there is an infectious agent capable of reproducing itself in the host animal.

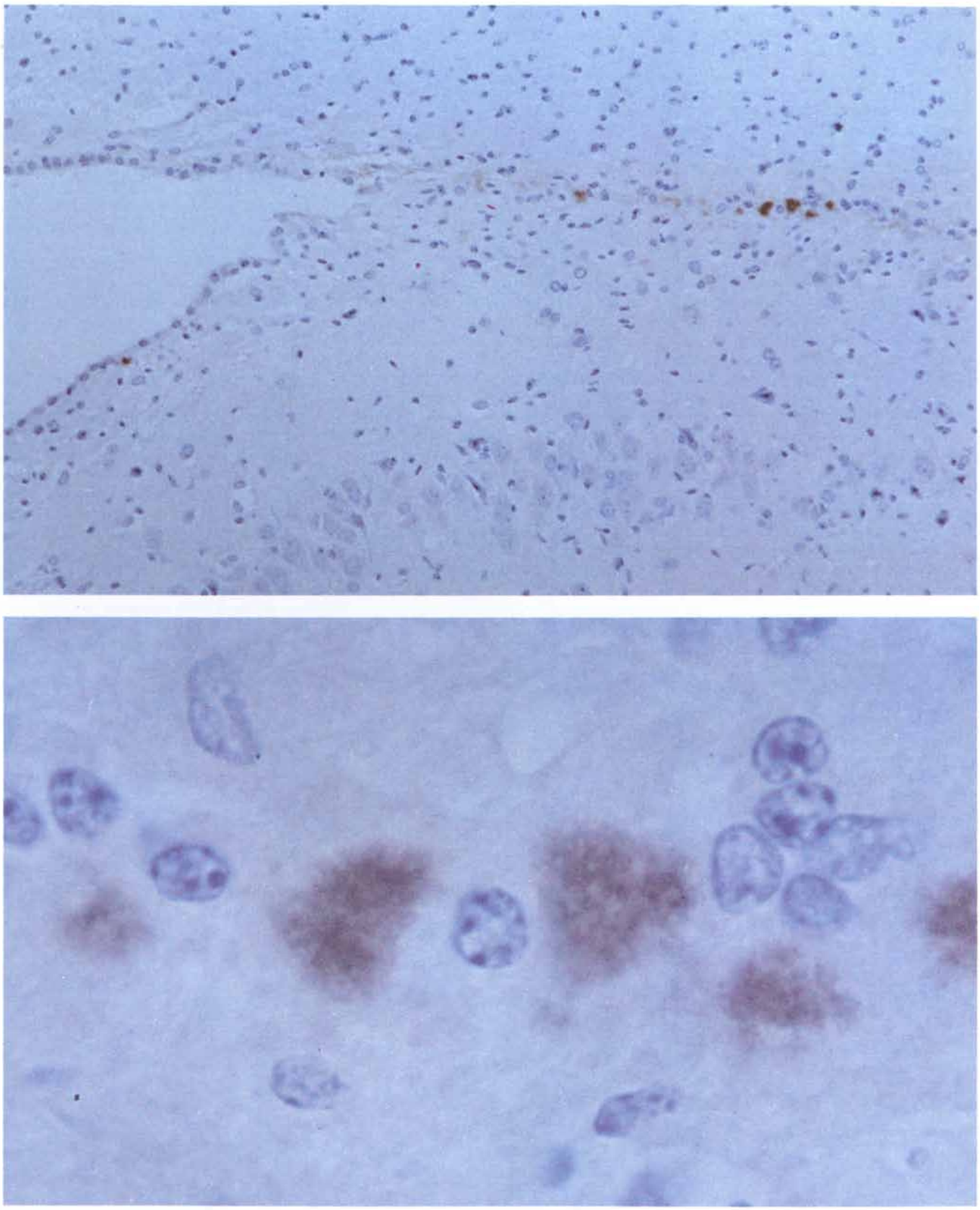
Scrapie remained an obscure veterinary disorder until 1959, when William J. Hadlow of the Rocky Mountain Laboratory of the National Institute of Allergy and Infectious Diseases suggested that scrapie and kuru might be related. The first medical descriptions of kuru had been published just two years earlier by Vincent Zigas and D. Carleton Gajdusek of the National Institute of Neurological and Communicative Disorders and Stroke. They had discovered the disease in New Guinea tribes said to practice a form of ritual cannibalism in which the brain of a deceased relative is eaten as an act of homage. The practice has since ended, and the incidence of kuru has diminished.

In 1966 Gajdusek, Clarence J. Gibbs, Jr., and Michael P. Alpers reported the transmission of kuru to apes. Two years later Gajdusek and Gibbs showed that Creutzfeldt-Jakob disease also can be transmitted to apes. More recently Colin L. Masters, Gajdusek and Gibbs have shown that Gerstmann-Sträussler syndrome is transmissible to apes and monkeys.

The clinical and pathological signs of scrapie, kuru, Creutzfeldt-Jakob disease and Gerstmann-Sträussler syndrome suggest they are all closely related. The initial symptoms of scrapie, kuru and Gerstmann-Sträussler syndrome are difficulty in walking and loss of coordination, indicating a functional disorder of the cerebellum. In kuru, dementia appears late in the course of the illness. Creutzfeldt-Jakob disease generally begins as a dementia, although a few cases show early signs similar to those of kuru. In none of the diseases is there any sign of an inflammatory process or fever, and counts of cells in the cerebrospinal fluid remain normal. These last observations are evidence that the immune system does not respond to the disease agent.

Pathological changes brought on by the diseases are confined to the central nervous system. A consistent indicator is abnormal proliferation of the astrocytes, a class of supporting cells in the brain. In neurons there is a depletion of dendritic spines, which have a role in the transmission of nerve impulses. In some of the disorders numerous vacuoles give the brain tissue a spongy appearance. Amyloid plaques, which are deposits of material with a fibrillar structure, have been observed in many cases of the diseases, but they do not seem to be a constant or obligatory feature.

In the past two decades considerable effort has been given to identifying the cause of scrapie. The unusual structure and properties of the agent and the slow and laborious biological assays needed to measure its concentration have impeded the work. By 1975 more than a dozen hypotheses had been pro-



PRIONS IN THE BRAIN of a hamster are identified by an immunological staining technique. The hamster had been infected with scrapie, the prototypical prion disease, which in nature affects sheep and goats. After an incubation period of roughly two months a section of brain tissue was exposed to antibodies with a specific affinity for a protein called PrP, the major constituent of the prion, and possibly the only constituent. The antibodies in turn were labeled with the enzyme peroxidase, which catalyzes the conversion of a colorless reagent into a dark stain. Thus when the reagent was applied to the tissue, stained structures seen by microscopic examination revealed the presence of prions. In the upper photomicrograph several stained bodies are visible; in the lower micrograph a few of the bodies are seen at greater magnification. They are tufts with a fluffy texture and are thought to be aggregations of prion "rods"; each rod is a condensation of perhaps 1,000 PrP molecules. Antibodies to PrP and prion rods have become available only in the past six months. The photographs were made by the author and his colleagues at the School of Medicine of the University of California at San Francisco.

posed for the nature of the scrapie agent; indeed, there were more hypotheses than there were experimental groups working on the disease.

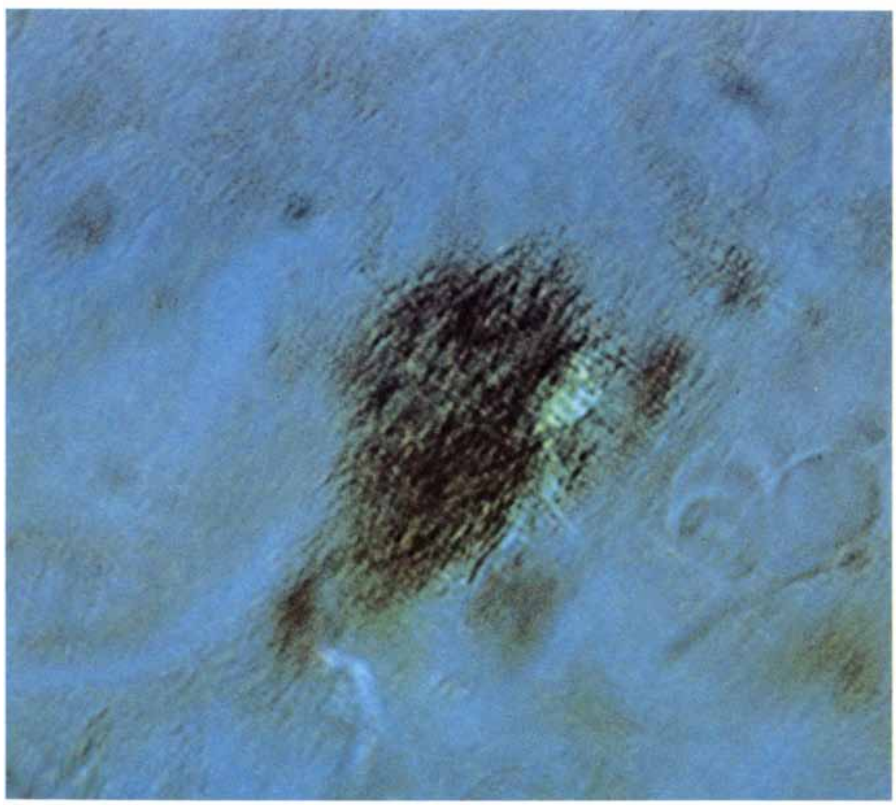
In trying to isolate an infectious agent whose structure and chemistry are unknown it is necessary to take an empirical approach. Typically a tissue sample from an infected animal is homogenized and then separated into fractions that differ in some physical or chemical property. The concentration of agent in each fraction is then assayed, and the purest fraction is singled out for further attention. In the case of scrapie the only way to measure the concentration of the agent has been by detecting its ability to induce the disease in animals.

For many years all such measurements had to be done by the method of end-point titration. Animals were inoculated with progressively more dilute specimens of material; the most dilute specimen capable of inducing the disease gave a measure of the concentration of the agent in the original material. In the early work with sheep and goats an entire herd of animals and several years of observation were needed to evaluate a single sample.

In 1960 Richard Chandler succeeded in transmitting scrapie to mice. End-point titrations in mice typically employed 10 dilutions, with each dilution being weaker than the one before by a factor of 10. Six mice would be inoculated at each dilution; those receiving a large dose would become ill in from four to five months, but in the animals given the most dilute solution capable of causing disease almost a year would pass before symptoms appeared. Hence 60 mice would have to be kept for a year before the end point could be determined. Although end-point titration in mice was an improvement over work in sheep and goats, this method of measuring the scrapie agent was still slower and more cumbersome than the methods used by Pasteur in his studies of viruses almost a century earlier.

In 1978 my colleagues and I found an alternative to end-point titration. Three years earlier Richard Marsh of the University of Wisconsin at Madison and Richard H. Kimberlin had described a form of scrapie in hamsters whose onset is about twice as fast as it is in mice. Studying the hamster disease, we found strong correlations between the concentration of scrapie agent and the rapidity of disease onset and between concentration and time of death. Thus instead of determining how much a sample could be diluted and still cause disease, we measured how fast a sample with a known dilution brought on disease symptoms and caused death.

The assay based on incubation times has been found to give an accurate measure of concentration for samples with a high titer of the agent. The gains in speed and economy have had a profound effect: we estimate that they have accelerated our work by a factor of 100. Instead of observing 60 animals for a year, we can assay a sample with just four animals in 60 days.



COLLECTIONS OF PRION RODS stained with the dye Congo red closely resemble amyloid plaques, which are features of several human and animal diseases that may be caused by prion infection. Amyloid stained with Congo red exhibits the optical phenomenon of birefringence, appearing green and gold when it is viewed between polarizing filters. The photomicrograph shows a doubly stained section of hamster brain in which a collection of prion rods can be identified by the presence of peroxidase-labeled antibodies to PrP; the rods have also taken up Congo-red dye and show green birefringence. The tissue here and on the preceding page is from the subependymal region of the brain, near the hippocampus. It is an area where amyloid plaques have been seen in rodents with scrapie and in human beings with Alzheimer's disease.

My work on scrapie began as a collaboration at the Rocky Mountain Laboratory with Hadlow and the late Carl M. Eklund. Our initial efforts to purify and isolate the agent utilized material from the mouse spleen, which we analyzed by centrifugation and end-point titration in mice. A centrifuge separates the components of a mixture according to their size and density. In the method we adopted a sample was spun at a particular speed for a specified interval, then the bioassay was carried out to determine how much of the infectious agent had sedimented and how much had remained in the supernatant fluid. The procedure was done for a wide range of speeds and times in a series of experiments that took almost two years. When we had finished, we repeated the study to make certain our results were reproducible.

In these early studies the greatest purification attained was about a thirtyfold enrichment of scrapie agent. One of the factors limiting the degree of purification was also one of our major findings: the infectious particles were shown to be extremely heterogeneous in size and density. Judging from the rate of sedimentation in the centrifuge, some of the particles were almost as large as mitochondria or bacteria; others seemed to be substantially smaller than the smallest viruses. The broad range of sizes implied that the scrapie agent can exist in many molecular forms; the observations could be explained, for example, if very small infectious particles aggregate to form much larger clusters. In an attempt to make more homogeneous preparations, as well as to facilitate the separation of the agent from cellular molecules, we added detergents. Although the detergents we employed had little effect on the observed heterogeneity, they did aid in purification.

The effort to purify the scrapie agent continued in my laboratory at the School of Medicine of the University of California at San Francisco. The story of our endeavor can be told briefly, but it represents a decade of painstaking and

sometimes frustrating labor. A large number of people have made important contributions. Among them I should like to mention Richard Baringer, Ronald Barry, Paul Bendheim, David Bolton, Karen Bowman, Patricia Cochran, Steve DeArmond, Darlene Groth, Michael McKinley and Daniel Stites.

It was at this point that we switched from mice to hamsters; in addition to providing a faster assay, the hamster brain has a titer of scrapie agent 100 times greater than that of the mouse spleen. Our purification method again began with detergent extraction and centrifugation, but we added three more steps: exposure to nucleases, exposure to proteases and analysis by gel electrophoresis. We had found that the infectivity of the scrapie agent is unaltered by digestion with nucleases, which are enzymes that catalyze the breakdown of nucleic acids. Hence treating the preparation with a nuclease eliminated most cellular nucleic acids while leaving the scrapie agent intact. Proteases, which cut the chain of amino acids in a protein, were used in a similar way to remove extraneous proteins. The final step in our revised procedure, gel electrophoresis, separates molecules according to the rate at which they migrate through a gel under the influence of an applied electric field. The rate at which a molecule moves is determined primarily by its electric charge, although size and shape also have an effect.

The purification method culminating in electrophoresis gave an overall purification by a factor of about 100, which was enough to establish several vital facts. The most important result was a convincing demonstration that the biological activity of the scrapie agent depends on a protein. We therefore introduced the term prion, for "proteinaceous infectious particle."

DISEASE	CAUSED BY PRIONS?	NATURAL HOST SPECIES	EXPERIMENTAL HOST SPECIES	INCUBATION PERIOD
SCRAPIE	YES	SHEEP, GOATS	MICE, HAMSTERS, MONKEYS	2 MONTHS TO 2 YEARS OR MORE
CREUTZFELDT-JAKOB DISEASE	YES	HUMAN BEINGS	APES, MONKEYS, MICE, GOATS, GUINEA PIGS	4 MONTHS TO 20 YEARS OR MORE
KURU	PROBABLY	HUMAN BEINGS	APES, MONKEYS	18 MONTHS TO 20 YEARS OR MORE
GERSTMANN-STRÄUSSLER SYNDROME	PROBABLY	HUMAN BEINGS	APES, MONKEYS	18 MONTHS OR MORE
TRANSMISSIBLE MINK ENCEPHALOPATHY	PROBABLY	MINK	MONKEYS, GOATS, HAMSTERS	5 MONTHS TO 7 YEARS OR MORE
CHRONIC WASTING DISEASE	PROBABLY	MULE DEER, ELK	FERRETS	18 MONTHS OR MORE

PRION DISEASES are classified in two categories according to the strength of the evidence implicating prions in their causation. Scrapie is a prion disease by definition, and there is substantial experimental evidence that prions also bring on Creutzfeldt-Jakob disease. Prions are considered the probable cause of at least four more diseases, including two human neurological disorders. The four diseases are known to be transmissible, their symptoms resemble those of scrapie and Creutzfeldt-Jakob disease, and they all induce similar changes in brain tissue.

We have since returned to centrifugation as the primary method of purification, but the technique is somewhat different from the one employed in the early studies. The specimen is placed on top of a sucrose solution arranged to form a gradient of progressively higher density. When the centrifuge is spun, components of the specimen migrate through the gradient until they reach a level where their own density matches that of the surrounding solution.

Our first version of the sucrose-gradient separation gave a thousandfold purification of the scrapie agent. We were thereby able to demonstrate that the bulk of the protein consists of a single molecular species, which we designated PrP, for prion protein. A larger-scale version of the purification procedure was then developed, based on a centrifuge with a "zonal rotor," a large vessel in which substantial quantities of material could be separated in a sucrose gradient. This latest separation technique yields an enrichment of 5,000 times.

With the greater purification it was possible to show that the most infectious fractions of the centrifuged material include essentially one protein, namely PrP. Further studies of the protein by electrophoresis have shown that it has an apparent molecular weight of between 27,000 and 30,000. We now have reason to think this estimate is probably too high. We have found that PrP is a glycoprotein, that is, a protein in which sugars are bound to the amino acid units. Using electrophoresis to measure the molecular weight of a glycoprotein often gives an erroneously high result. In any event, PrP is a comparatively small protein, less than half the size of hemoglobin.

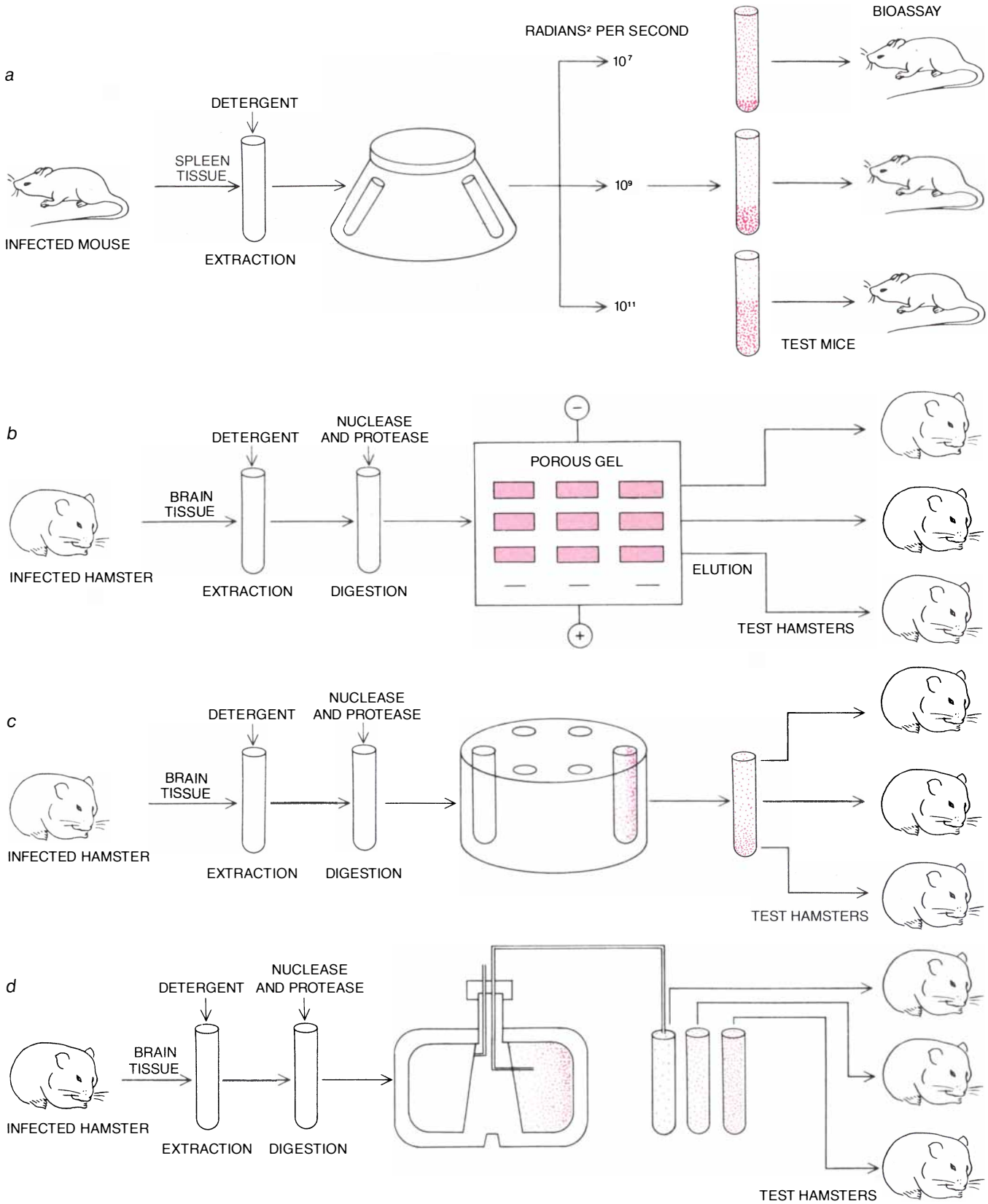
Electron micrographs of the purified material have revealed numerous rod-shaped particles much too large to be the individual prions. The rods had been observed in earlier experiments, but we had been unable to verify the obvious hypothesis that they are aggregates of prions; the possibility remained that they were not the disease agent itself but the product of some pathological change brought on by the disease. With our highly purified fractions we were able to establish that the rods are composed of PrP and hence can be considered prion aggregates.

The isolation of reasonably pure specimens in substantial quantity has proved to be the key to answering many puzzling questions about the prion. Quite recently it has enabled us to raise antibodies to prions in experimental animals, a feat we had been attempting for years without success. The crucial factor turned out to be the quantity of material injected into the animal, which we were able to increase by a factor of 10 (to roughly 100 micrograms). The availability of antibodies can be expected to work a great change in the pace and the technology of prion research. For example, it may greatly reduce the need for biological assays. The concentration of prions in a sample could be determined in hours rather than months by measuring the affinity of the antibodies for components of the material. The same affinity could be the basis of a new purification technique.

Study of the purer specimens has raised new questions as well as answering old ones. We have noted an extraordinary resemblance between dense collections of prion rods and the amyloid plaques seen in some cases of scrapie and Creutzfeldt-Jakob disease. It would seem reasonable to suppose the amyloid plaques are in fact deposits of prions in an aggregated state. The trouble is that amyloid plaques are also characteristic of disorders in which there is no strong reason to suspect a prion infection. The plaques are closely associated with Alzheimer's disease, which has not been shown to be transmissible.

Our conclusion that a protein is a component of the prion and is necessary for infectivity emerged directly from our work on purification. In the course of that work we examined numerous chemical reagents to determine whether or not they alter the infectivity of the prion. The essence of our findings is that substances known to disrupt proteins diminish prion infectivity. Substances that have no effect on proteins leave the infectivity unchanged.

The clearest evidence comes from ex-



PURIFICATION TECHNOLOGIES employed by the author and his colleagues in the past 10 years have yielded progressively more concentrated solutions of prions. Detergent extraction and centrifugation with a fixed-angle rotor (*a*) separated a specimen into fractions according to particle size and achieved a purification of 30 times. Gel electrophoresis (*b*) separates the components of a solution according to their electric charge and size. Electrophoresis in conjunction with detergent extraction and enzyme digestion yielded a hundredfold purification and led to the demonstration that a protein is essential for prion infectivity. In sucrose-gradient centrifugation molecules migrate through layers of sugar solution until they reach a level matching their own density. Small-scale experiments were done with a vertical-tube rotor (*c*). The resulting thousandfold purification led to the identification of PrP. Purifications by a factor of 5,000 are now being achieved by sucrose-gradient centrifugation in a zonal rotor (*d*), which can process much larger quantities. With more material of higher purity it has been possible to investigate the composition and structure of PrP and also to raise antibodies that specifically bind to it.

periments with proteases. A protease is highly specific: it has almost no effect on biological molecules other than proteins. In early studies with impure material, protease treatment gave equivocal results, but when enriched fractions became available, we were able to show convincingly that proteases reduce prion infectivity. Furthermore, the confusion over the early results has been resolved. It turns out that PrP is resistant to protease treatment compared with most cellular proteins. Consequently in an impure mixture with many other proteins most of the enzyme activity is directed toward competing substrates. We have exploited this effect in the purification of prions. Early in the procedure extraneous proteins can be digested with a protease without significantly degrading the prion protein.

Several other reagents damage proteins not by cutting the chain of amino acids but by unfolding it; the protein is said to be denatured. The detergent sodium dodecyl sulfate (SDS) is such a substance, and boiling a prion solution in SDS abolishes its infectivity. Phenol, urea and certain salts are also denaturing reagents; they too reduce the biological activity of prions.

All the reactions described above are irreversible. Another reagent, diethyl pyrocarbonate (DEP), modifies proteins chemically in a way that can be reversed by exposure to hydroxylamine. As in the case of proteases, DEP gave inconsistent results in early experiments with impure mixtures; we have since shown that it does diminish the titer of infectious agent in purified fractions. The infectivity can be restored by treatment with hydroxylamine.

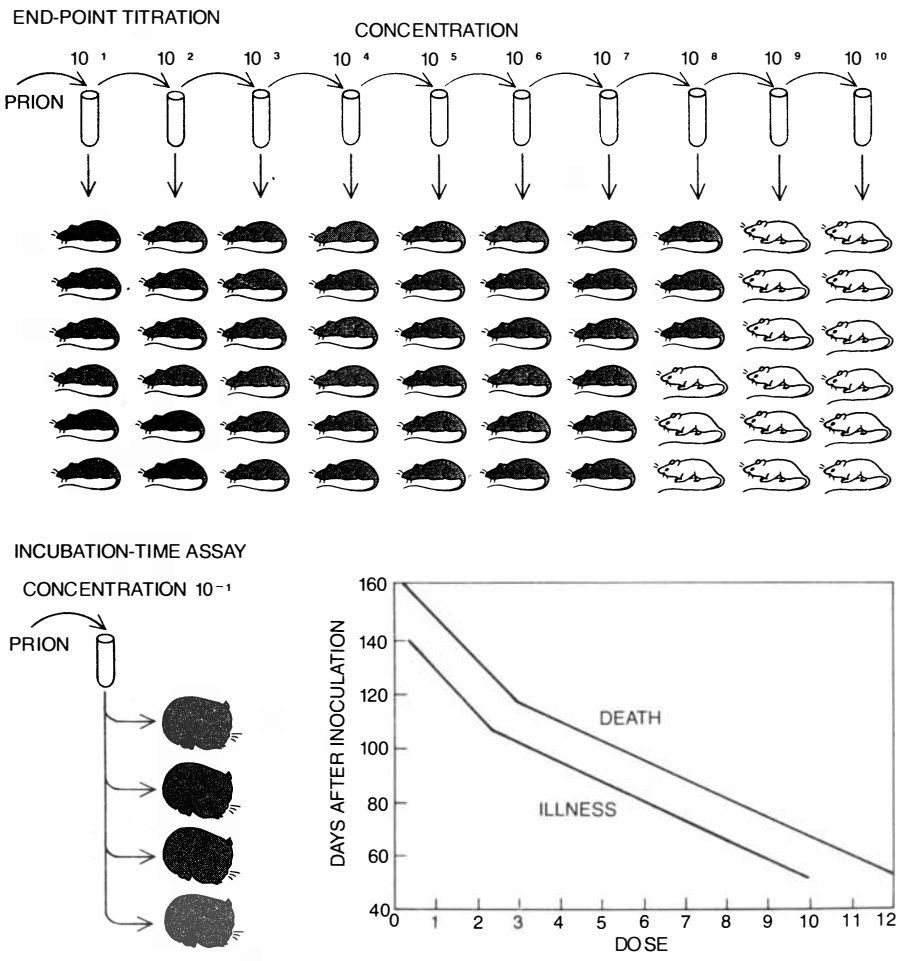
The presumed target of all these treatments is PrP. The possibility remained, however, that PrP was not a structural component of the prion but a pathological product of the scrapie infection. We therefore undertook a lengthy series of experiments, which we feel has established that the prion is indeed composed at least in part of PrP molecules.

The first line of evidence is simply that PrP has been found in every sample with a high titer of prions, whether it was prepared by centrifugation or by electrophoresis and even if it was purified from scrapie-infected tissue before the appearance of pathological changes. The concentration of PrP is directly proportional to the titer of prions. The rate of digestion by proteases is the same for PrP and for prions, and other experimental procedures that alter the prion titer also change the concentration of PrP. DEP, which we had already shown to be capable of inactivating prions, was found to bind directly to PrP. Finally, in purified material treated with protease, PrP is the only protein that can be detected, suggesting prions have just one major protein.

Having established the role of a protein in the infective process, we began to search for nucleic acid. The effort encountered consistent frustration. Almost 20 years ago the British investigators Tikvah Alper, David Haig and Michael Clarke irradiated crude homogenates of scrapie-infected tissue with both ultraviolet radiation and shorter-wavelength ionizing radiation. In general, ionizing radiation destroys cells and viruses through damage to nucleic acids; the probability of damage is roughly proportional to the size of the target molecules. Alper and her colleagues found that extremely high doses of radiation were needed to inactivate the scrapie agent. They concluded that the agent has no nucleic acid and is considerably smaller than a virus. Their conclusions were met with great skepticism, but in collaboration with James Cleaver of the School of Medicine of the University of California at San Francisco we have since repeated the ultraviolet-irradiation studies using purified preparations and have obtained similar results.

In many experiments we have attempted to inactivate prions by means of treatments that chemically attack nucleic acids. Experiments directly analogous to those with proteases employ a nuclease. We have exposed prion solutions to several nucleases, including enzymes that destroy both DNA and RNA, without detecting any significant diminution of prion infectivity.

One possible objection to such experiments is that the enzyme may not be able to gain access to the nucleic acids; many viruses are resistant to nucleases because the protein coat protects the DNA or RNA. In collaboration with John E. Hearst of the University of California at Berkeley we tried treating

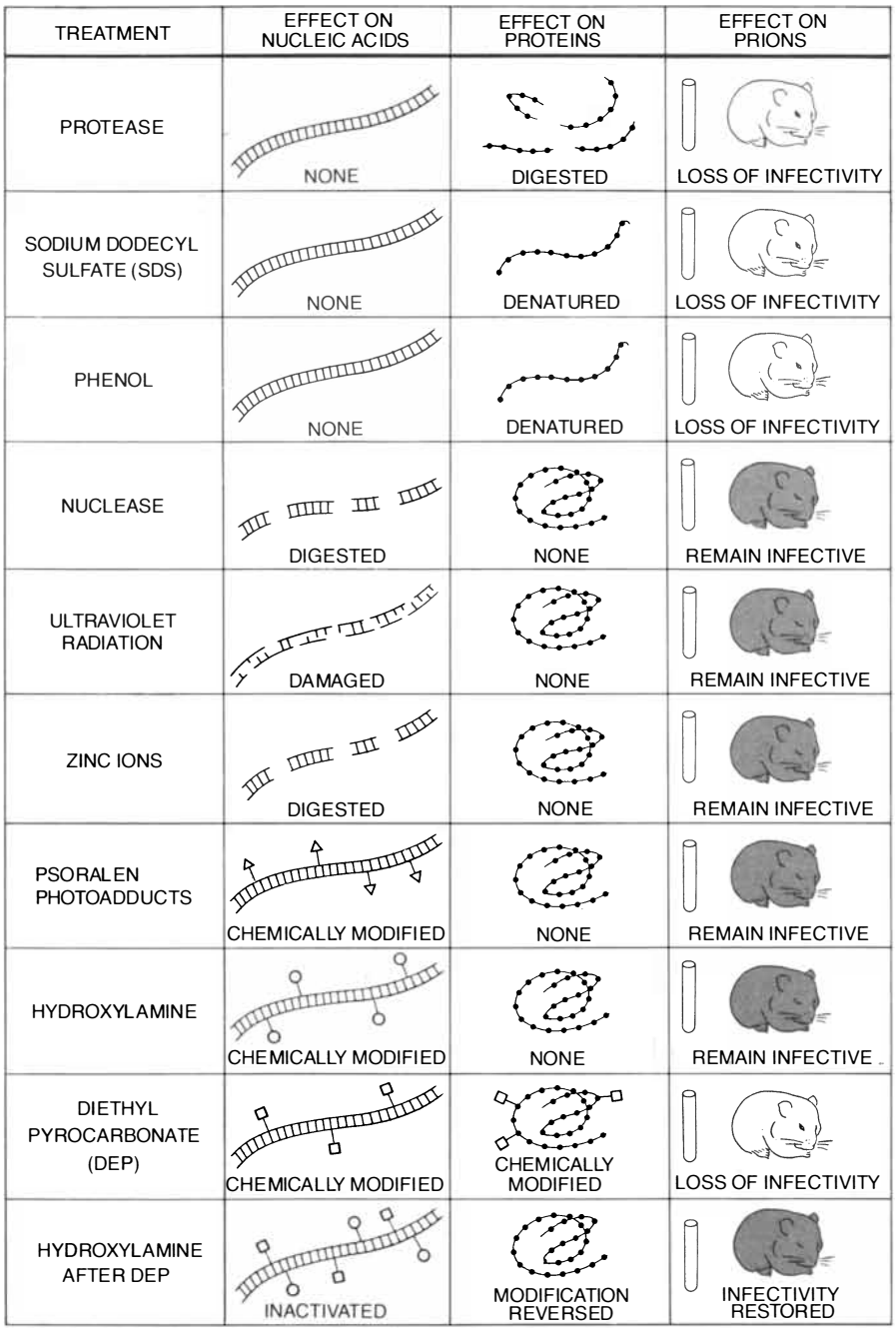


BIOLOGICAL ASSAYS OF INFECTIVITY have long been the only available methods of measuring the concentration of scrapie prions in a specimen. In the method of end-point titration 10 samples were prepared, with each sample being more dilute than the one before by a factor of 10. Each sample was injected into six mice, which were observed for up to a year for neurological signs of scrapie. The greatest dilution capable of causing disease gave a measure of the concentration of scrapie agent in the original specimen. The incubation-time method has reduced both the time and the number of animals needed to assay a specimen. It is based on the observation that the interval between inoculation and the first appearance of neurological signs as well as the interval between inoculation and death depends on the dose of agent injected. For highly purified samples measuring the time between inoculation and the onset of illness in four hamsters can indicate the concentration of scrapie agent in as little as 60 days.

prions with molecules called psoralens, which can pass through the protein coat of most viruses; on exposure to ultraviolet radiation the psoralens bind to the nucleic acid and inactivate it. Again we observed no loss of prion infectivity. Zinc ions, which catalyze the breakdown of RNA, were likewise ineffective.

Our work with DEP adds further evidence. DEP can inactivate nucleic acids as well as proteins, but only proteins can be restored to functionality by hydroxylamine. Hence the recovery of infectivity we observed after hydroxylamine treatment argues that DEP is not acting on a nucleic acid.

In collaboration with Theodor O. Diener of the U.S. Department of Agriculture we have compared the effects various reagents have on prions and on viroids. Because both entities seem to be much smaller than viruses, it was once thought prions might be similar to viroids in structure, that is, they might consist of "naked" RNA. Actually the two kinds of infectious agent are antithetical. Procedures that modify proteins inactivate prions but have no effect on viroids; treatments that attack nucleic acids destroy viroids but not prions.

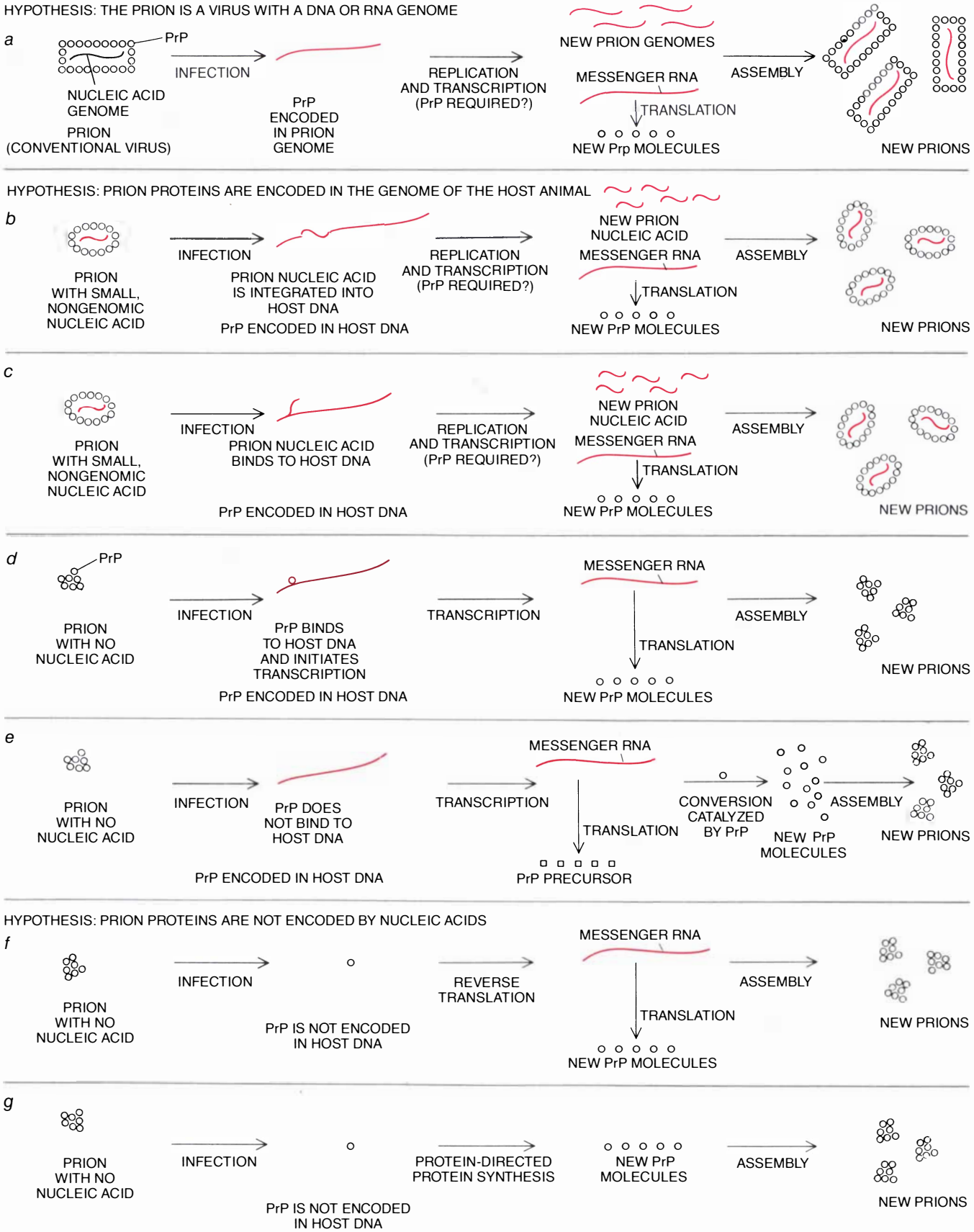


CHEMICAL NATURE OF THE PRION has been investigated by means of treatments that inactivate either proteins or nucleic acids. Proteases (enzymes that cut the chain of amino acids in a protein) can drastically reduce the infectivity of a purified solution of prions; so can certain detergents and other reagents that denature a protein by unfolding it. In contrast, nucleases (which digest the nucleic acids DNA and RNA) have no effect on prion infectivity. Prions are also highly resistant to ultraviolet radiation and zinc ions, which act mainly on nucleic acids. A nucleic acid in the prion might be protected by a coat of protein, but psoralens and hydroxylamine, which can penetrate the protein coat of many viruses, are also ineffective. The reagent diethyl pyrocarbonate (DEP) can modify both proteins and nucleic acids, but its action on proteins can be reversed by treatment with hydroxylamine, whereas its effect on nucleic acids is irreversible. Prions are inactivated by DEP and their infectivity is restored by hydroxylamine. All the chemical and physical evidence suggests the prion consists of a protein without nucleic acid, although the possibility of a small DNA or RNA cannot yet be excluded.

The question of the prion's size also bears on the nucleic acid issue. Individual prions seem to be very small. Therefore the amount of nucleic acid a prion could encapsulate is probably quite limited. The target-size studies done by Alper and her colleagues suggested the infectious scrapie particle might have a molecular weight of between 60,000 and 150,000. The remarkable heterogeneity of prions has made it difficult to determine the size of the smallest infectious particle by more direct methods. After attempting to break up aggregates of prions with detergents and other chemicals, we have investigated the size of the particles by sucrose-gradient centrifugation, by timing their passage through a chromatographic column and by passing them through a membrane filter with pores of a known size. All the studies have given results consistent with a molecular weight of between 50,000 and 100,000, but each method has potential pitfalls. Because of the many sources of uncertainty, the most that can be said for now is that the smallest infectious form of the prion may be 100 times smaller than the smallest viruses.

If the prion has a molecular weight of 50,000, its diameter would be about five nanometers, or five billionths of a meter. If it is constructed like a conventional virus, it might take the form of an approximately spherical shell of protein surrounding a core of nucleic acid. The shell could not be less than about a nanometer thick, which would leave room in the core for no more than about 12 nucleotides. Limits on the size of any prion nucleic acid can also be derived from other measurements. The prion's resistance to inactivation by ultraviolet radiation is consistent with a nucleic acid made up of from 12 to 50 nucleotides; our experiments with psoralens would not have detected a nucleic acid with fewer than 40 nucleotides.

The failure to detect a nucleic acid in prions cannot be taken as proof that it does not exist. It could still be concealed in some way by a surrounding structure or could be present in quantities too small to be detected. Nevertheless, it



PRION REPLICATION might be accomplished by any one of several hypothetical mechanisms. The possibility that the prion is a conventional virus with a DNA or RNA genome encoding the structure of PrP (*a*) is unlikely. In another hypothesis PrP is encoded by genes in the host animal. A small prion nucleic acid might be incorporated into the host genome (*b*) or might simply bind to the host DNA (*c*); in either case it could promote the expression of PrP genes. Alternatively, if the prion has no nucleic acid at all, PrP itself might stimulate gene expression by binding to the host DNA (*d*). The requirement of biologically active PrP molecules for the synthesis of new prion particles could also be explained if PrP catalyzes the conversion of a precursor molecule into PrP (*e*). Two further possibilities violate the "central dogma" of molecular biology, which holds that genetic information flows from nucleic acids to proteins. Reverse translation might create a nucleic acid sequence based on the sequence of amino acids in PrP (*f*), or PrP might be a template for its own synthesis (*g*).

seems reasonable to suggest that if the prion has any nucleic acid at all, it is likely to be less than 50 nucleotides long. In the standard genetic code three nucleotides are needed to specify each amino acid, and so the putative prion "genome" could not encode a protein with more than about a dozen amino acids. It should be noted that the molecular weight of PrP implies it has roughly 250 amino acids.

There is only one way to establish with certainty that the prion consists of nothing but protein: determine the complete sequence of amino acids in PrP, synthesize an artificial protein with the same sequence and then demonstrate that it has the same biological action as the natural protein. Surprisingly, in recent months the prospect of such a demonstration has begun to seem less remote, again because of the availability of purified preparations in large quantities. In collaboration with Leroy E. Hood of the California Institute of Technology and Stephen B. H. Kent of Molecular Genetics, Inc., we have identified the first 15 amino acids of PrP.

The partial sequence of PrP was determined by repeatedly employing a series of reagents that cut off the terminal amino acid of a protein, so that the amino acids in the chain are liberated one by one. In about half of the cleavage cycles we detected multiple amino acids, with one strong signal being accompanied by weaker ones. Initially we thought the minor signals indicated the existence of a spectrum of PrP molecules with slightly different amino acid sequences. We subsequently found, however, that the variant sequences differ from the major sequence only in their starting point. One minor sequence can be brought into alignment with the major sequence by moving the minor sequence forward four amino acid positions; in another case the minor sequence must be moved backward two units. The variations are observed because the PrP molecules have "ragged ends," presumably caused by treatment with proteases during purification.

Knowing even a small part of the amino acid sequence of PrP has opened up two important avenues of research. First, synthetic chains of amino acids that match the known PrP sequence have been constructed for use as antigens. Antibodies formed against the synthetic molecules can ultimately be employed in the purification of prions and in assays of their concentration as well as in probing their structure. Second, synthetic segments of nucleic acid that encode the known part of the amino acid sequence of PrP have been constructed. They are being used as probes for natural DNA that encodes PrP.

Surely the preeminent question about the prion is how it replicates. There are essentially three categories of possible replication mechanisms.

The first hypothesis is that the prion, in spite of all indications to the contrary, is a conventional virus with a genome of DNA or RNA that encodes the entire structure of the prion protein. On entering a cell the nucleic acid genome of the prion would be transcribed into messenger-RNA molecules that would serve as templates for the synthesis of prion proteins. The fact that a prion protein is obligatory for scrapie infection could be explained by assuming the prion is a "negative-strand virus," one in which the viral nucleic acid alone is not enough to cause disease; a viral protein, acting as an enzyme, is needed for transcription of the viral nucleic acid into messenger RNA. The virus hypothesis is quite unlikely in view of all the information we now have about the chemical and physical structure of prions.

In one sense the conventional-virus explanation is the most conservative hypothesis; the most radical idea is that the amino acids in PrP somehow specify their own sequence during prion replication. This might come about indirectly through the "reverse translation" of the protein into RNA or DNA, which would then be interpreted by the cellular apparatus in the usual way to make more protein. Such a process has never been observed, and it would be a clear violation of the central dogma, which again states that the flow of information in the cell is always from nucleotides to proteins.

One can also imagine a mechanism in which the amino acid sequence of PrP would serve directly as a template for the construction of a new protein molecule. Such protein-directed protein synthesis has never been seen, and enzymes capable of building a complex protein by this means are not known.

The third category of possible replication mechanisms includes the ones I consider the most plausible. In these schemes there does exist a DNA gene encoding the amino acid sequence of the prion protein. The gene is not carried by the prion, however; instead it is a component of the normal mammalian genome. Infection by prions would somehow activate the gene or perhaps alter it.

If the prion includes a small piece of nucleic acid, that could be the trigger for gene activation. This hypothetical small prion nucleic acid might be inserted into a host-cell chromosome just "upstream" of the PrP gene, or in other words just ahead of the point at which transcription of the gene begins. The inserted sequence could then serve as a promoter or enhancer of gene expression. Alternatively, if the prion consists exclusively of protein, PrP itself might bind to cellular DNA in a region that controls the transcription of the PrP gene. Most proteins that bind to DNA tend to repress gene expression, but the phenomenon of a protein that stimulates its own synthesis is not entirely without precedent.

One objection to the proposal that PrP is encoded by a host gene centers on the observation that there seem to be various "strains" of prions. If replication of the disease agent is nothing more than activation of a host gene, how can the same genetic line of animals serve as host to multiple prions? Any answers suggested now are necessarily speculative, but one possibility involves the rearrangement of genes. In the synthesis of immunoglobulins, for example, the reshuffling of genes gives rise to an enormous diversity of proteins.

The question of whether or not there is a host-animal gene for PrP is likely to be settled in the coming months. From the partial amino acid sequence of PrP we have been able to prepare DNA with a sequence of nucleotides complementary to one that would encode the known part of the protein. The synthesized DNA should bind to any DNA with the complementary sequence, and so it can be employed as a probe to find a

AMINO-TERMINAL AMINO ACIDS

SEQUENCING CYCLE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
MAJOR SIGNAL	Gly	Gln	Gly	Gly	Gly	Thr	His	Asn	Gln	Trp	Asn	Lys	Pro	Ser	Lys
MINOR SIGNALS				Thr	His	Asn		Trp		Lys	Pro				
			Pro	Trp		Gln				Thr	His		Gln	Trp	

INTERPRETED AMINO ACID SEQUENCE

Gly	Gln	Gly	Gly	Gly	Thr	His	Asn	Gln	Trp	Asn	Lys	Pro	Ser	Lys
		*	*	*	Thr	His	Asn	*	Trp	*	Lys	Pro		
*	*	Pro	Trp	*	Gln	*	*	*	Thr	His	*	Gln	Trp		

AMINO ACID SEQUENCE of PrP has come under investigation with the availability of substantial quantities of highly purified material. The sequence of the first 15 amino acids at the end of the molecule known as the amino terminus has been determined. Variant sequences can be brought into alignment with the sequence deduced from the major signal and interpreted as "ragged ends" caused by protease treatment in the course of the material's preparation.

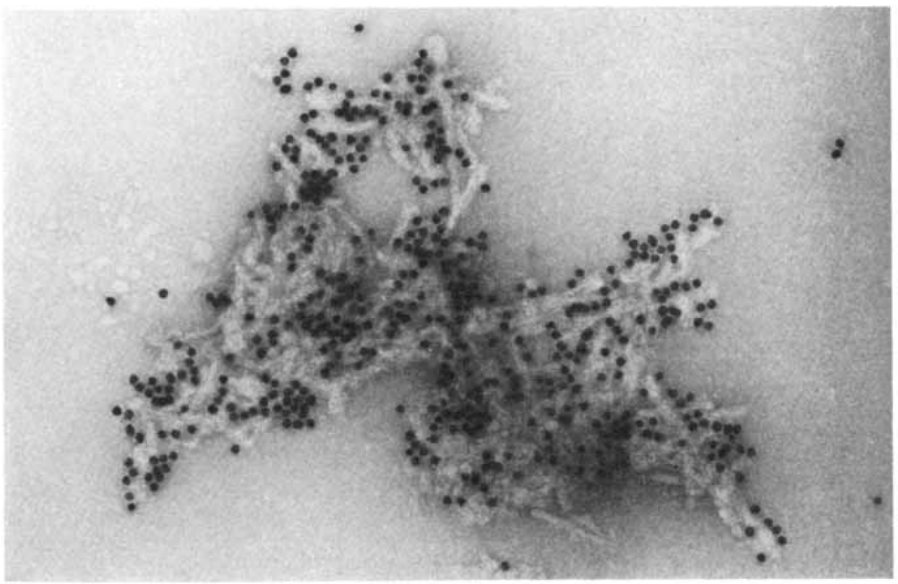
PrP gene in the cell. If prion proteins are indeed encoded by host genes, it may be more appropriate to speak of the synthesis of new prions as amplification rather than replication.

Because electron microscopy has revealed much about the structure and assembly of virus particles, many investigators have employed the electron microscope to search for a specific particle associated with scrapie infection. Spherical and cylindrical particles have been described in tissue sections and extracts. H. K. Narang found rod-shaped particles in sections of scrapie-infected brain tissue and showed that the particles are stained by substances that selectively bind to sugars. The latter findings are of notable interest, since PrP aggregates to form rods and is now known to be a glycoprotein. Henryk M. Wisniewski and his colleagues at the Downstate Medical Center of the State University of New York have found long fibrils in brain tissue infected with scrapie and Creutzfeldt-Jakob disease. They believe that the fibrils can be distinguished from amyloid, that they represent a filamentous animal virus causing scrapie and that they are an elongated form of prion rods.

The prion rods have recently given us further clues to the biological and medical significance of prions. The rods can be found in preparations consisting of one protein, namely PrP, and so they must be composed largely of PrP molecules. Our recent work with antibodies to PrP confirms that identification: the antibodies specifically bind to the rods. In electron micrographs the rods are typically from 10 to 20 nanometers in diameter and from 100 to 200 nanometers long, which suggests a single rod may consist of as many as 1,000 PrP molecules, probably stacked in a crystalline array.

Perhaps the most important and intriguing aspect of the prion rods is their resemblance to amyloid. The standard procedure for identifying amyloid is to stain a section of tissue with the dye Congo red. The amyloid binds the dye and appears red in light micrographs; in addition, when the stained amyloid is viewed through polarizing filters, it exhibits the optical property called birefringence, changing from green to gold as the orientation of the filters is changed. With George G. Glenner of the National Institute of Arthritis, Metabolism, and Digestive Diseases we stained clusters of prion rods with Congo red and examined them microscopically. They were red under ordinary illumination and showed green-gold birefringence between polarizing filters.

For more than 60 years amyloid plaques in the central nervous system were considered accumulations of waste material formed as a result of some disease process. Our findings suggest a quite different interpretation, namely that the plaques may be aggregations of prions in an almost crystalline state. They could be analogous to the inclusion bodies characteristic of many viral infections; inclusion bodies are crystalline arrays of virus particles.



CLUSTER OF PRION RODS is seen at a magnification of 100,000 times in an electron micrograph. Adhering to the rods are antibody molecules with a specific affinity for PrP. The antibodies have been labeled with colloidal gold beads, which appear in the image as conspicuous black dots. The structure of such a cluster of rods is much like that of an amyloid plaque.

The recent production of antibodies to PrP has allowed us to demonstrate that amyloid plaques in the brain of scrapie-infected hamsters contain prion proteins. Brain sections from the hamsters were stained first with the antibodies and then with Congo red. The same structures that bind the PrP antibodies also stain with the dye and exhibit green-gold birefringence. In collaboration with David T. Kingsbury of the University of California at Berkeley and the U.S. Naval Biosciences Laboratory we have also employed the antibodies to scrapie PrP to identify similar prion proteins in Creutzfeldt-Jakob disease. Proteins with the properties of PrP were found in the brain of both animals and human patients infected with the agent of Creutzfeldt-Jakob disease. These proteins too aggregate to form rod-shaped particles that have the characteristics of amyloid.

The disorder in which amyloid is an almost definitive sign is Alzheimer's disease; indeed, as the number of amyloid plaques increases, the degree of mental dysfunction rises. The possibility that Alzheimer's disease might be caused by prion infection is intriguing, but there is no reliable evidence that the disorder is transmissible or communicable. It is not spread by person-to-person contact. Gajdusek and Gibbs have repeatedly attempted to transmit the disease to experimental animals by inoculation; their results have been negative except in two possible cases, which they have not been able to reproduce.

If Alzheimer's disease is in fact caused by prions, two hypotheses might be proposed to explain the apparent failure of the transmissibility experiments. First, the infectious agent may not replicate in the species chosen for the experiments. Second, the incubation period may simply be too long for the disease to have been detected in the experiments; indeed, it may be longer than the lifetime of the animals. The second hypothesis is consistent with reports of an incubation period of two or three decades for kuru and Creutzfeldt-Jakob disease; it is also notable that Alzheimer's disease is commonest in older people and that the incidence rises with age. Of course, the possibility also remains that Alzheimer's disease is not caused by prions or any other infectious agent; many other causative mechanisms have been proposed.

The transmissibility of Creutzfeldt-Jakob disease raises questions of a different kind. That the disease can be transmitted in laboratory experiments is not in doubt, but it is not clear how the infection persists in a natural population. Creutzfeldt-Jakob disease has an incidence of about one case per million population throughout the world, which would not seem to be enough to sustain a chain of transmission from person to person.

Perhaps these puzzles will be resolved when the biochemistry of the prion is understood in more detail. If the prion is indeed a single protein and the product of a gene native to the host organism, the time may have come for a reconsideration of what is meant by the concept of infection.

Seismic Tomography

By analyzing many earthquake waves with this technique, it is now possible to map the earth's mantle in three dimensions. The maps throw light on the convective flow that propels the crustal plates

by Don L. Anderson and Adam M. Dziewonski

The outer layer of the earth, the lithosphere, consists of a dozen rigid, crust-bearing plates that ride on the underlying mantle, rearranging continents, forming mountains, creating and destroying oceans. What drives this constant remodeling? Ultimately the motive force is the convective circulation of the mantle. The mantle is solid rock but so hot that over geologic time it deforms easily and flows. New lithosphere is formed at midocean ridges, where hot magma from the mantle wells up between diverging plates. The new surface material spreads outward from the ridges and eventually plunges back into the mantle at oceanic trenches, where two plates collide. Although this model is widely accepted, the origins of the upwelling material and the fate of the subducted plates—in general, the details of the flow in the mantle—have remained unknown, beyond the reach of conventional geophysics and its analytic methods.

In the past few years a new analytic technique called seismic tomography has become available that promises to profoundly enhance knowledge of the earth's internal structure, including the pattern of flow in the mantle. Like its medical analogue, computer-aided tomography or CAT scanning, seismic tomography combines information from a large number of crisscrossing waves to construct three-dimensional images of the medium the rays have traversed. In the case of CAT scanning the medium is the human body and the source of energy is an X-ray generator. The best source of information on the earth's interior is the seismic waves triggered by earthquakes, because they are attenuated only slightly by their passage through the earth; an earthquake of moderate strength radiates waves that are recorded by seismometers all over the world.

Seismic investigations over the past 70 years have revealed much of the earth's average radial structure: the fact that it has a crust, an upper and a lower mantle and an outer and an inner core. Tomography is now adding considerable detail to this simple model. By determining how fundamental properties such as temperature and density vary with latitude and longitude as well as with depth, it is providing the first three-dimensional view of the mantle.

To understand how seismic tomography works one must first understand a few facts about earthquake waves and how they behave. The earth transmits seismic disturbances because it is an elastic medium: it resists deformation, and when a part of it is strained—compressed in volume or sheared in shape—a restoring force acts to return that part to its original condition. A seismic wave is just a traveling strain triggered by the release of stress by an earthquake. Seismic waves travel fastest through regions of the earth that are most resistant to deformation.

Every earthquake radiates waves that plunge through the body of the earth as well as ones that travel along the surface. The body waves fall into two categories: compressional waves and shear waves. Compressional waves are like sound waves in that they consist of periodic compression and dilation of the rock along the direction of travel. Because earthquakes are precipitated by slippage or shearing across a fault, however, they also radiate shear waves, which are analogous to electromagnetic radiation in that the direction of vibration is transverse, or perpendicular, to the direction of propagation. Like electromagnetic waves, shear waves can be polarized, in which case they vibrate in a single transverse direction.

Surface waves also are of two principal types. Rayleigh waves, the first type, have both a compressional and a shear component; they cause rock particles to move elliptically in the vertical plane that lies between the earthquake focus and the detector. In contrast, Love waves are polarized shear waves vibrating in the horizontal plane parallel to the earth's surface. Although both Rayleigh and Love waves travel along great-circle paths at the surface, they extend deep into the mantle and thus provide information on its structure.

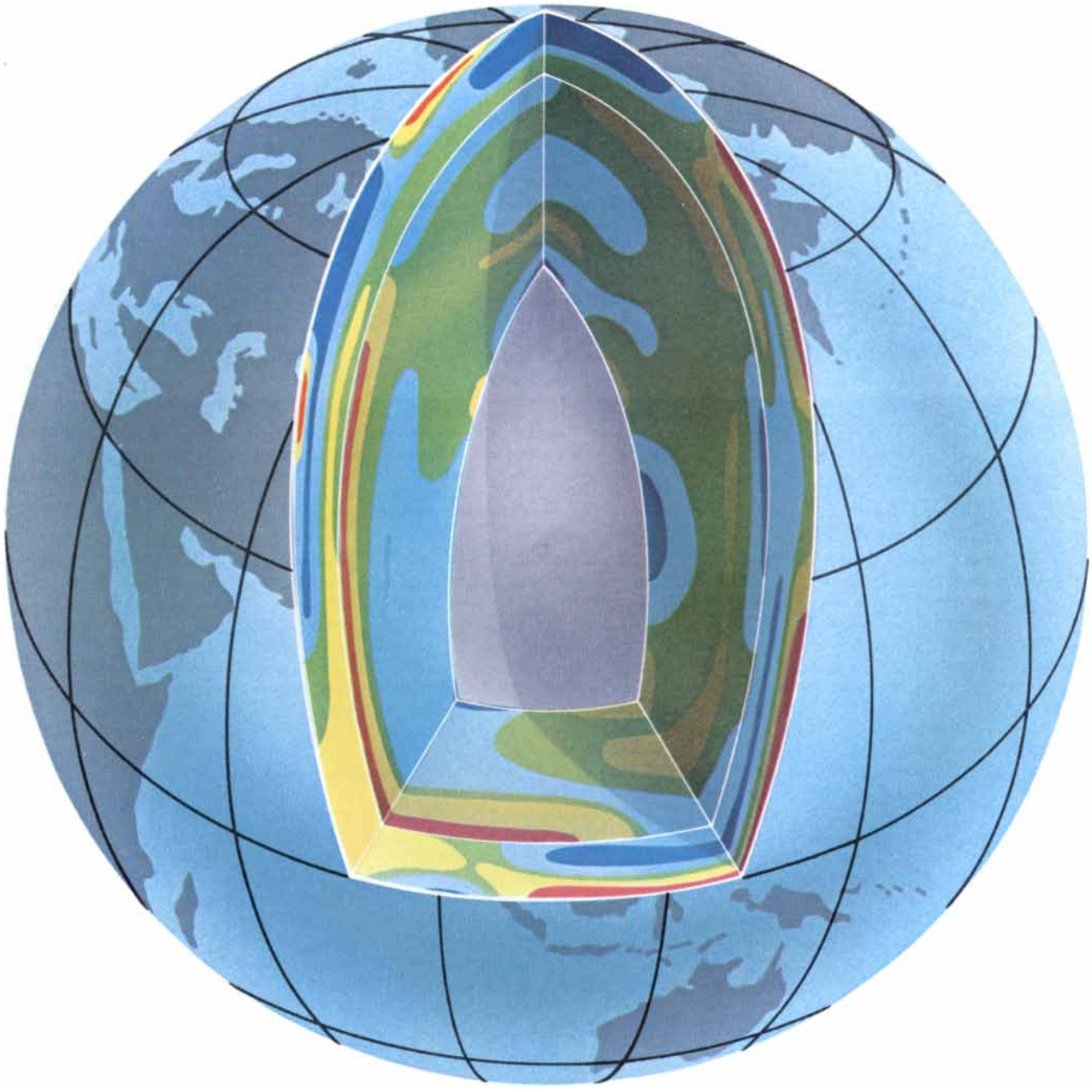
The velocity of shear waves is a function of the medium's rigidity, which is a measure of its resistance to shearing stress. Liquids, for instance, are not rigid and hence do not transmit shear waves; that is how it was discovered that the earth's outer core is liquid. In the case of compressional body waves and the compressional component of Rayleigh waves, velocity is a function of both rigidity and another elastic property of the medium: its incompressibility. (Similarly, sound, a compressional disturbance, travels faster through water than through air, which is more compressible, and faster still through ice, which is both incompressible and rigid.)

Cold material in general tends to be more rigid and more incompressible than hot material, and seismic waves therefore traverse cold regions of the earth's interior more rapidly. Hot material in turn has relatively low density and is generally associated with ascending flow in the mantle, whereas cold matter sinks because it is denser than its surroundings. Wave velocity also depends on a small-scale property of the medium: the orientation of crystals in the material. Mineral crystals of the kind that make up the mantle have three axes, and each axis exhibits a different degree of stiffness. The most rigid axis is "fast." If over a large region the fast crystallographic axes become aligned, say, by a current in the mantle, waves whose polarization (direction of vibration) or direction of propagation is parallel to the fast axes will accelerate in that region.

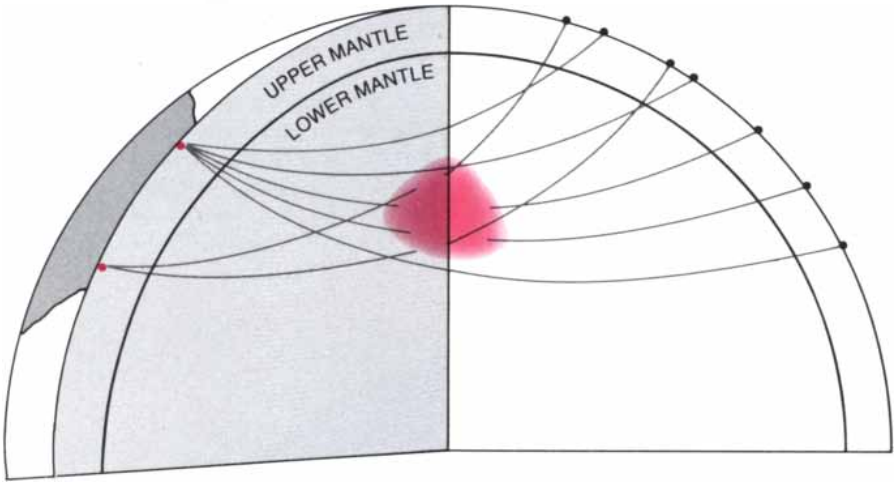
The velocity of seismic waves thus contains indirect information on the pattern of flow in the mantle. Extracting that information, however, is not easy. The velocity of a single ray, as computed from its arrival time at a seismic station, is simply an average over the ray's entire path and does not reveal where the wave has been slowed or speeded up. Moreover, the average is generally over a great distance, because large expanses

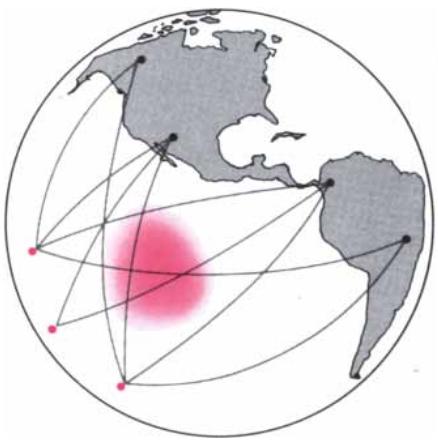
of the earth, particularly the oceans, are without seismic stations, and because earthquakes tend to occur only at plate boundaries. To draw inferences about such properties as density and temperature in the earth's interior it is necessary to combine information from many rays; the more the better.

Because of a vast expansion at the seismic data base, it is now possible to construct much more detailed images of the earth's seismic-velocity structure through the application of tomography. Recent improvements in the data base have been twofold. For the past seven years a global network of sensitive digital seismometers has been recording long-period surface waves in computer-readable form on magnetic tape. At the same time the International Seismological Centre (ISC) near London has been collecting reports from more than 1,000 ordinary seismic stations around the world. Most of these stations are relatively insensitive to surface waves, but they record body waves, which have a much shorter period, from earthquakes



EARTH MODEL derived from seismic tomography is far more detailed than one based on conventional seismology. Earlier studies determined seismic velocity only as global averages at various depths; the boundaries between the upper and the lower mantle at a depth of 670 kilometers and between the mantle and the core at 2,900 kilometers (*white lines*) mark zones where seismic velocity changes rapidly with depth. In contrast, tomography reveals lateral variations in velocity and, by inference, in the temperature and density of the mantle. The illustration shows vertical cross sections through the entire mantle along three great circles: along the Equator and from Hudson Bay through central Asia and through the western Pacific. Red and yellow indicate hot, upwelling regions, where seismic velocity is anomalously slow; blue regions are cold, dense and fast, and green indicates average velocity for a particular depth. Island arcs such as the Philippines and Indonesia show up as hot regions underlain by cold material in the upper mantle; these volcanic regions are subduction zones, where a cold, descending plate squeezes hot mantle upward. Large fast anomalies also appear near the core. Cross sections are based on studies done at Harvard University and were prepared with the help of Robert W. Clayton of the California Institute of Technology.





SEISMIC TOMOGRAPHY isolates velocity anomalies (*color*) in the mantle by combining information from many seismic waves traveling from earthquakes (*colored dots*) to seismic stations (*black dots*) along crisscrossing paths. Waves that miss the anomaly show the normal travel times for their surface distances. Waves that penetrate the anomaly are slowed down or speeded up. A dense mesh of intersecting rays can be used to define the anomaly and to measure the velocity with which it transmits waves; the structure of the anomaly must account for the observed deviations in travel time of all the rays that traverse it. The lower mantle is mapped with body waves (*top*) that dive into the earth. Long-period surface waves (*bottom*) that "see" to great depth offer the best coverage of the upper mantle.

all over the world. As a by-product of its effort to pinpoint the location of some 10,000 earthquakes per year, the ISC has collected arrival times for several million seismic rays.

Although medical tomographers have an advantage over seismologists in that they can control the source of the radiation as well as the detection apparatus, the analytic technique is fundamentally the same. In CAT scanning, X rays are used to map density variations in the human body and thereby reveal internal organs and other structures. The absorption of X rays is greatest in dense regions of the body, and so these regions show up on the image as shadows. On a conventional radiographic image it is often difficult to distinguish overlapping structures, particularly when they have similar densities; CAT scanning overcomes this problem by mathematically recombining information taken from many X rays sent through the body along different paths. The result is horizontal slices that when stacked show internal structure in three dimensions.

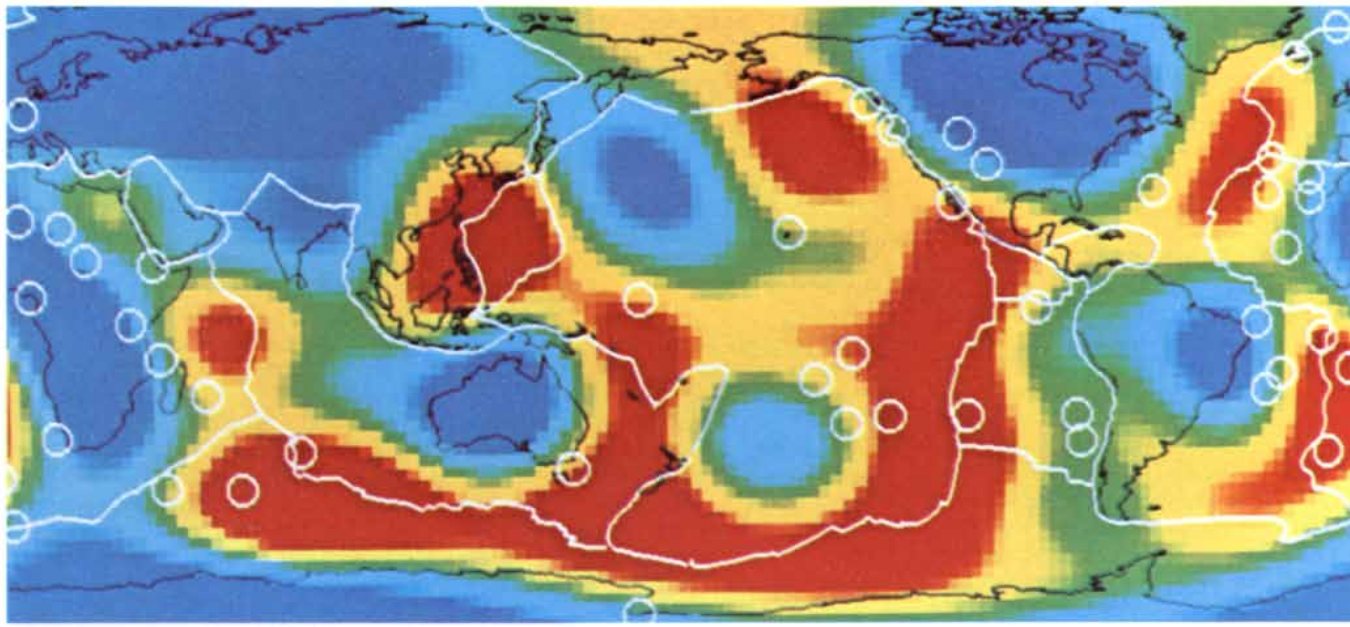
In tomographic studies of the earth's interior it is the velocity of seismic waves and not their absorption that is measured, and the resulting images are maps of "slow" and "fast" regions in the mantle. These anomalies are found, just as in CAT scanning, by combining the information from many crisscrossing rays. If the velocity of an individual seismic ray deviates from the expected value (taken from seismological tables that list average travel times for seismic waves according to the surface distance between the epicenter and the seismometer), the anomalous mantle mass that caused the deviation could lie anywhere along the ray's path. If, however, another ray crosses the first one at some point, then the measured velocity of the second ray provides a constraint on the velocity of the first ray at the point where they intersect. A dense mesh of many intersecting rays creates a network of mutual constraints that make it possible to map the velocity structure of the region covered by the mesh. The denser the mesh is, the greater are the resolution and the accuracy of the map.

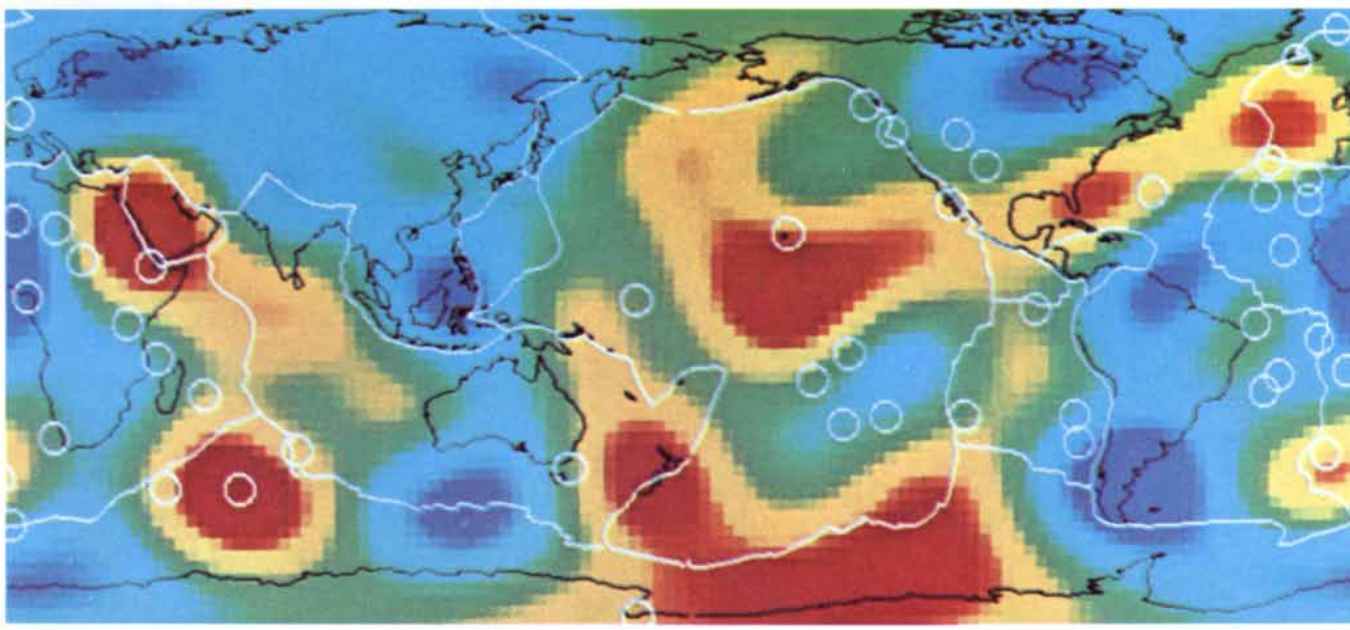
In practice finding velocity values that satisfy all the constraints requires a complex mathematical procedure and a large computer to carry it out. Essentially the technique consists in solving a set of simultaneous equations for each unit region of the mantle. On the right side of the equations are the known travel times of all the rays that traverse the region. On the left side of each equation is a series of terms with associated velocity parameters. Solving the problem means finding the values of the velocity parameters that produce the closest fit between the expansions of terms and the observed travel times. This is an example of the "inverse problem" in seismology: working backward from observations to a model of the earth's structure.

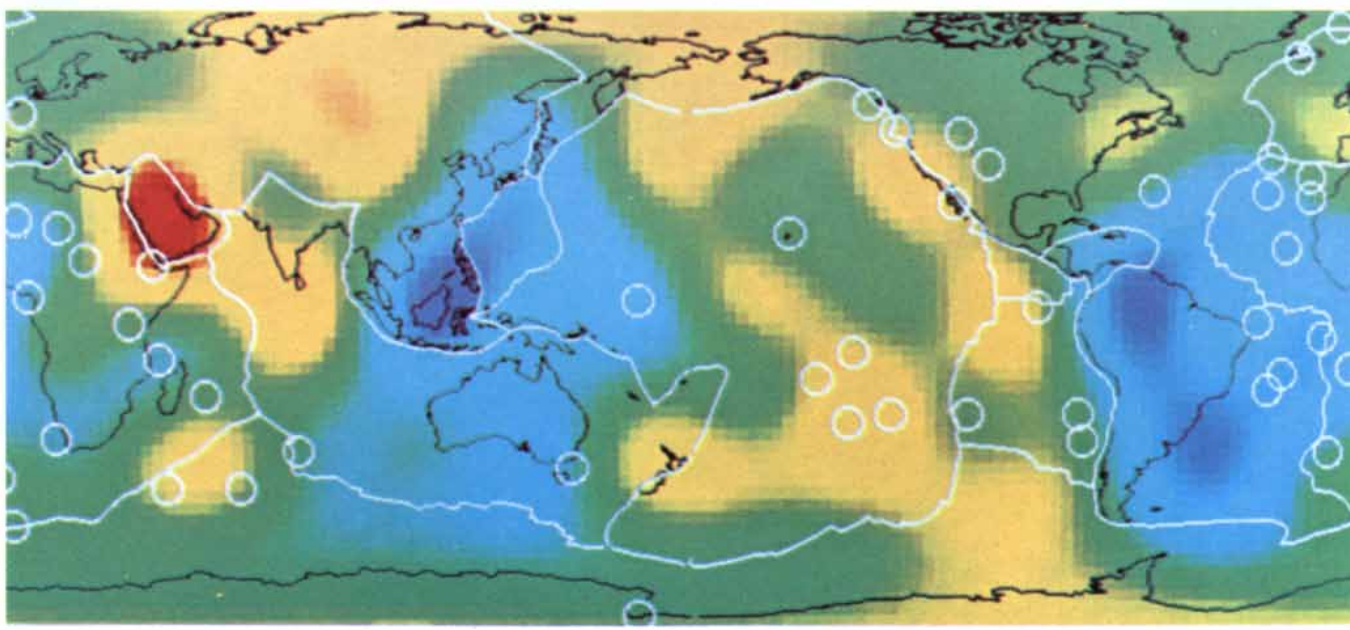
The body-wave arrival times collected by the ISC and the surface-wave data recorded by the network of long-period seismometers are complementary sources of information on the mantle. Body waves are the only direct way of probing the lower mantle, which extends from the bottom of the upper mantle at 670 kilometers to the boundary of the core, which lies at 2,900 kilometers. One of us (Dziewonski) has recently derived, from a data base of more than 500,000 rays, a model of the lower mantle that resolves structural features with horizontal dimensions of 2,000 to 3,000 kilometers and a vertical extent of 500 kilometers. Robert W. Clayton and his colleagues at the California Institute of Technology Seismological Laboratory have derived a more detailed model using a larger set of data and special techniques for enhancing resolution. The large-scale features identified in the two models are similar; both studies, for example, found evidence of large velocity anomalies near the core boundary.

A body wave that plunges into the earth from an earthquake focus and later reemerges at a seismometer follows a relatively short, nearly vertical path through the upper mantle, and with the current distribution of seismic stations it is impossible to achieve a sufficiently dense mesh of body-wave paths to image that layer. On the other hand, the entire surface of the earth is fairly well sampled by Rayleigh and Love waves that "see" deep into the upper mantle. Long-period, or low-frequency, surface

TOMOGRAPHIC MAPS show shear velocity of surface waves at three depths in the upper mantle. At 150 kilometers (*top*) surface tectonic features are still evident: midocean ridges in the Atlantic, the eastern Pacific off South America and the southern Indian Ocean have slow velocity (*red*); these are regions where hot mantle is welling up to the surface. Volcanic regions above subduction zones in the western and southwestern Pacific are also slow. At 350 kilometers (*middle*) there is less correlation between seismic velocity and surface features, but cold subducted Pacific lithosphere shows up as fast regions (*blue*) under the western Pacific and South America; these fast anomalies are even more pronounced at 550 kilometers (*bottom*). The darkest colors indicate seismic velocities that differ by 2 percent from the average (*green*) for that depth. White lines and circles represent plate boundaries and surface hot spots. Maps are based on a study by John H. Woodhouse and one of the authors (Dziewonski) at Harvard.







waves sample the mantle to greater depths than waves with relatively short periods, just as a long swell in the ocean reaches farther below the surface than short chop. The new long-period digital seismometers register surface waves whose velocity is affected by structure down to a depth of about 700 kilometers, slightly below the boundary between the upper and the lower mantle.

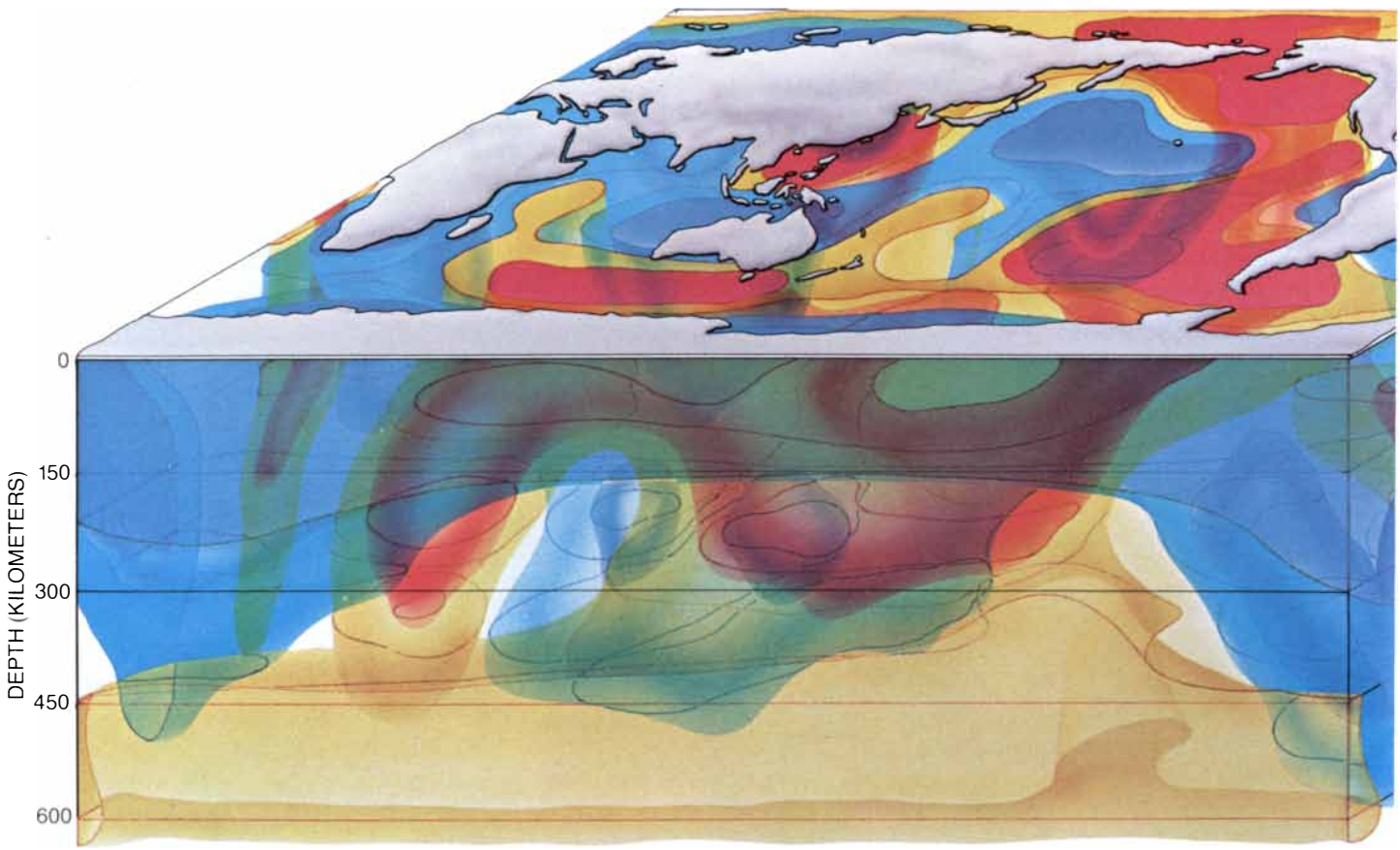
With our colleagues Ichiro Nakanishi, Henri-Claude Nataf and Toshiro Tanimoto at Caltech and John H. Woodhouse at Harvard University we have analyzed these surface-wave data and have constructed tomographic images showing the lateral and vertical extent of velocity anomalies in the upper mantle. So far we have calculated, for each of some 60 large earthquakes that occurred between 1977 and 1982 and that provide the best available geographic coverage, the velocity of surface waves of about six frequencies traveling from the event to each digital seismometer. The different frequencies give information on the velocity structure of the mantle within different, overlapping depth intervals. On the average we use data from about 20 seismic stations; because surface waves travel from earthquake to station along both the short arc and the long arc of a great circle, we obtain, for each quake and each frequency, measurements of average velocity along about 40 different paths. To prepare a map of the mantle at a particular depth or a vertical cross section along a particular great circle we typically combine between 400 and 1,000 different paths, converting the averages along many arcs into specific velocities in each region.

Plate-tectonic theory gives us some idea of what to expect in mapping the upper mantle. Under midocean ridges, volcanic regions and regions such as the Red Sea where rifts are forming in continental plates, seismic velocities should be slow; these are areas where hot, less dense mantle is welling up toward the surface, melting as it rises. Below stable continental "shields," where plates have remained at the surface for billions of years and so have had ample time to cool, one would expect seismic waves to travel anomalously fast, at least in the shallow mantle. At greater depths fast anomalies should be found in regions that have been cooled by the subduction of oceanic lithosphere, which has itself cooled while at the surface.

To a certain extent our results confirm these expectations. At a depth of 150 kilometers we have observed slow seismic velocities under most of the world's tectonic and volcanic regions, including the midocean ridges. In contrast, the Canadian, Brazilian, Siberian, African and Australian shields are all fast. The velocity differences, in fact, are greater than can be explained by temperature alone; there must also be some as yet undetermined lateral variations at this depth in mineral composition or in the extent of melting of the mantle.

The fact that ridges and volcanoes are underlain by hot mantle at shallow depths is not surprising. The more interesting question, much debated in geophysical circles, concerns the depth to which these hot thermal anomalies extend. Our maps suggest they reach at least to the 400-kilometer level, but the surface expressions are often offset by large distances from the mantle source.

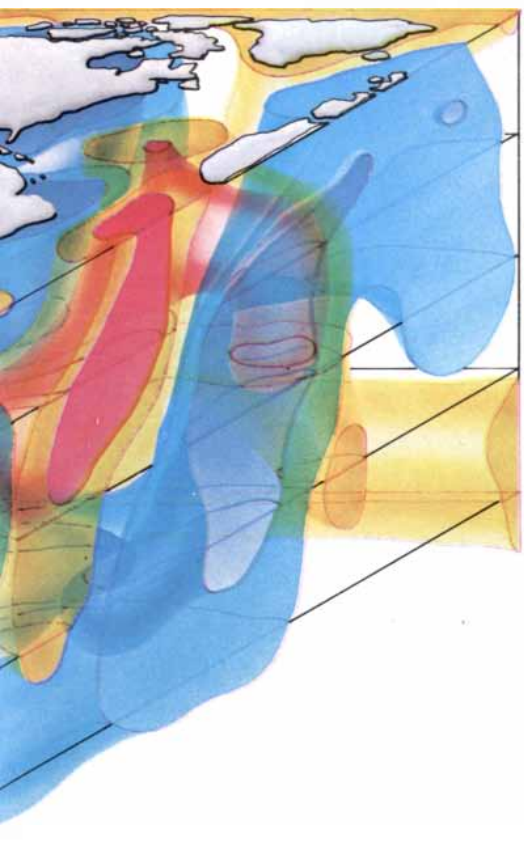
At a depth of 350 kilometers, for example, the globe-girdling midocean ridge system is no longer continuous but is broken up into isolated segments. The central Mid-Atlantic and the Southeast Indian Ocean ridges are actually underlain by fast material, and the slow anomaly associated with the East Pacific Rise has either vanished or is laterally offset.



THREE-DIMENSIONAL MODEL of the upper mantle was assembled from Harvard seismic-velocity maps at depth intervals of 50 kilometers. The model suggests that lateral movement of both hot and relatively cold material in the mantle is extensive. The hot anomalies (*red and yellow*) associated with midocean ridges can be traced to great depth, but the source is often offset from the surface expression; for example, the upwellings responsible for the East Pacific Rise and the Southeast Indian Ocean Ridge seem to come from a single source under the South Pacific. The same color at two different depths implies not that the seismic velocities of the two regions are

At 550 kilometers there is even less relation between mantle and surface features: most of the Atlantic and Indian Ocean ridges are fast, whereas the Siberian shield, cold and therefore fast near the surface, is slow. Clearly the mid-ocean ridge system is not simply the surface expression of vertical upwelling currents. Instead it seems to be fed by lateral transport of hot material from a few broad thermal anomalies in the upper mantle.

The tomographic maps also confirm some general predictions, based on plate tectonics, about the location of subducted lithosphere. Since the breakup of the supercontinent Pangaea some 200 million years ago and the subsequent formation of the Atlantic Ocean, the Pacific has been shrinking. North and South America have been encroaching on it from the east, and in the west Pacific lithosphere is being overridden by eastern Asia, Australia and various large islands. In mapping the velocity structure of the mantle at a depth of 350 kilometers we found fast regions under eastern Asia, South America and the central and southern Atlantic—right where one might expect to find cold, subducted Pacific lithosphere. At 550 kilometers the fast anomalies are larger and even more prominent; the one in the west includes most of the western Pacific and all of Australia.



equal, but rather that the velocities differ by the same relative amount from the averages at their respective depths. Green marks regions where a cold anomaly (*blue*) lies in front of or behind a warm, slow anomaly (*yellow*).

Interestingly, seismic velocities in the western Pacific at shallower depths—200 kilometers and less—are slow. This too is expected from plate tectonics. When cold oceanic lithosphere enters the mantle at a subduction zone, it displaces hot mantle upward. At the same time volatile compounds (mainly water) in the subducted sediment and crust lower the melting point of the mantle, and friction melts part of the mantle "wedge" above the descending slab. Hot, buoyant magma rises to the surface and forms volcanoes. The island arcs of the western Pacific, including Japan and the Philippines, are volcanic archipelagos situated above subduction zones.

Our results show that mapping seismic-velocity anomalies is an accurate way of locating hot and cold regions and hence of locating ascending and descending currents in the mantle. By themselves, however, the maps offer little information on the final leg of the convective circulation: the horizontal flow of subducted material from trench to ridge. Fortunately clues to the pattern of this horizontal flow can be extracted from seismic data by taking into account the fact that the velocity of seismic waves in the mantle also depends on the direction in which the waves travel through a given region. As a result of differences in crystal orientation waves propagate faster in a particular azimuth, or horizontal direction, than along other azimuths. This property of the mantle is called azimuthal anisotropy.

Much of the shallow mantle (the layer above a depth of 400 kilometers) is composed of olivine, a magnesium-iron silicate. Olivine crystals are anisotropic with respect to seismic waves: one axis of the crystals is significantly faster than the other two axes. If a mass of crystals were oriented randomly, the effects of their anisotropy would cancel, but field studies have shown that olivine crystals tend to be aligned by flow in the mantle. Over very large areas their fast axes are oriented parallel to the flow direction; this can be thought of as analogous to the magnetization of a piece of iron by an external magnetic field.

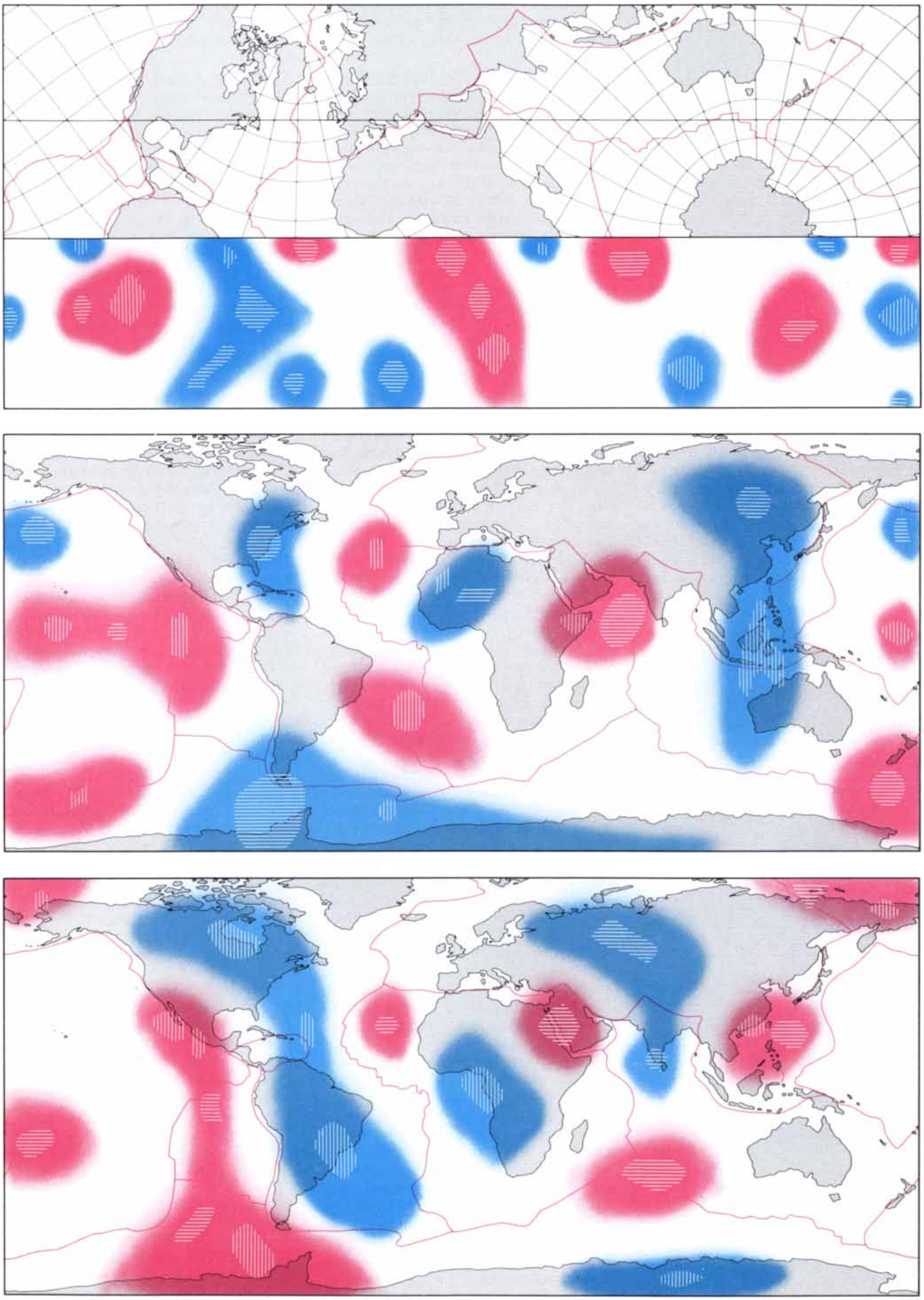
By mapping the fast directions of seismic waves it is thus possible to get an idea of the horizontal flow in the mantle. Tanimoto and one of us (Anderson) have prepared a map showing the fast directions of Rayleigh waves, which are most sensitive to physical properties in the 200-to-400-kilometer depth interval [*see illustration on page 68*]. Because Rayleigh waves travel fast in each direction along the fast crystallographic axis, the flow directions on the map have a 180-degree ambiguity, but by assuming that the flow is generally from subduction zone to ridge the ambiguity can usually be resolved. When they are combined with the seismic velocity maps, the maps of azimuthal anisotropy provide powerful new constraints on theoretical models of convection in the mantle. They show, for example, that the upper mantle under central North America is flowing from north to south, whereas under Siberia the flow is probably in the opposite direction.

By exploiting the anisotropy of the mantle in another way, we can distinguish on a single map between regions of horizontal flow and regions of vertical flow. Rayleigh waves vibrate in the vertical plane, whereas Love waves vibrate horizontally. In a region in which the crystals are aligned the velocities of a Rayleigh wave and of a Love wave traveling along the same path will thus be different. In general the horizontally polarized shear-wave velocity of Love waves should be greater in regions of horizontal flow, because the fast axes of the olivine will lie in the flow plane and in the direction of vibration of at least some of the Love waves passing through the region. Conversely, the Rayleigh waves' vertically polarized shear velocity should be greater in ascending or descending currents.

Maps combining the information on the "polarization anisotropy" of the mantle with seismic-velocity data pick up clearly the hot upwelling under mid-ocean ridges and continental rifts, as well as the downward motion at subduction zones in the western Pacific. Equally clear is the evidence they give of large-scale lateral transport of both cold material and hot material.

Until the advent of seismic tomography geophysicists had no direct way of mapping convection in the mantle. From data collected by seismometer networks in earthquake-prone regions they were able to determine seismic velocity, temperature and density as global averages at various depths, and in a few regions it was even possible to discern broad lateral variations in these properties. But the data and the analytic technique needed to construct global maps of lateral variations were lacking.

In the absence of direct observational evidence workers inferred what they could about mantle flow from indirect evidence gathered at the earth's surface: mainly variations in gravity and in surface elevation (or ocean depth). In general hot upwelling regions are expected to exhibit anomalously high gravity and high elevation. The problem, however, is that gravity and elevation are integrated functions of the way density varies with depth. In other words, a given gravity or elevation anomaly could be produced by an anomalous mass at any depth. As a result variations in these quantities cannot serve to locate unam-

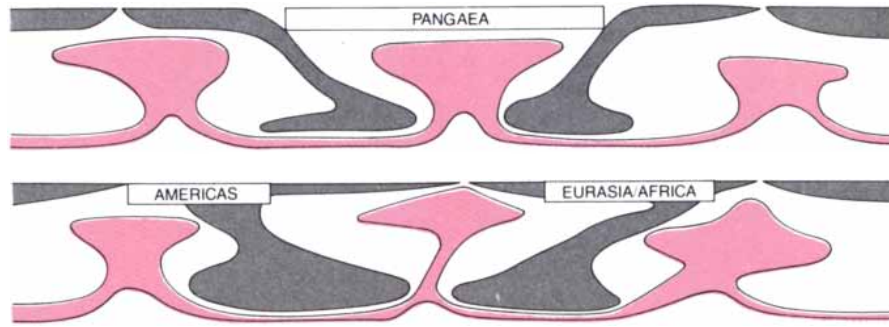


biguously the density anomalies that drive convection in the mantle, and so they are a poor guide to the nature of the convective flow.

Geophysicists have also placed heavy emphasis on numerical models of simplified convective systems. One significant simplification in most models has been the assumption that the convective flow is two-dimensional, with material moving in a vertical plane from ridge to trench at the surface and from trench back to ridge in the mantle. Most numerical calculations have also assumed that viscosity, or resistance to flow, is uniform throughout the mantle. Such assumptions were adopted in part because more complicated flow models would require calculations that strain the power of even the largest computers.

Not surprisingly, reality as it is now being revealed by direct seismic mapping of the mantle turns out to be a good deal more complex than the simple models. For one thing, it is three-dimensional: both cold and hot material flow in many lateral directions, and the spreading at the surface and the return flow in the mantle are by no means in the same vertical plane. Thermal anomalies under midocean ridges, under continental rifts and even under volcanic regions can be traced to great depth, but they are offset from the surface features and are not simply vertical sheets of rising magma. The thermal anomalies and the density variations they induce control the motions of the plates to some extent, but plate tectonics in turn affects the location of the anomalies: the subduction of cold lithosphere cools the mantle, and thick continental lithosphere at the surface insulates the mantle below and causes it to heat up.

Furthermore, in most areas of the world there is, contrary to the assumption of many numerical models, a low-viscosity layer in the upper mantle under the lithosphere. This weak layer partially decouples the plate from the mantle. As a result the moving plates drag mantle along with them and distort the underlying circulation. Some years ago Bradford H. Hager, now at Caltech, and Richard J. O'Connell of Harvard did flow calculations taking the weak layer into account; their results, unlike those of simpler models, are generally consistent with our map of horizontal flow based on azimuthal anisotropy.



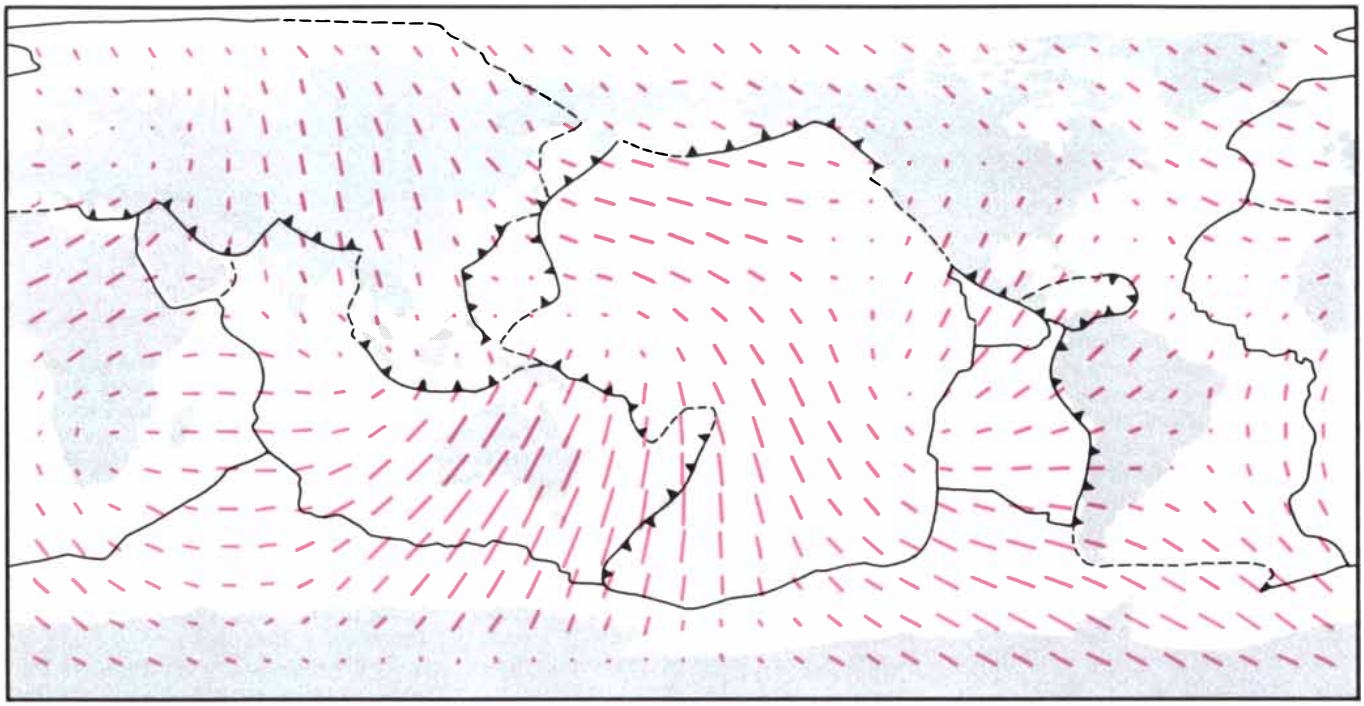
BREAKUP OF PANGAEA some 200 million years ago opened the Atlantic and has led to the steady shrinking of the Pacific. The process is illustrated schematically with vertical cross sections of the upper mantle before (*top*) and after (*bottom*) the breakup. The hot anomalies (*red*) are stable features; the one under Pangaea became the source of the Mid-Atlantic Ridge. As the Americas, Australia and Asia override Pacific lithosphere (*gray*), however, the subduction of surface material cools the mantle and distorts the hot upward flow. The subducted lithosphere creates fast anomalies that make the continents appear to have deep, cold "roots."

Seismic tomography allows a reversal of the previous sequence of analysis: instead of trying to infer the existence of density anomalies in the mantle from the earth's gravity field, investigators can use seismic maps of the density distribution to explain observed variations in gravity. Drawing on the results of two independent tomographic studies, one by Clayton and Robert Comer of Caltech and the other by one of us (Dziewonski), Hager and his colleagues have shown that the large-scale variations in the gravity field—broad gravity highs over the central Pacific and Africa and lows over the Indian Ocean and Antarctica—can be traced to large density anomalies in the lower mantle. Smaller-scale fluctuations in gravity, on the other hand, seem to be at least in part the result of density variations in the upper mantle. Our maps show that the gravity highs in the North Atlantic between Iceland and the Azores, in the South Atlantic centered on the island of Tristan de Cunha and in the southwestern Indian Ocean between Madagascar and Kerguelen are all underlain by slow regions in the mantle, corresponding to hot, upwelling plumes, at depths of 200 to 400 kilometers; a hot plume in the upper mantle also appears to contribute to the gravity anomaly in the central Pacific that Hager has linked to the lower mantle. Although the hot material has relatively low density, these plumes are nonetheless expected to produce gravity highs because they increase surface elevation and thus raise the center of gravity under a particular surface area.

The flow model Hager used to make his calculations assumes that subducted lithosphere sometimes penetrates the seismic boundary between the upper and the lower mantle. The question of whether material can flow across the 670-kilometer boundary, or whether instead the convective circulation in the mantle is divided into two separate layers, is a central issue in geophysics. Tomography has not yet settled the debate. Our maps show apparent continuities across the boundary in some regions, such as the anomalously hot central Pacific, but discontinuities in others. Moreover, the continuities could be explained without assuming that material is transferred across the boundary: the lower mantle may simply heat regions of the upper mantle as a stove heats a pot, by conduction. Unfortunately the boundary is in a depth interval in which the resolution of both surface-wave and body-wave tomography is poor. The solution is to do a tomographic analysis combining both types of data, and in particular to analyze body waves whose paths are confined to the upper mantle because they are reflected or refracted off the 670-kilometer boundary and other discontinuities.

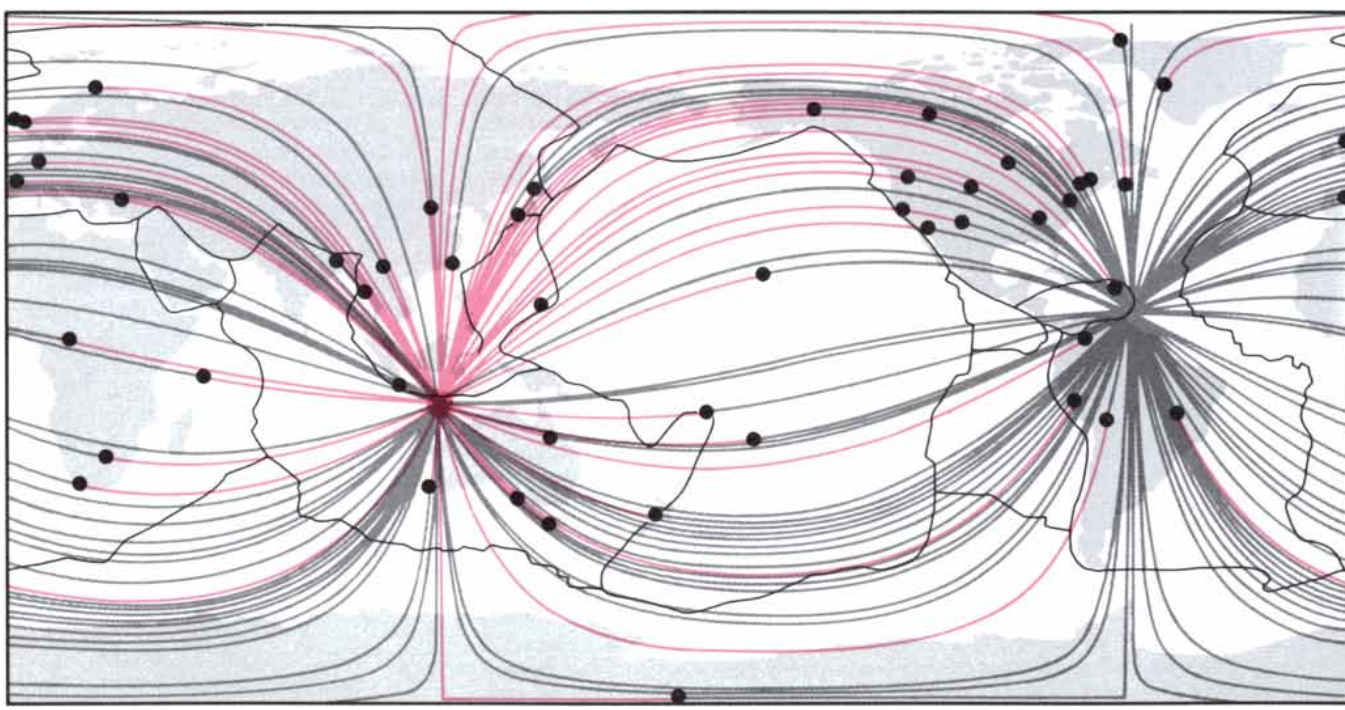
Although the results we have obtained are a striking improvement over conventional seismology, the sparseness of the present network of digital seismic stations limits the resolving power of surface-wave tomography, even in the shallow mantle. Only velocity anomalies that are either very large, about

FLOW PATTERN of the mantle can be mapped by virtue of the fact that it is anisotropic: the speed of seismic waves depends on their direction of travel and their polarization. Where the flow is horizontal the horizontally polarized Love waves are faster than the vertically vibrating Rayleigh waves. The converse is true in regions of vertical flow. On the maps at the left anisotropy information is superposed on ordinary velocity data showing the location of hot and cold anomalies; vertical hatching indicates vertical flow and horizontal hatching indicates flow in any horizontal direction. The vertical cross section (*top*), a slice into the earth along a great circle (*horizontal line*), suggests among other things that the thermal anomaly in the Red Sea area extends at least to the bottom of the cross section at 350 kilometers. At a depth of 250 kilometers (*middle*) parts of the midocean ridge system are characterized by hot, rising currents and the subduction zones of the western Pacific are marked by cold, descending material. Other ridge segments, however, such as the central Mid-Atlantic Ridge and the Southwest Indian Ocean Ridge, are not hot. The map at 100 kilometers (*bottom*) suggests that these regions are fed by lateral transport of hot material at that depth. The maps are based on studies conducted by Henri-Claude Nataf, Ichiro Nakanishi and one of the authors (Anderson) at Caltech.



HORIZONTAL FLOW in the upper mantle can be mapped by determining the fast propagation direction of Rayleigh waves. Such waves travel fastest either way along the fast crystallographic axes, which tend to be aligned by flow; the 180-degree ambiguity in the flow lines is partially resolved by assuming that the flow is generally from subduction zones (*toothed black lines*) back to ridges (*solid black lines*). The map was prepared at Caltech using 200-second waves, which are sensitive to structure at depths of 200 to 400 kilometers.

2,000 kilometers in radius, or very pronounced can be mapped. Recently seismologists from nearly 50 universities formed the Incorporated Research Institutions for Seismology (IRIS), a nonprofit corporation with the objective of modernizing seismological research facilities. Among other projects IRIS plans to improve both the quality and the quantity of the digital data by increasinq the number of long-period seismic stations to at least 100 and by operating about 1,000 portable seismometers. A detailed study of a subduction zone using the portable seismometers in conjunction with the global array of permanent digital seismometers might determine, for example, whether descending plates do indeed plunge into the lower mantle. Together the new permanent and portable seismic stations should produce noticeably sharper global images of the mantle and of the hidden flow that shapes the surface of the earth.



DIGITAL SEISMOMETERS (*black dots*) record long-period surface waves that travel from an earthquake (*colored dot*) along both the minor arc (*color*) and the major arc (*gray*) of a great circle. Installing new stations, particularly on the ocean floor and in land areas such as Siberia that are not near earthquake zones, would tighten the mesh of intersecting rays and improve the resolution of tomography.

The birth of a new generation.

A new coal gasification complex opened April of this year in Kingsport, Tennessee, making Kodak's Eastman Chemicals Division the first manufacturer in the United States to use modern coal conversion technology to produce industrial chemicals commercially from coal.

This method of manufacturing a new generation of chemicals saves us the equivalent of one million barrels of oil annually. A fact that makes us pleased. And just a little proud.

Using nearby Appalachian coal mines as a ready source of supply, the new plant can gasify about 900 tons of coal a day. That amounts to annual production of around 500 million pounds of acetic anhydride—a chemical needed to produce cellulose acetate, used in the manufacture of photographic film base, fibers, plastics and other products. Conventional technology, which uses ethylene produced from petrochemical feedstocks as the raw material, would consume the equivalent of a million barrels of oil to achieve the same results.

While some of the syngas processes were developed by others and licensed to Eastman Chemicals Division, new technological breakthroughs developed by us made the venture work.

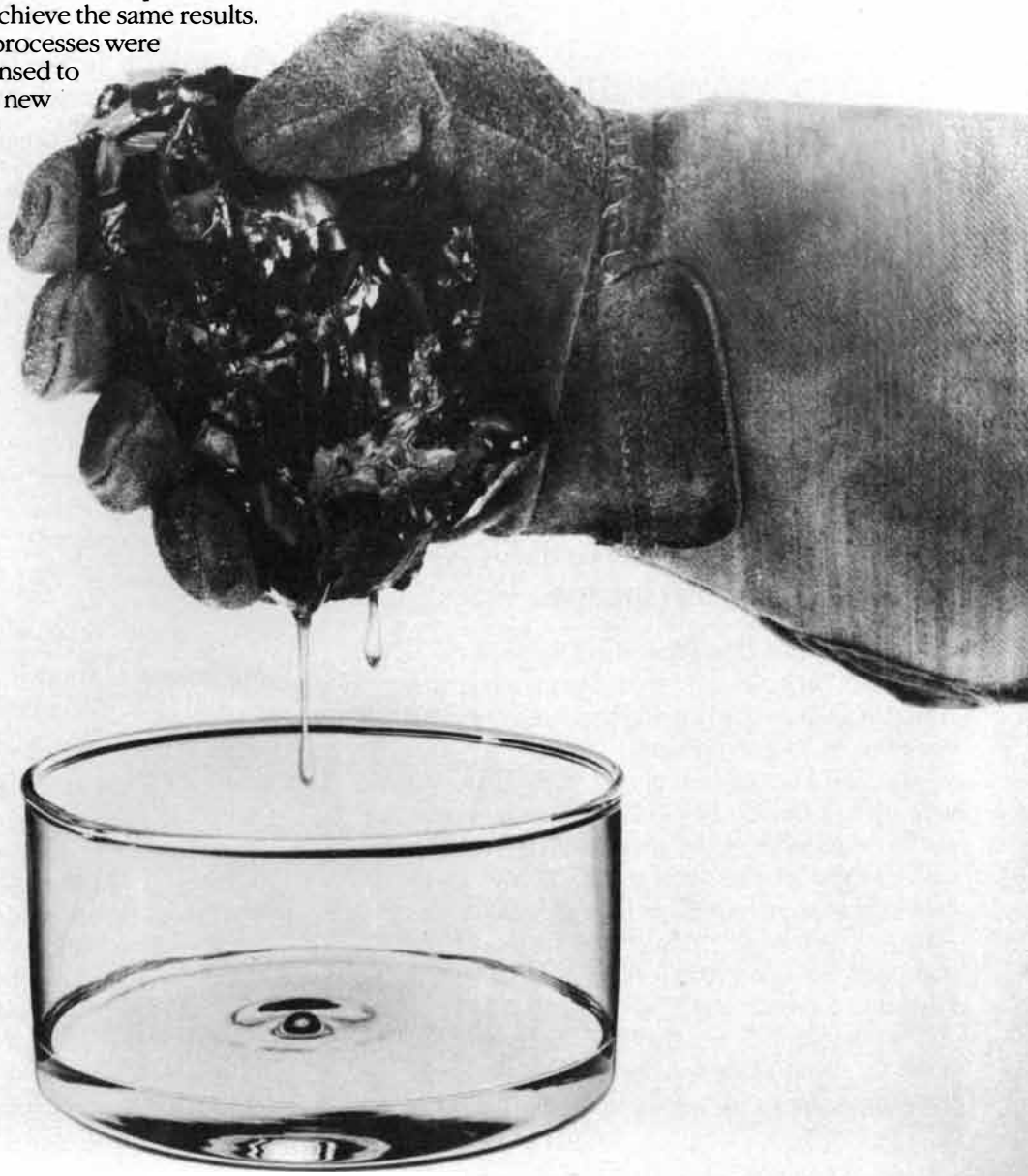
Our innovations include: an energy-efficient method for producing refined methyl acetate, and a new process and catalyst system to produce acetic anhydride.

These achievements in chemical engineering and technology are steps in the right direction. Not only for ourselves (as coal gasification appears to be more economical), but for our country, too. We're now one step closer to independence from foreign oil.

To learn more about construction of a coal-gasification facility, write for our brochure *"Eastman* Chemicals from Coal" to Tennessee Eastman Company, Public Relations Dept, PO Box 511, Kingsport, TN 37662.

Kodak

Kodak. Where technology anticipates need.



Introducing the most powerful

The new IBM Personal Computer AT.

Hold on to your hat.

The new IBM Personal Computer AT (for Advanced Technology) is based on the advanced 80286 16-bit microprocessor. This remarkable computer will run many of the programs written for the IBM PC, up to three times faster. You'll be able to recalculate large spreadsheets in seconds and retrieve files in a flash.

It's got the power (and price) to surprise you. In many ways.

Compatibility, expandability, networking too.

With the IBM Disk Operating System, the IBM Personal Computer AT can use many programs from the fastest-growing library in the personal computer software industry.

The IBM Personal Computer AT is also available with up to 3 million bytes of user memory to run multiuser, multitasking operating systems such as XENIX™. Volume upon volume of information is available at your fingertips. You can customize your system to store up to 20,000 pages of information at one time. And its keyboard helps you use all of this computing power more easily.

This new member of the IBM PC Family is a powerful stand-alone computer that can also be both the primary file server and a station on your network. With the new IBM PC Network (which is so easy to connect you can do it yourself), the IBM Personal Computer AT can share information with IBM PCs, PC/XTs and IBM *Portable* PCs.

IBM Personal Computer AT Specifications

User Memory
256KB-3MB*
Microprocessor
16/24-bit 80286*
Real and protected modes*
Auxiliary Memory
1.2MB and 360KB diskette drives*
20MB fixed disk drive*
41.2MB maximum auxiliary memory*
Keyboard
Enlarged enter and shift keys
84 keys
10-foot cord*
Caps lock, num lock and scroll lock indicators
Display Screen
IBM Monochrome and Color Displays
Operating Systems
DOS 3.0, XENIX*
Diagnostics
Power-on self-testing*
Parity checking*
CMOS configuration table with battery backup*
Languages
BASIC, Pascal, FORTRAN, Macro Assembler, COBOL, APL
Printers
Supports attachment of serial and parallel devices
Permanent Memory
(ROM) 64KB
Clock/calendar with battery*
Color/Graphics
Text mode
Graphics mode
Communications
RS-232-C interface
Networking
High-performance, high-capacity station on the IBM PC Network*

*Advanced Features for Personal Computers

Get a hands-on, hats-off demonstration.

The new IBM Personal Computer AT has the power, compatibility and expandability many PC users need, at a very appealing price.

For more information contact your authorized IBM PC dealer, IBM Product Center or IBM marketing representative. For a store near you call 1-800-447-4700. In Alaska or Hawaii call 1-800-447-0890.

IBM®

personal computer IBM has ever made.

XENIX™ is a registered trademark of Microsoft Corporation.

SCIENCE AND THE CITIZEN

Winter's Blast

The idea that nuclear war could lead to catastrophic global climatic change seems to have caught the attention of the Reagan Administration. In February, George A. Keyworth II, the president's science adviser, asked Alan D. Hecht of the National Climate Program Office of the National Oceanic and Atmospheric Administration to coordinate a plan for a large-scale study of the possibility. The plan, which is expected to be presented to Keyworth this month, calls for $50 million to be spent over the next five years to investigate the impact of various levels and patterns of nuclear attack on global climate. At several national laboratories such computer studies are already under way.

Arousing the Administration's concern is a prediction by several astronomers and atmospheric scientists that a full-scale nuclear war could eject at least 100 million metric tons of soot, dust, smoke and various chemical substances into the atmosphere (see "The Climatic Effects of Nuclear War," by Richard P. Turco, Owen B. Toon, Thomas P. Ackerman, James B. Pollack and Carl Sagan; SCIENTIFIC AMERICAN, August). Such airborne material could form a globe-enveloping cloud reaching as high as 20 kilometers. The absorption of sunlight by the cloud could lower the temperatures at the surface of the earth by an average of 40 degrees Celsius. Richard P. Turco of R & D Associates, Inc., in Marina del Rey, Calif., has called such a disaster nuclear winter; it could lead to the extinction of many species, including man.

One major purpose of the study is to repeat the kind of calculations done by Turco and his co-workers, using more sophisticated models of the atmosphere. In the original investigation only the vertical mixing of the atmosphere with smoke and dust was allowed to vary with time; in the current study investigators will look at mixing in all three dimensions.

Another major goal is to predict how long the particles will stay in the atmosphere. The rate at which such particles fall back to the earth is determined in part by their size, and average size depends on the rate at which they clump together. To make such a prediction suspended matter in the atmosphere must be sampled at various heights during and after large fires. The U.S. and the U.S.S.R. may cooperate in such field studies since neither side would be risking revelation of sensitive technology.

A third goal is to study the scattering and absorption of radiation by the suspended particles. Both the visible solar radiation and the infrared, or heat, radiation emitted by the earth's surface as well as by the smoke particles themselves must be investigated.

Studies of the biological and ecological consequences of a nuclear winter are also planned. Current models of animal physiology and heat balance make it possible to estimate the effects of prolonged cold and darkness on various species. It is much harder to assess the response to such stresses by an ecosystem. How well will seeds survive in the soil? Will plants regenerate after photosynthesis has temporarily ceased?

A nuclear winter has far-reaching implications for military strategy. If any militarily useful first strike would precipitate a nuclear winter, the threat of such an attack would be even less credible than it is now. No plausible plan for civil defense and military follow-up could meet such conditions. The populace would require several years of protection from the cold; troop movements would have to be carried out in almost total darkness.

It is not clear that $50 million is enough to produce the answers being sought. The major revisions needed for the existing computer models of the atmosphere and the detailed field studies that can give initial conditions on which the models can be based are both expensive undertakings. Fortunately several presumed noncombatants, including Australia, Canada and New Zealand, plan to begin their own investigations.

Capitalism for the Poor

The Reagan Administration's views on population have raised eyebrows among many demographers and most political leaders in the rest of the world. The most comprehensive airing of the Administration's demographic concepts was given before the United Nations International Conference on Population, held during the summer in Mexico City. In brief the Administration maintains that desirable population growth rates follow "naturally" from the introduction of a free-market economy. A corollary to the position was stated explicitly by James L. Buckley, the chief U.S. delegate to the conference: Centrally planned economic systems are to blame for the world's population ills.

Given the fact that China, the U.S.S.R. and many of the nations of eastern Europe have all reduced their population growth rates, many demographers are inclined to regard the statement as intended primarily for the consumption of conservatives in the U.S. Aside from the proclaimed elimination of funding to all private organizations that promote or counsel abortion, it is not thought the statement will have much effect on U.S. population policies with respect to the developing nations of the world.

Nevertheless, to the extent the Administration's position can be taken as the basis of policy, workers in the field believe it seriously misrepresents the lessons of demographic history. The "natural" demographic transition to which the Administration makes appeal is the historical shift of the U.S., Japan and the nations of Western Europe to lower fertility and mortality rates during the economic development brought about by the industrial revolution. It is true that in most of these countries improving economic fortune led to lower mortality rates before the fertility rate was lowered. An oversimplified demographic model of 30 years ago held that lower fertility inevitably follows a declining mortality. It is now known, however, that the strict temporal sequence of the earlier model cannot always be applied. The essential lesson of the demographic transition is simply that if economic development is to succeed, population growth must be rather slow.

In contrast, population growth in the developing countries today has often been explosive. Mortality rates are much lower than they were in 18th-century Europe, and they have dropped precipitously. Furthermore, the current declines in mortality are brought about primarily by rapid innovations and improvements in public health; they certainly do not reflect conditions of domestic economic improvement.

Moreover, according to Charles F. Westoff of Princeton University, the historical decline in the birth rate in France was virtually simultaneous with the decline in the death rate. The fertility rate in many developing countries has also failed to conform to the early demographic model. Although the fertility rate has fallen in many developing countries, the decline in the rate has stalled at high levels in some countries; it appears to have stopped altogether in Costa Rica, India and Sri Lanka.

The consensus among demographers is that decline in fertility and the consequent slowing in real population growth in developing nations is affected by a number of factors in addition to economic development. As Samuel H. Preston of the University of Pennsylvania points out, the status of women and the prevailing religious beliefs of a culture may be at least as important.

According to the *World Development Report 1984,* recently published by the World Bank, the correlation between current economic development and fer-

PROJECTION

FINALLY. A GIANT SCREEN TV AS BIG AND BRIGHT AS ALL OUTDOORS... INDOORS.

INTRODUCING FULL SPECTRUM PROJECTION TELEVISION. ONLY FROM RCA.

Now you can see a picture that's as bright and as beautiful as it is big.

Because RCA's new 45-inch set (measured diagonally) features RCA's Full Spectrum audio and video technology, making it the first giant screen TV to process 100% of the audio/video signal. All the color. All the sound. All the time. Surrounding you with bright, sharp color. And clean, crisp sound.

FULL SPECTRUM

COMPLETE WITH FULL STEREO, VIDEO MONITOR CAPABILITY AND MORE.

All Full Spectrum sets feature a built-in stereo broadcast receiver. (Stereo sound dependent on TV broadcast signal.) Plus video monitor circuitry for easy hookup of audio/video components.

Other standard features include 127-channel tuning (including cable) plus RCA's Digital Command Center, the total control remote control that operates an entire RCA video system. And all these features fit in a cabinet measuring just 42¼" x 45".

See this high performance wonder at your RCA dealer. Or, for more information and a free copy of our "Living With Video" book (a $2.50 retail value), write: RCA Consumer Electronics, Dept. 32-312CC, P.O. Box 7036, Indianapolis, IN 46207-7036.

RCA

TECHNOLOGY THAT EXCITES THE SENSES.

Simulated TV picture

TELEVISION

tility rates is far from perfect. The World Bank report shows that in China, Colombia, India, Indonesia, South Korea, Sri Lanka and Thailand fertility rates are below those that would be expected for the level of income. In Algeria, Jordan, Mexico, Morocco, most of the countries of sub-Saharan Africa and Venezuela fertility rates are much higher than they would be if they were controlled by income level alone.

There is a striking irony in the Administration's choice of Hong Kong and South Korea as examples of rapid fertility decline. According to Westoff, a major factor in these declines is a high rate of abortion.

Galactic Dynamo

The Milky Way galaxy, it appears, generates a dipole magnetic field. Writing in *Nature,* Farhad Yusef-Zadeh and Don R. Chance of Columbia University and Mark R. Morris of the University of California at Los Angeles report observing at the galactic center an arc of continuous filaments, visible at radio frequencies. The arc lies near Sagittarius A, a large, complex object that may contain a black hole. The continuous filaments of which it consists are separated from one another by dark bands and extend for a distance of at least 130 light-years; each filament is only three light-years wide (*see false-color map below*).

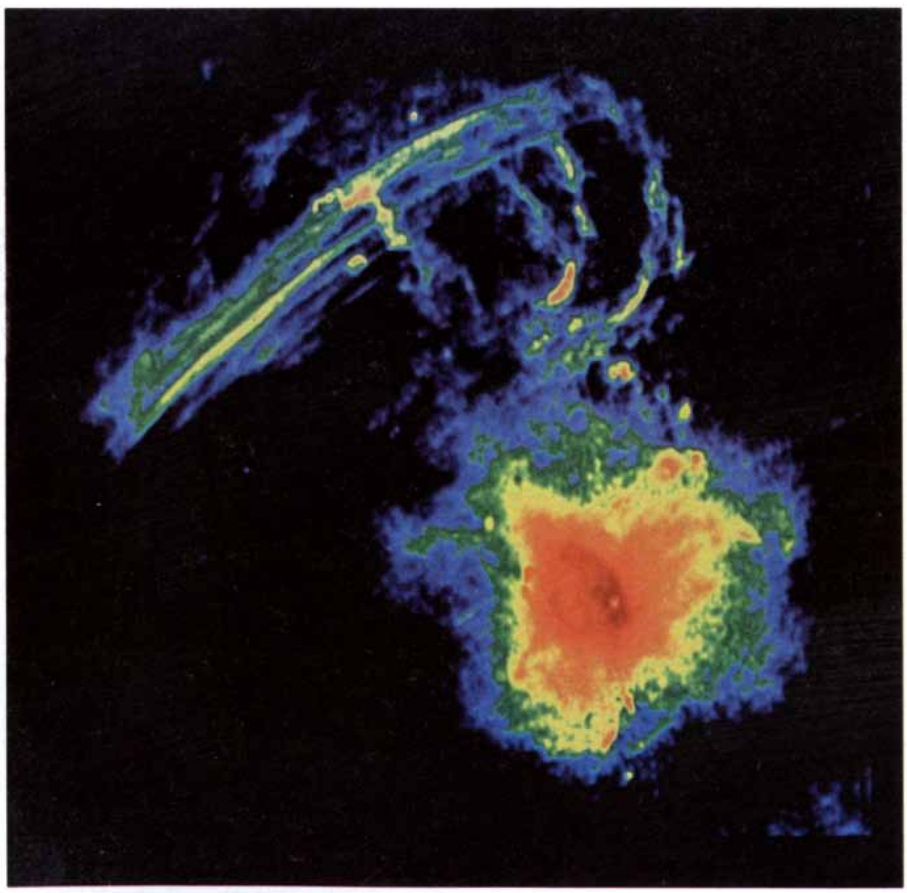
Two facets of the observation strongly suggest that the filaments represent lines of force in a galactic magnetic field. First, the investigators say, the arc is "clearly a physically unified phenomenon." Second, at two points along the filaments a significant fraction of the radio emissions are polarized: the electric field vector vibrates in a direction roughly perpendicular to the axis of the filaments. Polarized emissions are characteristic of synchrotron radiation, which is generated by electrons captured in a magnetic field as they gyrate along its lines of force.

If this interpretation is correct, Yusef-Zadeh, Chance and Morris have the first evidence that the galaxy, like the sun, the earth and some of the planets, has a dipole magnetic field. It is probable that the galactic magnetic field, like its smaller counterparts, is generated by a dynamo mechanism.

How could such a striking feature have been overlooked in a prominent region of the sky? According to Yusef-Zadeh, the problem with earlier radio observations of the arc, made in The Netherlands, was that the telescopes were widely separated from one another. This kind of arrangement sees small features but tends not to detect large ones; the images, or maps, it produces show the trees, but they miss the forest. Thus the emissions from the arc were interpreted as coming from many discrete sources, presumed to be clouds of hydrogen ionized by ultraviolet radiation from young, massive stars within the clouds. (Such thermal emissions would be unpolarized.)

Yusef-Zadeh, Chance and Morris compensated for instrumental myopia by arranging the 27 radio telescopes in the Very Large Array in New Mexico in two different patterns: one pattern in which the distance between any pair of instruments was as much as 11 kilometers and one in which the spacing was about one kilometer. This array, tuned to receive energy at a frequency of 20 centimeters, produces a map of the galactic center that is sensitive to both large and small features.

Future studies may reveal whether the Milky Way is unique in its magnetic properties, or whether other galaxies produce their own fields.



Radio map of the Milky Way center at the 20-centimeter wavelength showing Sagittarius A and filaments generated by the galactic magnetic field

Energy Crisis

Almost anyone who has recovered from the flu or some other viral illness has experienced a lingering fatigue that persists for days or weeks. In a few extreme instances the malaise lasts for months or even years. George K. Radda and his colleagues at the University of Oxford report in *The Lancet* that they have found what may be the basis of this aftereffect, an infection-related metabolic abnormality in the muscles.

Muscle derives energy for work from two sources. Ordinarily the oxidation of carbohydrates and fatty acids liberates energy that converts adenosine diphosphate (ADP) into adenosine triphosphate (ATP), a higher-energy compound. ATP in turn transfers energy to the contracting muscle fibers, reverting as it does so to ADP. When oxygen supplies are inadequate, cells fuel ATP production through a less efficient anaerobic process called glycolysis.

In normal muscle glycolysis generates bursts of energy for strenuous muscular activity such as sprinting, bicycling or playing tennis. Glycolytic metabolism cannot be sustained for long: it yields lactic acid as a waste product, which causes fatigue and muscle pain as it accumulates. Somehow, the Oxford group says, viral infections appear to trigger conversion to this form of metabolism before levels of exertion are reached that require it, and the effect can persist for a long time after the infection.

Radda and his colleagues observed such alteration of metabolism in the arm muscle of a man who, since an attack of chickenpox four years earlier, had experienced crippling fatigue and muscle pain after mild exertion. They

detected the altered metabolism by means of nuclear-magnetic-resonance (NMR) spectroscopy.

NMR spectroscopy allows the observation of metabolic processes in living tissue because certain elements, notably hydrogen, carbon 13 and phosphorus, that are common in organic compounds have an innate magnetism. Consequently, when they are subjected to a strong magnetic field, their nuclei align with the magnetic lines of force. Radio energy is then applied. For each element a particular frequency of radio energy causes the nuclei to resonate so that they flip into an orientation opposed to the magnetic field. As they do so they absorb the radio energy. The frequencies that are absorbed identify different compounds or elements; the intensity of absorption indicates the abundance of each substance.

Radda and his colleagues had their patient exercise his forearm muscles by squeezing a rubber bulb while his arm lay extended in the NMR apparatus. They found that his muscle cells contained normal proportions of ADP and phosphocreatine, a substance involved in the reconstitution of ATP from ADP. Large quantities of lactic acid, however, soon accumulated in the muscle cells as the man exercised. Radda and his group are now conducting NMR studies of working muscle in patients suffering from flu; they report finding the same inappropriate accumulation of acid.

Femtopulse

Lasers that emit intense, extremely short pulses of light are the best means of observing such ultrafast processes as the rippling of energy through a molecule. A first pulse excites the molecules in a sample material; subsequent, weaker pulses measure the time the molecular system takes to return to its initial condition and even reveal some of the energy transfers that occur in the process. The shorter the pulse is, the more subtle and complex are the effects that can be observed.

Writing in *Applied Physics Letters,* Erich P. Ippen, Andrew M. Weiner and James G. Fujimoto of the Massachusetts Institute of Technology have recently reported the generation of the shortest light pulses yet produced: visible, red-orange flashes that last for a mere 16 femtoseconds. (A femtosecond is 10^{-15} second.)

The apparatus employed by the M.I.T. group is an adaptation of one that has become standard in the analysis of ultrafast processes: the mode-locked laser. When the thousands of modes, or discrete frequencies, in which an organic-dye laser oscillates are "locked" in phase, they interfere constructively at some points and destructively at others. The resulting output is a train of short, intense pulses. The duration of a pulse is inversely proportional to the bandwidth of the modes contributing to it.

The most efficient mode-locked laser, a colliding-pulse ring developed by Charles V. Shank and his colleagues at AT&T Bell Laboratories, yields 65-femtosecond pulses. The technique used by both the Bell and the M.I.T. workers to improve on this result—until recently Shank's group held the record with a 30-femtosecond pulse—consists of two steps: stretching the bandwidth in a glass fiber and then compressing the pulse in time by passing it through a pair of diffraction gratings.

"Chirping," or bandwidth stretching, is possible because the speed of light in glass varies slightly with intensity. This nonlinear effect is noticeable only when the intensity of the light is great. Ippen and his colleagues achieve high intensity by amplifying the pulse from a colliding-ring laser and channeling it into a glass fiber a few micrometers in diameter. The speed of the wave components that make up the front of the pulse increases, and their frequency declines; at the back of the pulse the converse is true. In this way Ippen and his co-workers have increased the bandwidth of the pulses by a factor of four.

Ordinarily the stretching of the frequency spectrum would tend to stretch out the pulse in time as well. Through arrangement of the two diffraction gratings, however, the blue-shifted trailing edge as it emerges from the glass fiber can be made to travel a path that is shorter than the path traveled by the red-shifted leading edge. The trailing edge thus tends to catch up with the leading edge, and the duration of the pulse is compressed. The M.I.T. workers have exploited the fourfold increase in the bandwidth of the 65-femtosecond pulses, achieved through chirping, to its theoretical limit, cutting the duration of the pulses by a factor of four.

"A" for Atropine?

For 134 years Nathaniel Hawthorne's *Scarlet Letter* has been regarded as a classic study of the psychological ravages of unacknowledged guilt. Now a physician proposes that it is really the story of a patient's murder by his physician. In an article in *The New England Journal of Medicine* Jemshed A. Khan of the University of Missouri School of Medicine argues that "surreptitious atropine poisoning accounts for the bizarre behavior and ultimate demise of the Reverend Arthur Dimmesdale." The poison, Khan contends, is administered by Roger Chillingworth.

Dimmesdale, a highly regarded divine in 17th-century Boston, conceals his complicity while his partner in adultery, Hester Prynne, is publicly humiliated and condemned to wear a scarlet letter *A* on her bodice. Over a period of seven years the minister's health fails progressively, presumably as the result of a profound gnawing guilt insidiously nurtured by Chillingworth, Dimmesdale's physician and constant companion. Unbeknownst to Dimmesdale, Chillingworth is Hester's estranged and cuckolded husband; he has perceived the minister's guilt and is monomaniacally bent on revenge. At the end Dimmesdale dies, publicly proclaiming his guilt and revealing, "imprinted in the flesh" of his chest, "his own red stigma."

Nathaniel Hawthorne

The author in 1850, when **The Scarlet Letter** *was published*

Chillingworth was not a malevolent forerunner of Freud but an old-fashioned murderer, Khan implies. He marshals an array of evidence. Chillingworth had a clear motive, ample opportunity and the knowledge necessary for preparing and administering repeated small doses of a poison. Hawthorne notes that the physician was expert in herbal medicine and that his study was filled with equipment with which he converted "weeds... into drugs of potency." Khan cites references to poisonous plants: apple of Peru, nightshade, dogwood and henbane. Henbane and nightshade contain atropine. Khan thinks Hawthorne could have learned about such matters in the course of his extensive reading in the Salem Athenaeum; his reading list includes several botany books, in one of which references to deadly nightshade and henbane closely follow each other.

Khan cites clinical evidence as well. He lists the manifestations of atropine intoxication and finds many of them reflected in Dimmesdale's symptoms. Dimmesdale frequently places a hand on his heart, as if in pain. Is this a sign of "angina exacerbated by atropine-induced tachycardia [rapid heartbeat]?" Atropine can produce muscular incoordination; Dimmesdale had a "listless-

ness of his gait," according to Hawthorne, and "tottered on his path." Atropine can induce visual disturbances and hallucinations; that effect might explain the occasion on which Dimmesdale sees an immense *A* marked out in the sky in lines of red light. Atropine's suppression of salivation can cause speech difficulties and make it hard to swallow food; Dimmesdale's voice became tremulous and "his form grew emaciated." The scarlet letter revealed on Dimmesdale's chest at the end, Khan concludes, was "a diffuse nonpunctate erythematous rash"—the typical rash of atropine poisoning.

Khan's imaginative interpretation will be scrutinized critically by scholars of American letters. Philip Young, professor of English at Pennsylvania State University, for one, disagrees with Khan and offers an even more surprising interpretation of the novel. In addition to general arguments based on the nature of Hawthorne's major concerns in *The Scarlet Letter,* Young cites two statements in the book. As Dimmesdale is dying Chillingworth cries repeatedly: "Thou hast escaped me!" Clearly, Young argues, Chillingworth was not a murderer: he did not want Dimmesdale to die. Moreover, Hawthorne appears to specifically dismiss the poisoning hypothesis. He writes that whereas some townspeople thought the physician had caused the *A* to appear on Dimmesdale's chest "through the agency of... poisonous drugs," those "best able to appreciate the minister's peculiar sensibility and the wonderful operation of his spirit upon the body" believed the letter "was the effect of the ever-active tooth of remorse."

Young characterizes his own interpretation as a "more radical notion than Dr. Khan's." He thinks Hawthorne had in mind "a sin far graver than adultery." In Hawthorne's mind, Young says, the letter was not an *A* but an *I*—for incest. Young's evidence for this major reinterpretation of *The Scarlet Letter* is adduced in his book *Hawthorne's Secret: An Un-Told Tale,* which has just been published by David R. Godine. It is Young's thesis that the relation between Hester and Dimmesdale reflects Hawthorne's preoccupation with his family history, which included several instances of incest, and Hawthorne's sense of his own relation with his sister Elizabeth.

Little Mystery

The discovery of a new fundamental particle has put theoretical physicists in the position of a mechanic who assembles an engine only to find that there is a part left over.

The particle, the zeta, discovered at the Deutsches Elektronen-Synchrotron (DESY), is not predicted by theory. Therefore its discovery could significantly change existing concepts of the structure of matter. Elliott D. Bloom of the Stanford Linear Accelerator Center and Hans Bienlein of DESY led the group that discovered zeta in the course of an investigation of the decay of the upsilon meson.

The upsilon meson has a mass equivalent to an energy of about 9.5 GeV (billion electron volts). In some instances its decay yielded a photon with an energy of about 1.1 GeV, and a shower of other particles that appear to be the decay products of a briefly extant companion to the photon. This companion, the mass of which was 8.3 GeV, is zeta.

It has been suggested that zeta is the Higgs particle, whose existence was predicted by Peter Higgs of the University of Edinburgh. The Higgs particle plays an essential role in the unification of the electromagnetic and weak fields. Zeta's characteristics, however, do not precisely match those of the Higgs particle. Further work will be needed to define the significance of zeta.

Emotional Chemistry

Do depressive emotional illness and its polar opposite, mania, have a biochemical origin? According to current theories the answer is yes. Such affective, or emotional, disorders, theory holds, result from abnormalities in the release or uptake of neurotransmitters.

In *The New England Journal of Medicine* N. Suzan Nadi of Wayne State University and her colleagues at the National Institute of Mental Health present evidence that supports the theoretical view. Their finding concerns the neurotransmitter acetylcholine and its receptor. Both are essential elements of the cholinergic, or parasympathetic, nervous system. The balance between this system and the adrenergic, or sympathetic, system (the main neurotransmitter of which is noradrenalin) regulates involuntary functions such as heart rate.

Both kinds of neurotransmitter and receptor are found in the brain. Here, it has been proposed, affective illness results from an imbalance in the two systems: depression from an excess of acetylcholine or of its receptors; mania from corresponding abnormalities of the adrenergic system.

Nadi and her colleagues sought to confirm that an increased number of cholinergic receptors is associated with depression. Because of the difficulty of sampling brain tissue, they assayed cholinergic-receptor density in skin fibroblasts, or connective tissue cells, which resemble neurons in some respects.

In all the 18 depressive or manic-depressive patients whose skin the investigators biopsied they found receptor numbers that stood well above the average in the 12 normal controls. The excess persisted through many cell-culture generations. It was found both in former patients who were clinically well and in acutely ill patients.

The investigators suggest a genetic excess of cholinergic receptors predisposes people to depressive and manic emotional disorders. They concede it is difficult to understand how the same biochemical abnormality could accompany both depression and mania.

To confirm their hypothesis Nadi and her colleagues plan to measure the number of skin receptors in the children of manic-depressive parents, a group with a high risk of future affective illness. Follow-up studies, says Nadi, will enable the researchers "to see whether or not the receptor number predicts who will eventually develop mental illness."

Gene Extractor

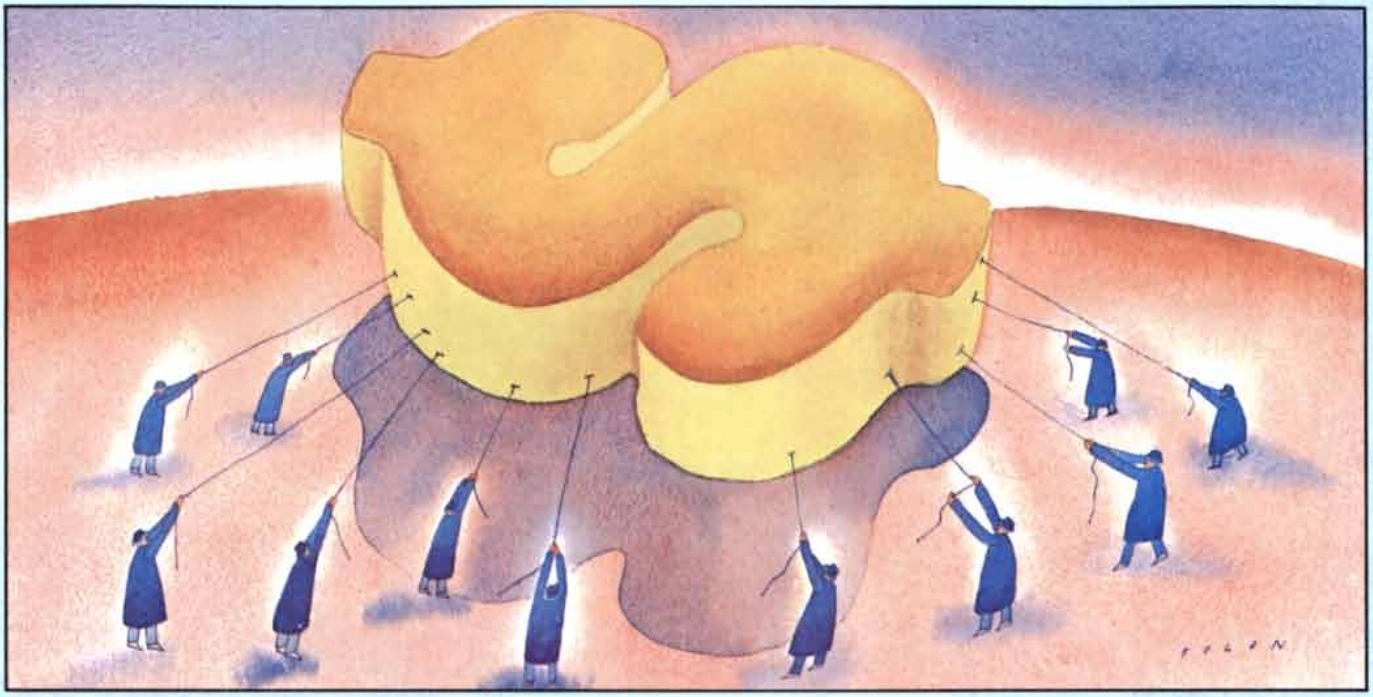
Molecular biologists may have a new tool for their kit: an enzyme that can separate an organism's complement of DNA into its component genes. The enzyme could be a major improvement on the classic method of isolating genes in which the molecular message is reconstructed using messenger RNA (mRNA) as a template.

The promising new enzyme is found in the mung bean plant, and so it bears the name mung bean nuclease. Its capacity to cut genes apart was first observed by Thomas McCutchan of the National Institute of Allergy and Infectious Diseases. McCutchan's group is collaborating with a team headed by Louis H. Miller at the National Institutes of Health in work aimed at developing a malaria vaccine.

In the vaccine project at the NIH it became important to isolate from the human malaria parasite *Plasmodium falciparum* a cell-surface protein known as the circumsporozoite (CS) protein. The nuclease proved superior to the classical method for obtaining the gene. A group led by Vincenzo Enea and Ruth S. Nussenzweig of New York University Medical Center cloned the CS-protein gene using the mRNA method. Because sporozoites cannot be grown in the laboratory, accumulating enough mRNA to isolate the gene was and is a laborious process; it took the N.Y.U. group two years to obtain enough mRNA to clone the CS-protein gene.

Mung bean nuclease, on the other hand, operates directly on the genes. Therefore it can be used to extract the CS-protein gene from a stage of the parasite that will grow in the laboratory. Using the nuclease, McCutchan's group was able to clone the gene for the CS protein in about eight months.

According to McCutchan, mung bean nuclease could have several significant scientific and clinical applications. It could be employed to produce a complete "library" of the genes of a particu-



TOMORROW'S ELECTRICITY

Can we hold down the cost?

How can the price of America's electricity be kept reasonable? It turns out that nuclear energy may be part of the answer.

The cost of electricity depends mostly on two things: the cost of the power plant, and the cost of the fuel used to generate the power.

How nuclear electricity saves money

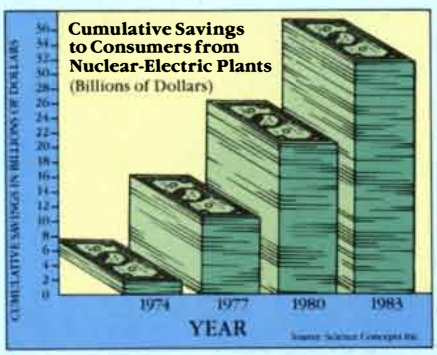
Nuclear power plants cost more to build but much less to fuel than power plants burning coal, oil, or natural gas. That's because 1) nuclear plants need relatively small amounts of uranium because of its high energy content, and 2) uranium is a lot less expensive than the fossil fuels it would take to produce an equal amount of energy.

There are 85 nuclear power plants operating in 28 states. These plants have saved American consumers more than $30 billion over the last ten years, compared to what the electricity would have cost coming from coal and oil.

Even the nuclear power plants now being built, which are more expensive because of higher construction costs, can save money compared to the fossil-fuel alternatives. The U.S. Congress's Office of Technology Assessment makes this estimate: A new nuclear plant will save consumers at least $3.5 to $5 billion over the next 30 years, compared to what the electricity would cost from an oil-fired plant. In the long run, "electricity ratepayers would be better off" with the nuclear plant.

Nuclear power is cheaper energy for tomorrow

After a while, a nuclear plant's lower fuel cost more than makes up for its higher cost of construction.



This graph shows how much money has been saved over the past 10 years by using nuclear power. If the nuclear-generated electricity had come instead from coal-fired and oil-fired power plants (75% coal and 25% oil), then America's electric bill would have been more than $30 billion higher. And the nuclear plants now operating will continue to save $4 to $6 billion annually.

Critics of nuclear power point to the relatively high price of electricity from a new plant's early years of operation, as the cost of construction is being paid for. But they ignore the cumulative savings a typical nuclear plant provides to ratepayers over its full operating life. Nor do they mention the proven savings of the plants already on line.

"On the basis of sheer economics, nuclear power can still be a bargain for virtually every region in the country," says Peter Navarro, energy researcher, Harvard University, and author of *The Policy Game*.

New fuels and technologies may one day play a bigger role in America's electric system. But realistically, they'll still be only part of a full team of national energy resources. With coal and nuclear energy as principal players, that team will be able to do its job, efficiently and economically, for many generations to come.

For a free booklet on the electrical future of America, write to the U.S. Committee for Energy Awareness, P.O. Box 37012 (E6), Washington, D.C. 20013. Please allow 6-8 weeks for delivery.

Information about energy America can count on today

U.S. COMMITTEE FOR ENERGY AWARENESS

We didn't have to make it the most advanc Cadillac ever.

But we did.

We know that doing things we don't have to do helps make a Cadillac a Cadillac. For example, Cadillac didn't have to give the 1985 Fleetwood and De Ville the performance of the world's only transverse-mounted V8 in a front-wheel-drive production car. But we did.

Cadillac didn't have to give these six-passenger cars even more front seat headroom and legroom than last year. But we did.

Cadillac didn't have to add a new standard center high-mounted stop lamp to give following drivers an additional warning of braking. But we did.

Cadillac didn't have to use an aerodynamic door-into-roof design for reduced wind noise and easy entry and exit. But we did.

Cadillac didn't have to include the convenience of power windows which can operate up to ten minutes after you turn off the ignition (or until a car door is opened). But we did.

1985 Fleetwood

Cadillac didn't have to protect these cars with the most extensive anti-corrosion treatments in our history. But we did.

And Cadillac didn't have to back these cars for 4 years/50,000 miles* But we did.

You don't have to fall in love with the most advanced Cadillac ever. But you will.

Cadillacs are equipped with engines produced at facilities operated by GM car groups, subsidiaries or affiliated companies worldwide. See your dealer for details *Limited Warranty. In some cases, a deductible applies. See your dealer for details.

Best of all...it's a Cadillac.

lar organism. The enzyme might also be a valuable tool for exploring the genetic origin of metabolic disorders and other clinical conditions.

Finally, McCutchan notes: "The enzyme is interesting in its own right. It could be very informative about the structure of DNA. For example, it could be that the signal that triggers the cutting at the ends of the genes has a role in controlling gene expression."

An unanswered question is whether the enzyme will perform its selective cutting on the genes of multicellular organisms, which differ slightly in structure from the genes of unicellular organisms. In view of the great potential of the mung bean enzyme, it is possible that such questions will get quick answers.

Pollution Solution

Bacteria that thrive on the hydrogen sulfide (H_2S) that emerges from hydrothermal vents on the deep-ocean floor may provide a means of dealing with the problem of acid rain and other forms of sulfur pollution.

The suggestion comes from two microbiologists, Craig Taylor and Holger Jannasch, who are doing experiments with sulfur bacteria at the Woods Hole Oceanographic Institution. Such bacteria are natural denizens only of the ocean deep. They are chemosynthetic, that is, they get from a chemical reaction (the oxidation of hydrogen sulfide) the energy they need for metabolism and growth. Taylor and Jannasch report they have maintained a culture of analogous bacteria for two years. They believe the technique can be scaled up so that the bacteria would turn certain forms of industrial waste into a source of energy while reducing pollution.

An assemblage of microorganisms would be established in gaseous or liquid waste containing hydrogen sulfide, a precursor of acid rain, and carbon dioxide. Such wastes are common in the gasification and burning of coal, the production of geothermal energy, the treatment of sewage and other processes. The bacteria would oxidize the hydrogen sulfide and other sulfur compounds and utilize the energy to convert carbon and nitrogen into more bacteria. This biomass could then be employed to make synthetic fuel (methane) or chemical feedstocks. It could also serve as a food in the aquaculture of such marine organisms as mussels.

Electron Snooper

A transistor so sensitive that it can "hear" a single electron has been developed. The minute device, the dimensions of which are one micrometer by .1 micrometer, has been produced at AT&T Bell Laboratories. It is designed to enable investigators to study the troublesome phenomenon of electron trapping in the silicon chips from which transistors are made.

As an electron moves through the crystalline lattice of a transistor, it can be trapped by irregularities in the structure. As electrons are trapped and escape from such a defect, low-frequency noise is generated that interferes with signal propagation in the transistor.

The device reveals in detail precisely what happens to a single electron in a single trap, according to William J. Skocpol, a member of the group that developed the highly sensitive device.

One result of experiments with the device has been the confirmation of a hypothesis that the electron traps in a transistor are near the interface of the silicon and oxide materials. The experiments also yielded the first measurements of the distance of traps from the interface and of the energies of trapped electrons. In the long run the device may aid the manufacture of more efficient semiconductor chips.

The Antibody Knows

In its early stages a tumor exists first as a single, malignantly transformed cell and then as a microscopic collection of such cells. If it could be detected at this point, eradication might be possible before the tumor grows large enough to threaten health and life.

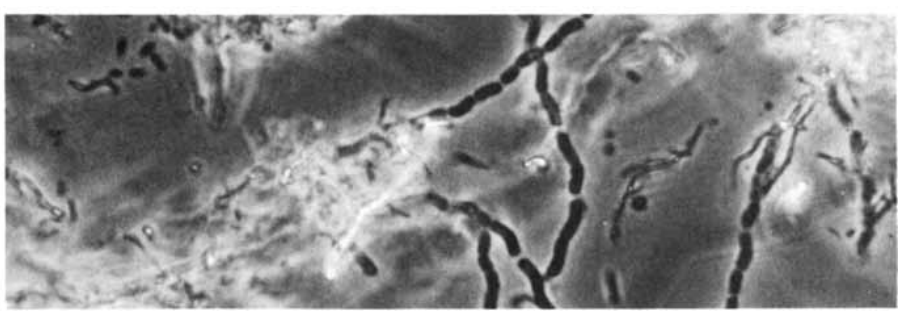
Work recently reported in *The New England Journal of Medicine* by Philip J. Moldofsky and his colleagues at the Fox Chase Cancer Center and the Wistar Institute of Anatomy and Biology indicates that such detection may be becoming a part of conventional practice. Moldofsky and his colleagues employed monoclonal antibodies to detect microscopic metastatic tumors in the lymph nodes of a patient who was known to have cancer of the colon.

Since a monoclonal antibody will bind to one specific cell-surface antigen only, it can be used to locate cells that bear such an antigen. Moldofsky and his group used a monoclonal antibody for an antigen found on the surface of human colorectal carcinoma cells; when they labeled samples of this antibody with a radioactive isotope of iodine and administered them intravenously, they found that some of the labeled antibodies accumulated in the patient's lymph nodes. Visual and X-ray inspection had shown no abnormality, but when surgeons removed one of the lymph nodes and examined it microscopically, they found that the tumor had indeed spread there.

Monoclonal antibodies can also be used for diagnosis outside the body of the patient. A blood sample can be exposed to tumor-specific antibodies in a test tube; antibodies that have bound to tumor antigens are then found by filtering and scanning the serum sample. Monoclonal-antibody-imaging kits for such use may soon be available to physicians in the U.S.

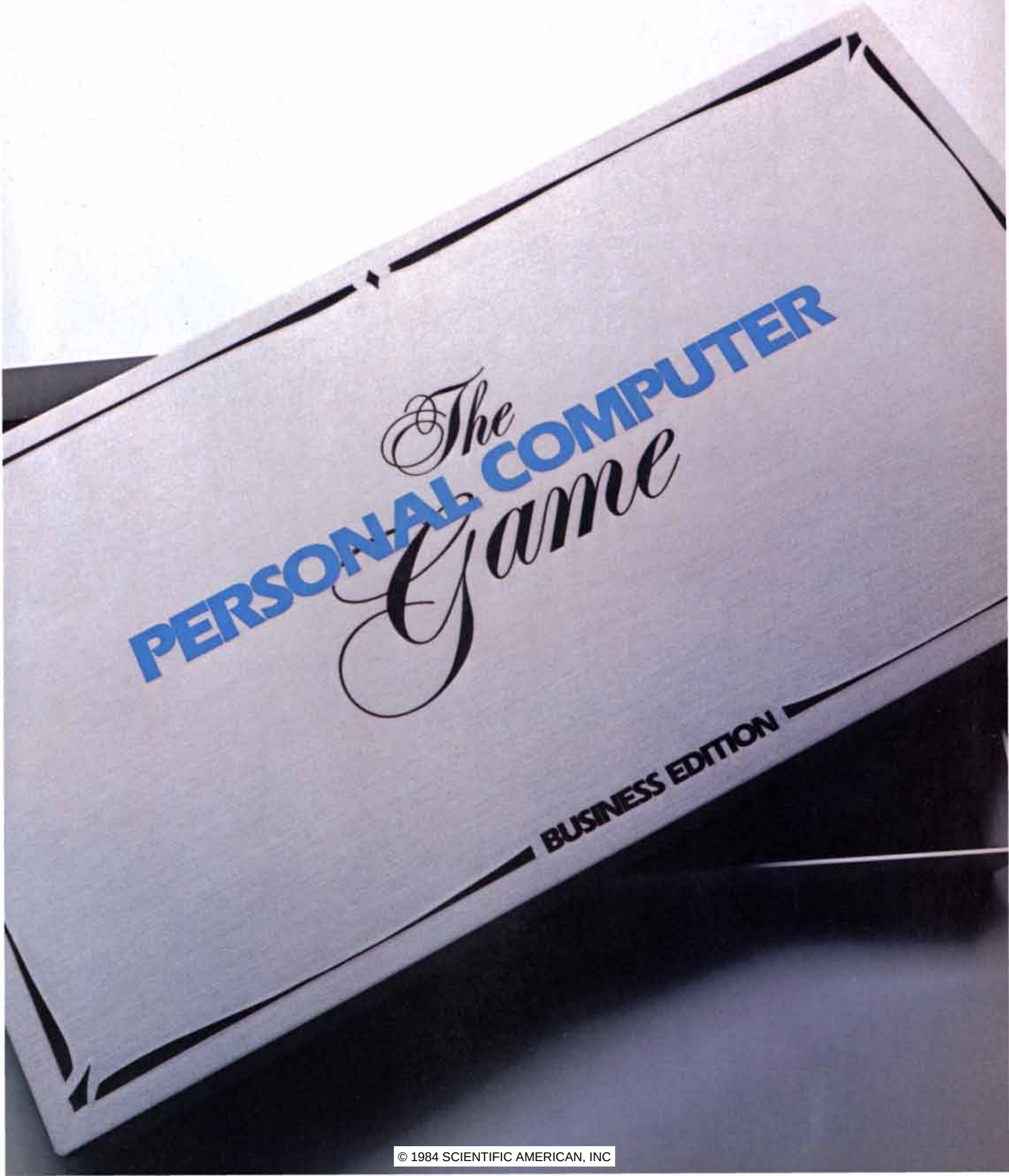
The difficulty of making monoclonal antibodies is the chief impediment to their widespread use. The process begins when mice are injected with samples of a tumor. Their *B* lymphocytes (cells that produce antibodies) are then isolated and fused with myeloma cells (immortal tumor cells) to produce what is called a hybridoma. The hybridomas are screened to identify the ones making antibodies that will bind to the target antigens, and these cells are grown as single clones. Since some antibodies that bind to tumors will also occasionally bind to normal cells, each potential antibody must be tested to ensure that it is truly tumor-specific. Antibodies can be too specific as well: those that bind the tumor cells of one person may not respond to the tumor cells of another. The development of a good clone can take several years.

The detection of metastasized tumors suggests the possibility that monoclonal antibody tests could be used to screen for malignancy in apparently healthy individuals or in former cancer patients who are in danger of relapse. Such screening, however, could raise a significant problem: it is believed the immune system regularly recognizes and destroys malignant cells. Early diagnosis, if followed by therapeutic intervention before the patient's immune system has had a chance to fight back, might needlessly expose some patients to the potentially deleterious side effects of surgery or chemo- and radiotherapy.



Sulfur-metabolizing bacteria that may help to mitigate the effects of acid rain

AND NOW THE GAME BEGINS



INTRODUCING THE AT&T

PERSONAL COMPUTER.

The personal computer game is ready for business—and now it's your move. To win the game your company needs a fast, flexible, reliable personal computer that enables you to call the shots.

Go directly to the new AT&T Personal Computer.

The AT&T Personal Computer is more than just a superior piece of hardware. Behind it is a commitment to a better way of managing information and communications. One that can pay off handsomely in increased productivity and profits.

Because our AT&T PC is designed to be flexible enough to meet all your business needs—today and in the future.

Whether you use it as a stand-alone workstation or as a team player in a fully integrated system, you'll find our PC a high-performance machine. And a hard-working addition to your office.

Its fast processing and high-resolution graphics will help make any computer task a computing pleasure. In addition, you can expect more standard features. More expansion slots. More options for future growth.

OFF-THE-SHELF SOFTWARE DOES THE JOB

Driven by the MS-DOS* operating system, the AT&T PC runs the most popular off-the-shelf software for a wide range of business applications. Financial analysis. Forecasting. Budgeting. Word processing. Inventory. The AT&T PC does it all.

And its flexibility means that when it's time to expand, our PC will actually make your computer growing pains painless.

With our unique PC Interface it can be linked to the more sophisticated, higher capacity world of the UNIX** System V Operating System—the AT&T Computer "brain" that is emerging as the operating system standard for multi-user, multi-tasking machines.

THE AT&T TRADITION CONTINUES

Some things about our PC cannot be measured in bits and bytes, but are of immeasurable value.

For instance, the unmatched service and support of AT&T. The built-in reliability—and outstanding quality—of our products. The century-long tradition of technological innovation and personal attention to detail.

Think about it. Then make your move—to the AT&T Personal Computer, from AT&T Information Systems.

To get in on the game, call your AT&T Account Executive or 1 800 247-1212.

AT&T INFORMATION SYSTEMS. WHEN YOU'VE GOT TO BE RIGHT.

AT&T

*MS-DOS is a trademark of Microsoft Corporation.
**UNIX is a trademark of AT&T Bell Laboratories.

Cartilage

It holds spaces for tissues in the embryo and then it cushions the body. Its fundamental properties of strength and resilience are now explained in terms of the tissue's molecular structure

by Arnold I. Caplan

There is a tissue that performs its crucial functions throughout the body although it lacks most of the attributes of other body tissues: it has no nerves, no blood vessels and no lymphatic system. It is a tissue whose properties are established not by the properties of its cells but by what the cells secrete: an elaborate network of giant molecules the cells deposit around themselves to form an extracellular matrix. The matrix includes some of the largest proteins made by any cell in nature. It also includes great volumes of water. Indeed, it is the structuring of water that yields the tissue's properties.

The tissue is cartilage, a fundamental biological material that helps to shape the body and then helps to support it. In the embryo, cartilage forms a framework on which the major bones of the body (except the bones of the head) take form. Then, in the growing body, cartilage composes the zones called growth plates, which are found at the tips of bones. The bones extend their length from them, in what amounts to a continuation of processes initiated in the embryo. Throughout life, cartilage provides a covering for the bones that withstands compressive loads yet enables the bones to move smoothly against one another. It also cushions the bones of the vertebral column. Throughout life, moreover, the breakage of a bone is followed by the reemergence of cartilage and its replacement by new bone, in what seems to be a recapitulation of the embryonic events.

The study of cartilage promises to illuminate not only the development of parts of the body but also the process of aging and the nature of diseases such as osteoarthritis. One aspect of that study is the analysis of the molecular structure of cartilage. It is growing in importance, because the functions of cartilage—its provision of a framework for replacement by bone and its cushioning of joints—can now be explained, at least in part, in terms of the tissue's molecular structure. Cartilage consists of cells called chondrocytes, which synthesize and deposit around themselves a matrix of giant molecules. In turn the chemical and structural properties of the molecules produce tensile strength (the ability to distribute weight and resist breakage) and resilience (the ability to resume an original shape and texture after a deformation). The tensile strength of cartilage is due to collagen: a rigid, rodlike molecular assemblage found also in places such as the skin, the lungs and the eyes. The resilience of cartilage is due to a remarkable family of molecules called proteoglycans.

Limb Development

Cartilage first appears in the embryo at a time when the limbs have begun to extend. Its onset is quite precisely timed; in the limb buds of a chick embryo (an organism whose limb buds are easily accessible to experimenters) the onset occurs after 4.5 days of incubation. At that time limb-bud cells in the mesoderm, the embryo's middle layer, become irreversibly committed to one of several developmental pathways leading to different types of limb tissue: the cells will express themselves as cartilage, bone or connective tissue.

The commitment is cued by a number of factors, most of them outside the cells themselves. Some of them are provided by the developing vascular system. Initially the limb bud of the chick embryo is nurtured by a uniform spray of capillaries [*see illustration on page 92*]. Soon, however, one of the capillaries near the center of the limb becomes enlarged. This vessel will form the major limb artery. Near it, in the core of the limb, all capillaries disappear, probably because the cells at the very center of the limb produce a substance that inhibits vascularization. (This substance seems to be present in cartilage throughout the life of the tissue.) By the fourth day of the embryo's development (half a day before limb cells become committed to particular developmental paths) the disappearance is complete. Toward the periphery of the limb, the major artery ramifies into a complex capillary bed; thus the core cells and the peripheral cells of the limb enjoy quite different supplies of nutrients, oxygen and whatever substances are circulating in the blood. Cartilage develops in the avascular core of the limb.

In 1967 Edgar Zwilling and I, working at Brandeis University, undertook to mimic the "microenvironments" that influence cells from place to place in a developing embryo. Our intent was to identify circumstances that promote the differentiation of cells into chondrocytes. We found that "seeding" undifferentiated limb-bud mesodermal cells into a culture dish at high density (five million cells in a 35-millimeter dish) induced most of the cells to become chondrocytes. At intermediate density (two million cells per dish) very few chondrocytes—usually none—developed. We also found that the exposure of high-density cultures to high concentrations of oxygen or to nicotinamide (vitamin B_3) prevented cartilage from forming. Nicotinamide was of interest to us because it is found in the embryo. The vascular system supplies it to the embryonic chick limb from storage sites in the yolk of the egg in which the embryo develops. Nicotinamide affects chemical reactions in both the cytoplasm and the nucleus of mesodermal cells and so affects the process of differentiation.

Some recent experiments done in my laboratory at Case Western Reserve University also employed chick limb-bud mesodermal cells that had not yet expressed themselves. In one such experiment hyaluronic acid, an important constituent of the extracellular matrix of embryonic tissues, was chemically bonded to the bottom of culture dishes. Hyaluronic acid consists of long repetitions of a sequence of two sugar molecules: *N*-acetylglucosamine and glucuronic acid. Its appearance or disappearance from developing tissues seems to be related to the transitions the tissues make from undifferentiated structures to differentiated ones. Limb mesodermal cells were plated onto the hyaluron-

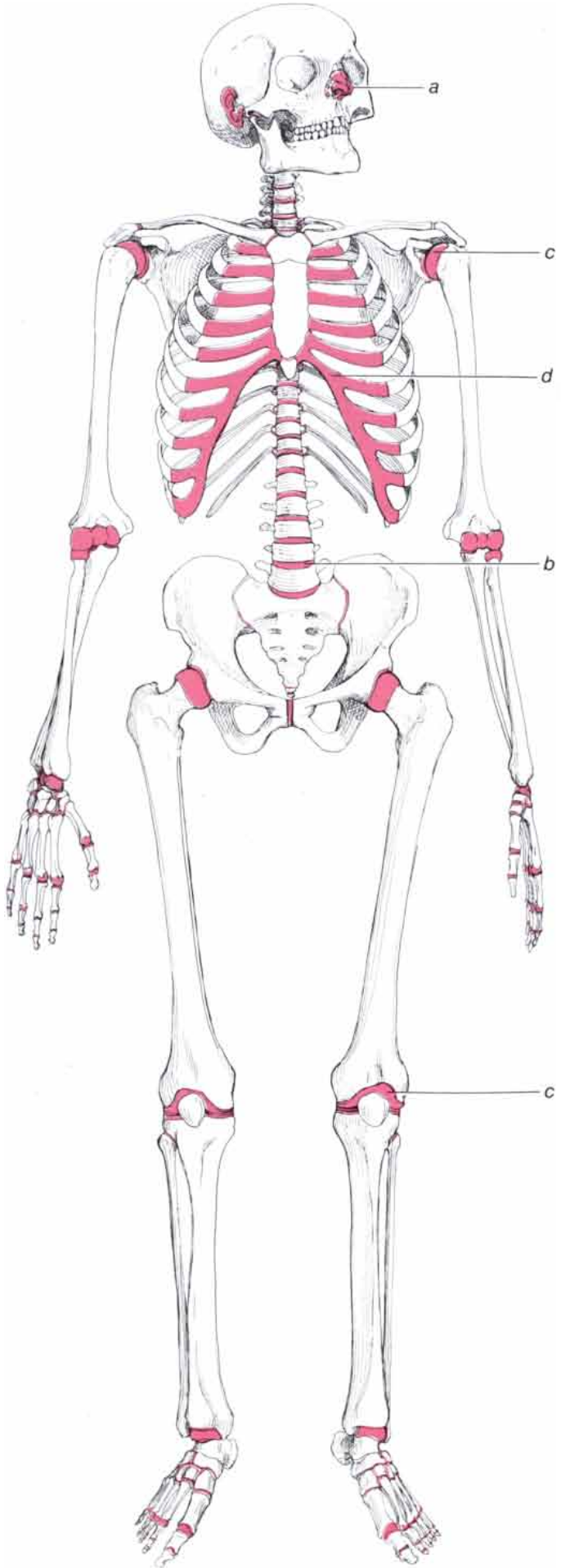
ic acid at the intermediate packing density of two million cells per dish. Most of the cells became chondrocytes.

In sum, the circumstances affecting the differentiation of limb-bud cells in a culture dish include the packing density of the cells, the substances to which the cells attach themselves and the identity and quantity of nutrients in the medium bathing the cells. An additional external factor emerged from still another experiment. Here cells from embryonic chick limbs were cultured at the intermediate packing density and exposed to extracts from demineralized adult bone or to the demineralized bone itself. (In demineralized bone, calcium and phosphates are leached out by dilute acid.) Again the cells became chondrocytes.

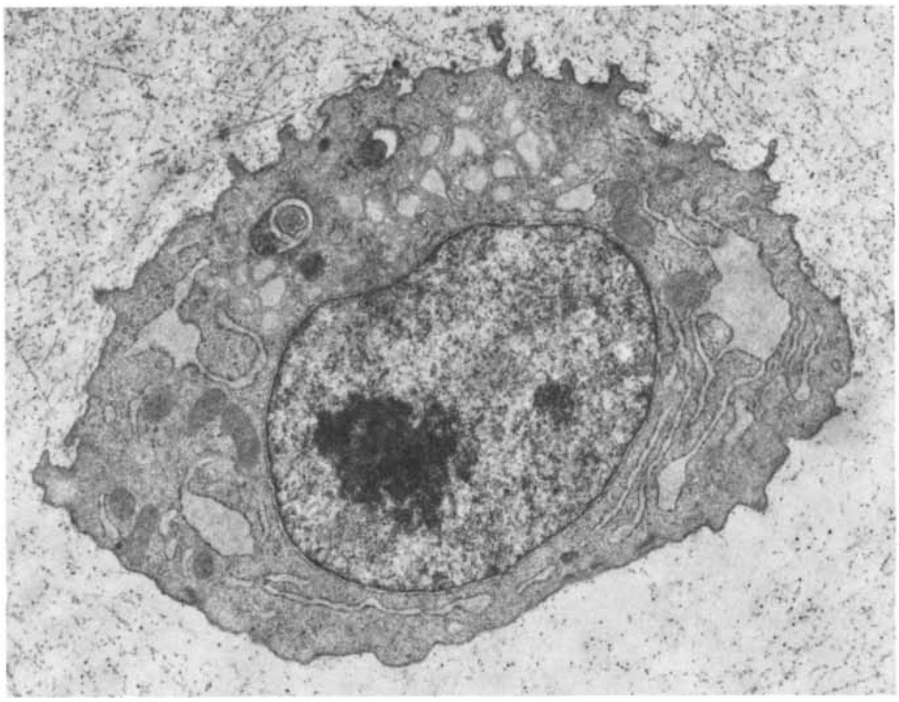
We suspect that the differentiation of cells into chondrocytes under the influence of demineralized bone or bone extracts in a culture dish is much like what occurs when a bone breaks in the body. In particular we infer that adult bone contains substances that signal cells from the circulatory system to clean out the breakage site, substances that summon undifferentiated cells to populate the site and multiply, and substances that cause these undifferentiated cells to become chondrocytes. The chondrocytes then produce a cartilaginous matrix that fills the break and joins the bone fragments together. Finally, the cartilage undergoes a developmental program in which it is calcified and then replaced by new living bone.

Cartilage Macromolecules

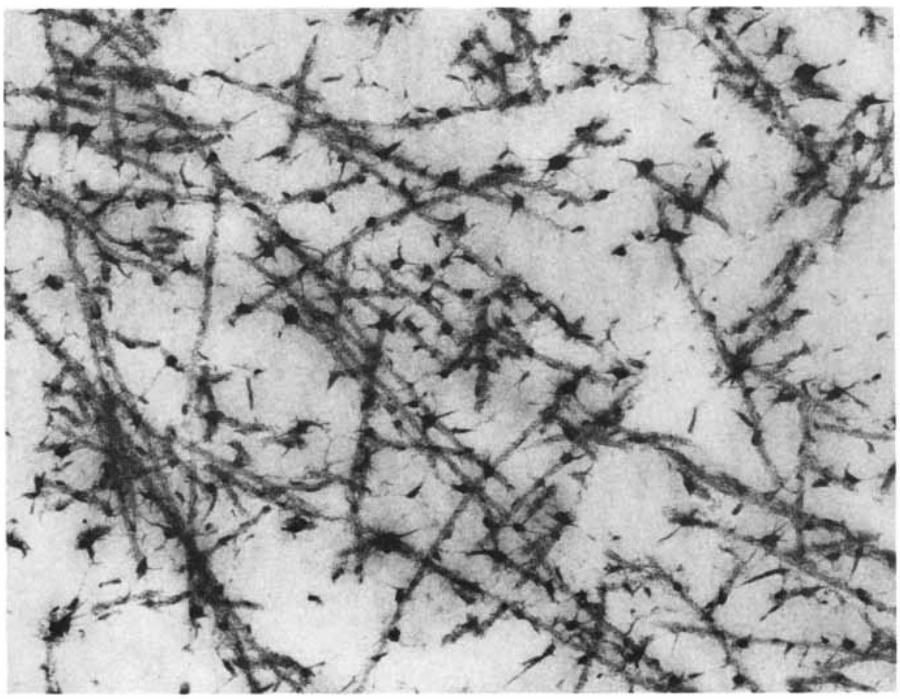
The molecular components of cartilage—the molecules now known to be responsible for the tensile strength and resilience of the tissue—have been identified over a period lasting for more than a century. In 1837 Johannes Peter Müller obtained solutions of what he called chondrin by steaming cartilage at high pressure. Half a century later C. S. W. Krukenberg isolated the chief component of chondrin and identified it as "chondroitin sulfate." Then in 1955 Eugene A. Davidson and Karl Meyer of Columbia University showed that

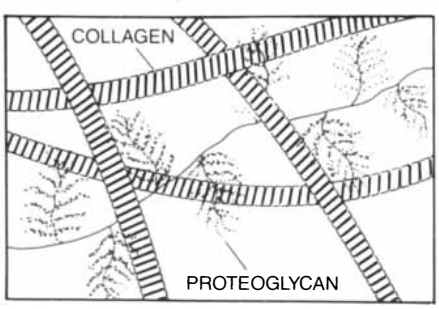


DISTRIBUTION OF CARTILAGE (*color*) in the human skeleton shows that the tissue has four varieties. The nose (*a*) includes morphological cartilage: a genetically sculpted tissue that holds a shape in the body. The intervertebral disks (*b*) include fibrocartilage, a tissue notable for its tensile strength, or ability to resist breakage. The joints (*c*) include articular cartilage, which (in addition to resilience, or ability to rebound from compression) enables bones to slip over one another smoothly. The front of the rib cage (*d*) is elastic cartilage, which enables the cage to expand with the lungs. The molecular structure of each type of cartilage governs its properties.



CHONDROCYTE, a cell that populates cartilage, was grown in a culture dish by "seeding" the dish with cells from the limb of a chick embryo. The nucleus of the cell is large and round; surrounding it are flattened cisterns called endoplasmic reticulum. They are the sites at which cells synthesize proteins. In particular the chondrocyte secretes and deposits around itself a matrix of giant molecules. Here the matrix appears as a feltwork surrounding the cell. The chondrocyte is shown in an electron micrograph at a magnification of about 9,000 diameters.





EXTRACELLULAR MATRIX of immense molecules made by chondrocytes gives cartilage its properties of resilience and tensile strength. In this electron micrograph, shown at an enlargement of about 61,000 diameters, the matrix produced by cultured chick-embryo chondrocytes was stained with the dye Alcian blue. The straight lines are fibrils (rodlike aggregates) of the protein collagen. The dark-staining bodies are collapsed masses of giant molecular constructions called proteoglycans. The drawing schematizes the matrix.

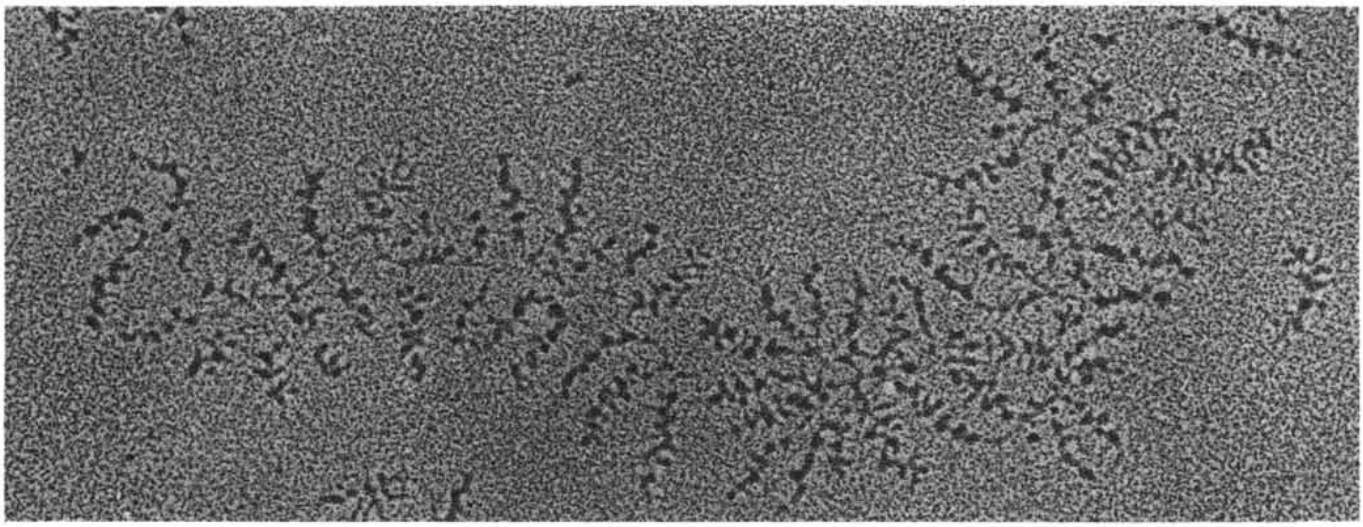
chondroitin sulfate is a repeating disaccharide, a type of polysaccharide, or polymer made up of sugars. Specifically it consists of glucuronic acid and sulfated *N*-acetylgalactosamine in alternation.

The repeating disaccharide proved to be linked at one of its ends to a long protein by a four-sugar chain that came to be known as the linkage region. In 1964 John Gregory, Torvard Laurent and Lennart Rodén of the University of Uppsala established that the linkage region consists of xylose (bound to a particular amino acid, serine, on the long protein), then a sequence of two galactose molecules and finally glucuronic acid (to which the chondroitin sulfate polymer is attached). In 1953 another component of cartilage had been characterized, at least partially, by Davidson, Meyer, Alfred Linker and Bernard Weissman. Called keratan sulfate, the substance was thought at first to be a self-contained constituent of cartilage. Later, however, it became clear that keratan sulfate and chondroitin sulfate attach to the same long protein. Keratan sulfate is a repeating disaccharide consisting of galactose and sulfated *N*-acetylglucosamine in alternation.

A major breakthrough came in 1969, when two graduate students at Rockefeller University, Vincent C. Hascall, Jr., and Stanley W. Sajdera, borrowed techniques that aid the study of nucleic acids such as DNA. By means of the techniques they extracted intact from cartilage the macromolecules containing chondroitin sulfate. The macromolecules are known as proteoglycans. They are well worth describing; proteoglycan turns out to be an elegant example of the relation between structure and function in molecular biology. Indeed, one can now say precisely how proteoglycan makes cartilage resilient.

Proteoglycan Structure

The central, organizing molecule of the proteoglycan in cartilage is the repeating disaccharide hyaluronic acid [*see illustration on pages 90–91*]. From this central strand extend strands of protein. They are among the largest proteins produced by any cell; each protein has a molecular weight of 250,000 to 350,000 units (that is, it has the weight of as many as 350,000 hydrogen atoms). The way each protein is anchored to the central strand of hyaluronic acid is remarkable. At one of its ends the chain of amino acids composing the protein folds itself into a ball: a globular region weighing 60,000 to 90,000 units. This region binds to the hyaluronic acid. The binding is not covalent, that is, no chemical bonds are made between the two. Instead the globular end of the protein makes an elaborate fit with a stretch of five disaccharide units along the length of the hyaluronic acid. In addition a



CLOSE VIEW OF PROTEOGLYCAN is offered by this electron micrograph, shown at an enlargement of 57,000 diameters. A central strand of the polymer hyaluronic acid runs through the field of view. It supports numerous projections, each consisting of a "core protein" to which bushy protrusions are linked. The structure of a core protein is shown schematically in the illustration on pages 90 and 91.

distinct protein known as link protein, weighing 40,000 to 60,000 units, interacts with both the hyaluronic acid and the protein to stabilize the interaction.

Each protein is itself the place of attachment for numerous polysaccharide chains; thus the proteins are termed the "core proteins" of proteoglycan. Three regions on each core protein are distinguished by the differences among their polysaccharide chains. The first of these regions includes the globular end where the core protein meets the central strand of hyaluronic acid. It has rather few polysaccharide chains, and the ones that are present tend to be small sugar chains called *N*-linked oligosaccharides. They are similar in structure to the ones found on many other proteins secreted by cells. Then comes a region rich in keratan sulfate polysaccharide chains and in a class of small sugar chains called O-linked oligosaccharides. The latter are probably unused initiation sites for keratan sulfate chains. Each chain is bound to a particular amino acid on the core protein, namely serine or threonine. The final region is rich in chondroitin sulfate chains. Each chain is bound to the core protein (in particular to the amino acid serine) by means of the linkage region of xylose, galactose, galactose and glucuronic acid.

Some 20 or 30 different enzymes are needed to prepare sugars and couple them to the core protein. In the synthesis of chondroitin sulfate, for instance, each sugar is connected to the next one by separate enzyme-controlled reactions. A first enzyme controls the binding of xylose to serine. A second exzyme couples xylose to galactose, a third couples galactose to galactose, a fourth couples glucuronic acid to galactose. That completes the linkage region. The next two sugars in the sequence begin the repeating disaccharide structure of chondroitin sulfate. First *N*-acetylgalactosamine is joined to the linkage region, then glucuronic acid is joined to *N*-acetylgalactosamine. The chain is assembled—one sugar at a time—by two types of enzymes working in alternation. The mechanism that terminates the synthesis of the chain is unknown. At some point sulfation enzymes substitute a sulfate group (SO_4) for one of the hydroxyl (OH) groups in each *N*-acetylgalactosamine. The sulfate group can be positioned at one of two positions (called the 4 or the 6 position) on the sugar. One enzyme works at the 4 position; a different enzyme works at the 6 position.

Overall a central strand of hyaluronic acid can have as many as 100 core proteins extending from it. In turn each core protein emits about 50 keratan sulfate chains and roughly 100 chondroitin sulfate chains. The total weight of a proteoglycan monomer (a core protein and the chains covalently bound to it) is 1.5 to 2.5 million units; the total weight of a proteoglycan aggregated on hyaluronic acid can amount to many tens of millions of units.

The Structuring of Water

The crucial point about the gigantic, elaborate structure of proteoglycan is that the sugars in the repeating disaccharide chains of keratan sulfate and chondroitin sulfate carry negative electric charges. In each disaccharide pair of chondroitin sulfate, for example, one of the sugars, glucuronic acid, has a COO^- chemical group. The other sugar, *N*-acetylgalactosamine, is sulfated: it includes an SO_4^{-2} group. In essence, therefore, proteoglycan must be considered an extremely large but highly ordered array of electronegativity.

Now, water is a small electric dipole: the distribution of electrons in the molecule tends to favor the oxygen atom at the center of H_2O, so that negative charge tends to be concentrated there, with positive charge at the hydrogen atoms. As a result the electrostatic forces of attraction and repulsion govern the interaction of water molecules and the interaction of water with other charged molecules. In particular, water becomes organized in multiple interacting layers or shells around a focus of electric charge; the number of layers depends on the strength of the charge. This means that proteoglycan structures large volumes of water—many times its own weight, in fact. Additional water is trapped in the interstices of the chondrocytes' extracellular matrix, much as water is trapped in the holes of a sponge. In both these ways proteoglycan organizes water so that water becomes the chief constituent of cartilage.

The resilience of cartilage results directly from this water-structuring property. When pressure is put on cartilage, water is forced away from the "charge domains" of the sulfate and carboxyl groups. The negative charges of those groups then come into close proximity, and the repulsive forces of the charges act against further compression. When the pressure is released, water returns to the charge domains. In fact, the nourishment of the cartilage at the joints of the body comes not in blood vessels—there are no blood vessels—but in the flow of liquid brought on by the compressions and relaxations arising from body movements. Little wonder that long periods of inactivity can weaken joint cartilage by making it thin and fragile.

The extracellular matrix of cartilage does not consist of proteoglycan alone but of proteoglycan and collagen. Indeed, in the body as a whole proteoglycan is less abundant than collagen. Although chondrocytes synthesize both,

BREAKTHROUGH: WEAVING LIGHTNESS AND STRENGTH INTO AIRPLANES.

The lighter an airplane is, the farther it can go or the more it can carry.

The problem: How to reduce weight while maintaining strength. Our solution: Use lightweight, high-strength carbon fibers.

We cut sheets of carbon cloth—thin filaments, woven together—to a precise shape. We build them up, layer by layer, to give them strength. Soft and pliant, these stacks of composite cloth are easily shaped to aerodynamic forms, then cured under pressure at high temperatures. The result: Wings and other parts that are lighter and more resistant to corrosion and have longer life than comparable metal parts.

Because of our use of carbon-epoxy composites for more than 25% of our AV-8B Harrier II structure, U.S. Marines have a plane that can land or take off vertically—and go twice as far or carry twice as much as earlier models.

We're making breakthroughs not only in aerospace but also in such fields as health care, information processing and lease financing.

We're McDonnell Douglas.

MCDONN
DOUGLA

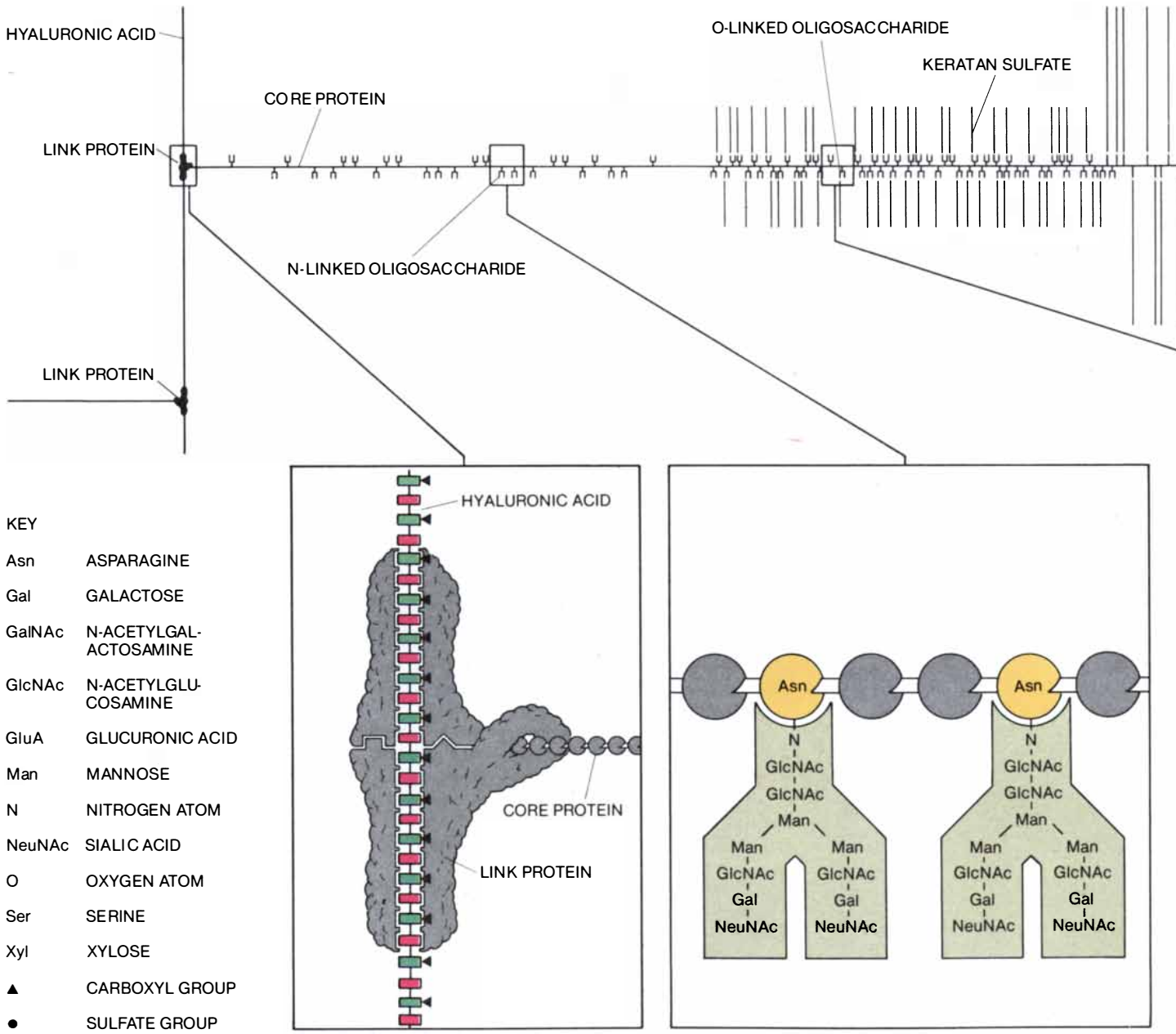
the syntheses of the two appear to be under separate control. Collagen begins, in any case, as procollagen, a protein consisting of about 1,100 amino acids, where every third amino acid is glycine. About a third of the intervening amino acids are proline or hydroxyproline, which are rare in other proteins. Proline and hydroxyproline are notable, however, for their compact and rigid structure, and these properties suit them well for the tight twisting that builds procollagen into collagen.

In particular, a procollagen strand twists into a compact and rigid left-handed helix; then, in the type of collagen characteristic of cartilage, three of these helixes twist into a right-handed superhelix. Outside the cell the ends of the individual procollagen molecules are clipped. The result is a rodlike macromolecule 1.4 nanometers (1.4×10^{-9} meter) in diameter and 300 nanometers long. Groups of these rodlike molecules are further organized into longer and thicker rods called fibrils, which range from 10 to 200 nanometers in diameter depending on the tissue they are in.

It is this last, large-scale organization of collagen into fibrils that gives cartilage tensile strength. The collagen fibrils in cartilage form a meshwork much like the steel girders in a building. Within this meshwork proteoglycan monomers and link proteins synthesized by chondrocytes assemble with hyaluronic acid into macromolecular aggregates surrounded by massive volumes of water.

The Aging of Cartilage

Until recently little had been discovered about the details of how cartilage changes through life and how these changes bear on aging and on illnesses such as joint disease. For these reasons Hascall and his colleagues at the National Institute of Dental Research had joined with my laboratory to make a detailed study of the structure of proteoglycans produced by chondrocytes at different stages of development. We employed cultured chick limb-bud cells because we could monitor the emer-



MOLECULAR STRUCTURE OF PROTEOGLYCAN produces the resilience of cartilage. In the drawing at the top of this illustration the central strand of hyaluronic acid is at the left. It is a repeating disaccharide, that is, it consists of alternations of two sugars. From the central strand project the long proteins known as the core proteins of proteoglycan. One such protein runs across the drawing at the top. It is joined to the hyaluronic acid by an elaborate fit and anchored by a "link protein." Three regions of the core protein are distinguished. The first includes oligosaccharides (short sugar chains) linked to the protein by a nitrogen atom. The second includes kera-

gence of newly expressed chondrocytes and watch them age, becoming "adult" chondrocytes and finally "old" or "senescent" ones. (I use quotation marks because the "aging" occurs in cell culture over a period of two to three weeks. It cannot be taken for granted that the sequence in a culture dish duplicates the sequence in intact living tissue.) On specific days of cell culture we exposed the cells to sulfate groups labeled with the radioactive isotope sulfur 35; the cells incorporate the groups exclusively into the proteoglycans they are producing.

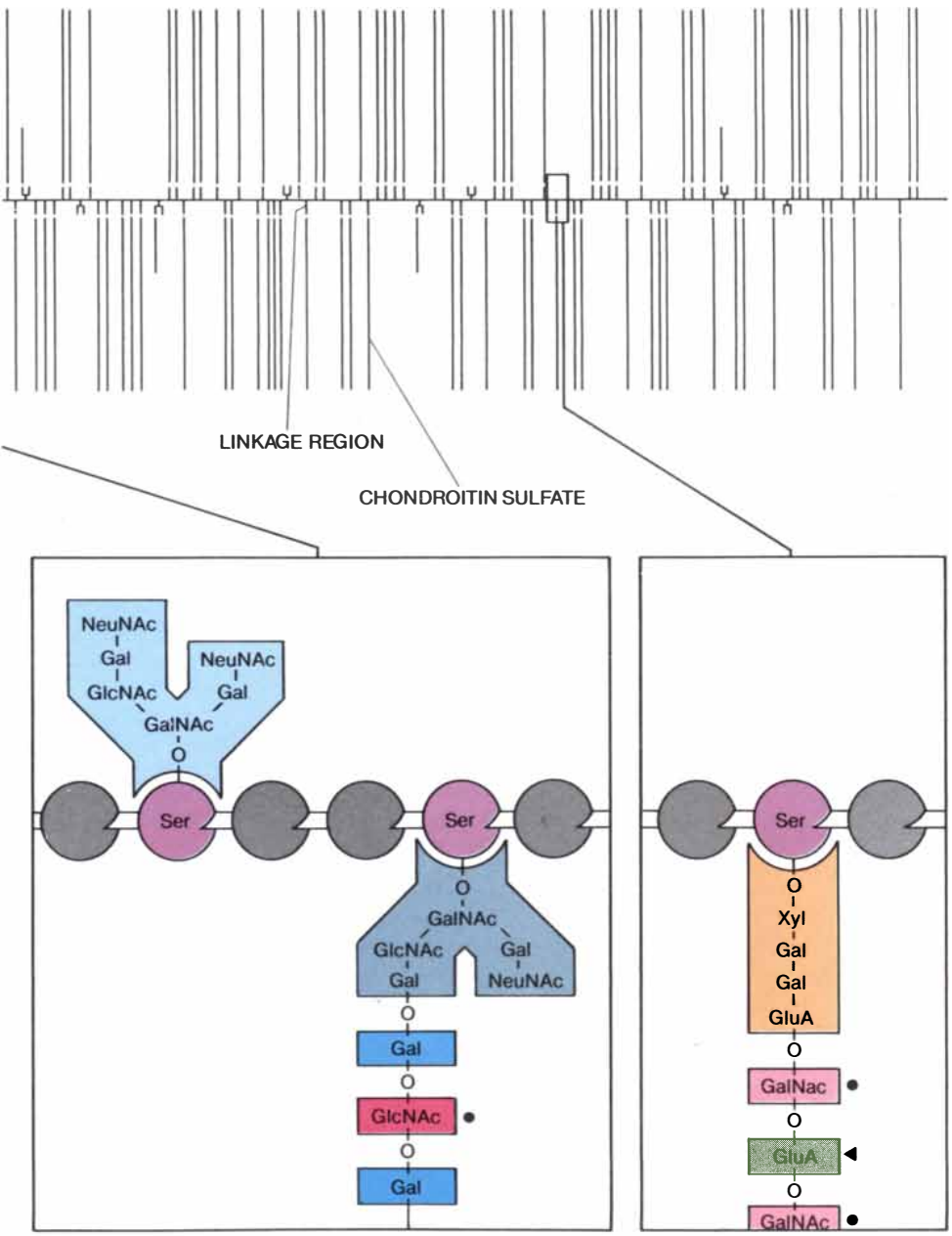
In this way we uncovered a temporal pattern of proteoglycan synthesis. The newly expressed chondrocytes synthesize a proteoglycan with distinctively long chondroitin sulfate chains: they are twice as long, on the average, as the ones older cells make. In addition the chondroitin sulfate chains have a relatively great proportion of chondroitin-6-sulfate. As chondrocytes age, the proportion of chondroitin-6-sulfate progressively lessens and the proportion of chondroitin-4-sulfate correspondingly increases. The keratan sulfate chains also show a difference. Young cells make them as very short polysaccharide polymers; the older cells synthesize much longer chains. In sum, both the length and the chemistry of the polysaccharide chains in proteoglycan vary with the developmental age of the cells over the two to three weeks in which the developmental program of the cells takes the cells from embryonic to "senescent." To put it another way, the cells synthesize proteoglycan with distinctly different chemistries rather than making a standard proteoglycan that is clipped or otherwise modified once it has taken up a position in the extracellular matrix.

In the limb bud of an embryo the transition from young chondrocytes to old ones culminates in the calcification of the cartilage matrix, the hypertrophy and death of the chondrocytes, the entrance of blood vessels into the tissue and the replacement of the cartilage tissue by bone cells and bone matrix built up on the remnant-scaffolding of the chondrocytes' extracellular matrix. This makes it reasonable to suspect that in anticipation of the replacement the hypertrophic chondrocytes synthesize proteoglycan distinctly different from the proteoglycan mature chondrocytes have produced. Remarkably, the replacement program differs from the one we found in chondrocytes progressing from "young" to "old" in a culture dish. Indeed, the hypertrophic chondrocyte makes a proteoglycan with long chondroitin sulfate chains sulfated predominantly in the 6 position, much as if it were a young chondrocyte again.

On the other hand, the progression from "young" to "old" proteoglycan in the cartilage made by cultured limb-bud cells has been verified in cartilage from young and old steers and from the young and the very old human body. For example, chondrocytes liberated from the joint cartilage of young and old steers synthesize proteoglycans that are distinctive in structure: the ones from the older animal have shorter chondroitin sulfate chains and longer keratan sulfate chains.

There is also a broad-scale difference: the proteoglycans from older animals are smaller overall, and structure less water around themselves, primarily because of the smaller size of the chondroitin sulfate chains. I suspect this difference may be responsible, at least in part, for the development of some forms of osteoarthritis in older people. After all, if the proteoglycans made in cartilage by aging chondrocytes have a lessened capacity to structure water, the resilience of the cartilage must be compromised, and with it the cushioning of bones at the joints. And if the pressure on the cartilage exceeds the capacity of the tissue to respond, the tissue must fragment at sites of high stress. Such fragmentation may be sensed as injury by the body's immune system, which then responds by generating inflammation at the site. The result would be swelling and some of the painful symptoms of osteoarthritis.

Could osteoarthritis be reversed or kept from progressing by inducing young chondrocytes to populate the car-

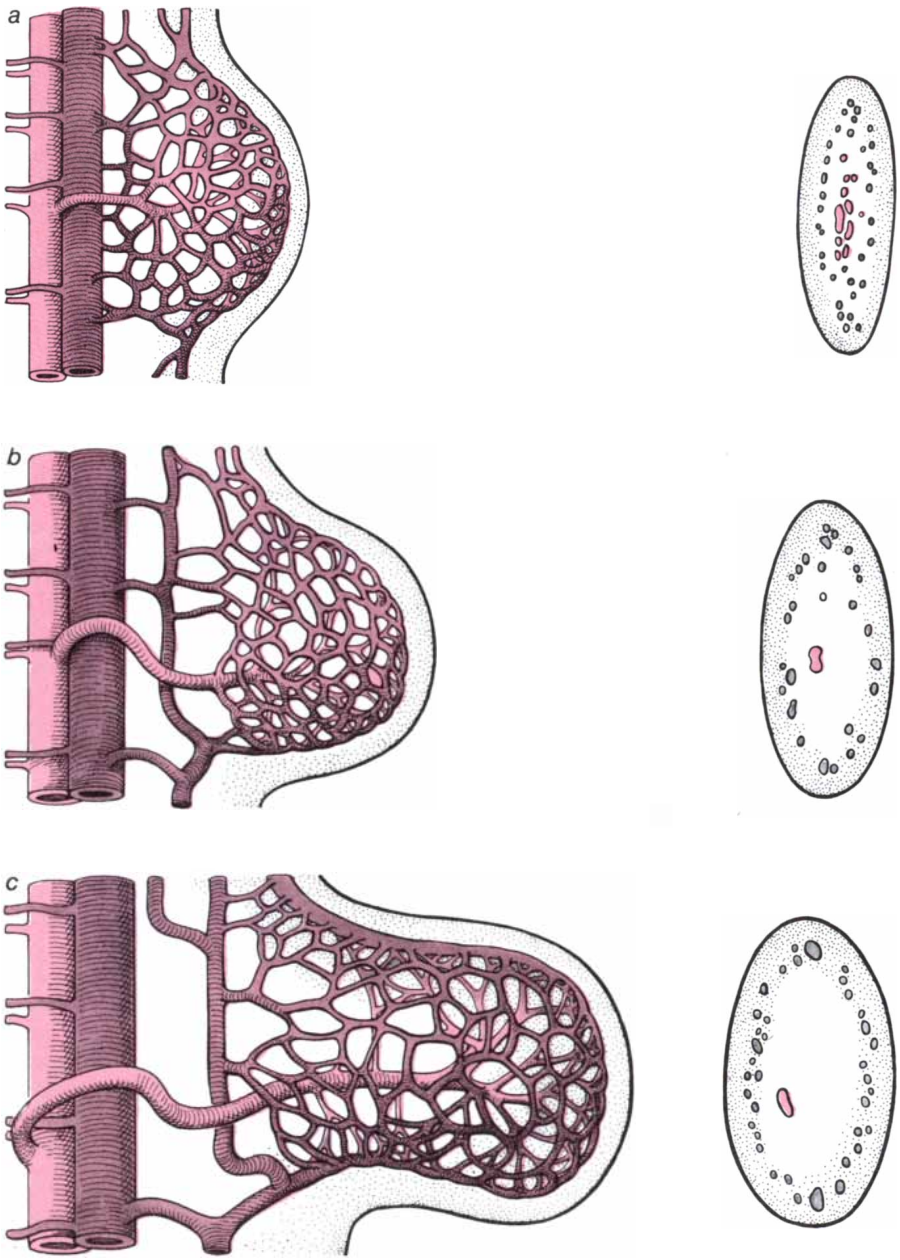


tan sulfate chains. They too are repeating disaccharides. It also includes oligosaccharides linked to the protein by an oxygen atom. The third includes chondroitin sulfate chains. They again are repeating disaccharides. Proteoglycan is studded with negative electric charge, which orders the surrounding water. When pressure is applied, water is driven out of the interstices of the tissue; when pressure is removed, the water returns, restoring the shape of the tissue.

tilage at joints and synthesize proteoglycans with a youthful water-structuring capacity? The hypothesis is attractive, but unfortunately it relies on an oversimplification of the complexity of osteoarthritis. It also denies the complexity of the biological mechanisms governing cartilage. It should be said, for example, that the progression of chondrocytes from "youth" to "senescence" in a culture dish is not in itself a clue to the life cycle of chondrocytes intact in the body. Most notably, it still is not known how many generations of chondrocytes populate cartilage throughout the life of an animal.

Even so, one can argue that the changing pattern of synthesis of proteoglycans on the part of chondrocytes must be governed by at least three distinct biological clocks. By "clock" I mean either the rate at which a biosynthetic "program" of change is played out or the rate at which an entire tissue changes. The controlling mechanisms are probably different. For example, the cellular mechanism governing the synthesis of proteoglycan molecules by chondrocytes is probably different from the mechanism governing how long a chondrocyte exists.

The first clock, then, is a fast one; it controls the development of cartilage in the embryo, thus producing tissue that serves as a scaffold for replacement by bone. The second clock is a slow one; it enables cartilage at the growth plates of the long bones of the body to make those bones grow longer. The growth of the human body ends with the teen-age years; hence the slow clock runs down during the second decade of human life. The third clock is a very slow one; it functions at joint-cushioning cartilage and plays out a changing biosynthetic program over periods that range (in the human body) up to 60 years or more.



LIMB BUD of the chick embryo suggests one of the principles controlling the emergence of cartilage in the embryo. Initially the limb is nourished by a uniform spray of capillaries (*a*). Later one of the central vessels enlarges to form the limb's major artery (*b*). In this way two regions emerge: an avascular core and a vascularized periphery (*c*). Cartilage develops in the core. The development of cartilage has an inverse relation to the development of blood supply.

Two Experiments

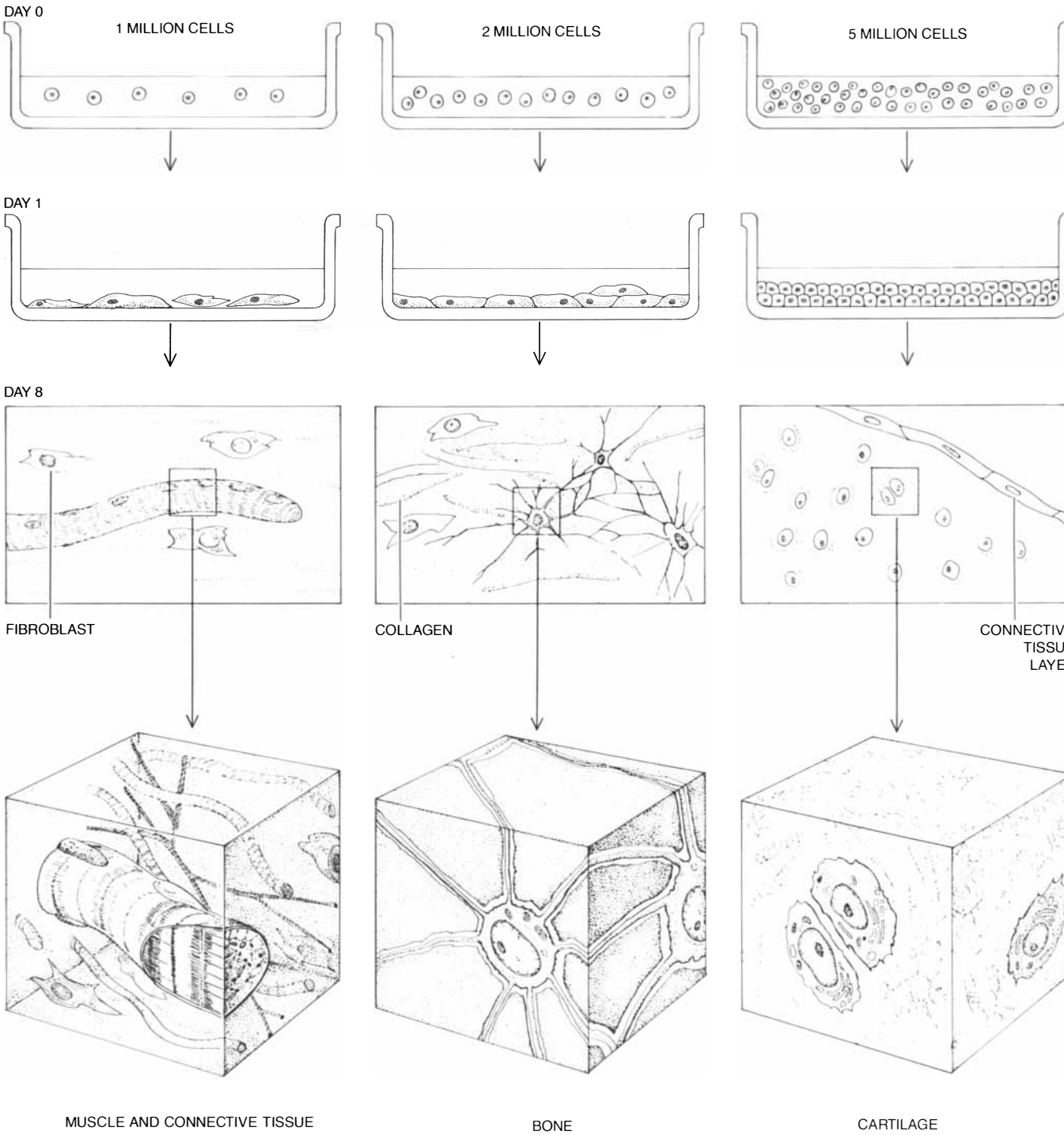
What controls the clocks? In an effort to turn the clocks back—to restart the developmental program governing the expressional pattern of cartilage—we have done two types of experiment. In the first type we explored the suggestion that the extracellular matrix is somehow responsible for the expressional pattern of chondrocytes. Accordingly we freed mature chondrocytes from the extracellular matrix by sampling cartilage tissue from eight-day-old high-density cultures of chick-limb mesodermal cells and treated the samples with enzymes that break up proteins in the matrix. The freed cells were then reseeded—naked, as it were—into new culture dishes. By exposing the naked cells to sulfate groups labeled with sulfur 35 we were able to examine the proteoglycans the naked cells made. We found that the naked chondrocytes synthesize the same proteoglycans they had synthesized before their liberation. We also found that the naked chondrocytes continue to play out their biosynthetic program. The absence of an extracellular matrix does not cause the clock to restart.

In the second type of experiment we attempted to adjust the clocks by making the extracellular matrix abnormal. Here we treated the cultured cells with beta-umbelliferyl xyloside, a xylose substitute that enters chondrocytes and competes for the enzyme joining galactose to xylose. (The xylose is already bound to the core protein of proteoglycan.) In this way the treatment interferes with the making of the four-sugar linkage region connecting chondroitin sulfate to the core protein. When the xylose substitute is present in chondrocytes, chondroitin sulfate chains are joined to the substitute instead of becoming part of proteoglycan; thus the chondroitin sulfate (bound to the xylose substitute) can be recovered from the medium bathing the chondrocytes.

The result is impressive. The cells in high-density limb-bud cell cultures exposed continuously to the xylose substitute become committed to differentiation into chondrocytes, but the proteoglycans they make are missing most of their chondroitin sulfate and so are smaller than normal proteoglycans. Ac-

cordingly the cells come to sit on one another instead of being separated from one another by a normal extracellular matrix. We removed the xylose substitute from the medium at a particular time, say on the eighth day of cell culture, and exposed the cells to radioactive sulfate groups. The newly synthesized proteoglycans (the ones marked with radioactivity) had the same structure they would have had if they had been made by normal eight-day-old cultured limb-bud chondrocytes.

The cells had never synthesized a normal proteoglycan and had never resided in a normal extracellular matrix. Indeed, their metabolism had been perturbed by an abnormal agent. Even so, their clocks had kept accurate developmental time and the programmed changes in the pattern of proteoglycan synthesis had occurred. Evidently the chondrocytes proceed through irreversible changes in their expressional programs, so that it is normally impossible to restart the program or initiate the program at any point other than the beginning. The effects of thalidomide and other teratogens (agents producing defects in embryos) seem to bear that out. Thalidomide causes abnormal morphologies, but the abnormalities are connected to normal structures. For



CELLS IN CULTURE differentiate into different types of tissue depending on how densely the cells are "seeded" into culture dishes. At low seeding density (one million cells in a dish 35 millimeters in diameter) the cells flatten themselves on the bottom of the dish (*left*). For the most part they do not touch one another. After a few days muscle fibers develop in a matrix of connective tissue. At intermediate seeding density (two million cells) the cells pave the bottom of the dish in a cobblestone pattern (*middle*). Bone cells develop, depositing minerals around themselves. At high density (five million cells) the cells form a layered pile (*right*) and differentiate into chondrocytes.

example, an arm may be malformed whereas the hand and the shoulder are normal. Evidently the programmed development of the body was forming the arm at the time the teratogen had access to the body's intracellular biosynthetic machinery. The teratogen interfered with arm development, but the expressional program continued to play itself out, so that as the teratogen was eliminated a normal hand could form.

The proteoglycans produced at a given time become part of a cartilage tissue whose own structure is time-dependent. This means that the structure of cartilage is complex in two time-dependent ways. It also subverts the hope that osteoarthritis might be arrested simply by repopulating a knee joint, for example, with young chondrocytes. To be sure, the young chondrocytes might generate an extracellular matrix of the young, resilient type. This matrix, however, would have to be integrated into a tissue that had changed greatly over the long span of time since young cartilage occupied the site.

The study of mutant mice provides a case in point. In a type of mutant mouse called brachymorphic, a defective sulfation enzyme causes a lack of complete sulfation of the chondroitin sulfate in embryonic proteoglycans. The consequence is a small fetus with misshapen limbs. Suppose the normal sulfation enzyme could be restored to the newborn mouse. It would sulfate newly synthesized proteoglycans. Unfortunately the new proteoglycan molecules would be required to integrate themselves into an extracellular matrix more suited to accommodate mutant proteoglycans. To repair the mouse's malformations a way would presumably have to be found in which the defective enzyme could slowly yield to its normal counterpart.

Cartilage Factors

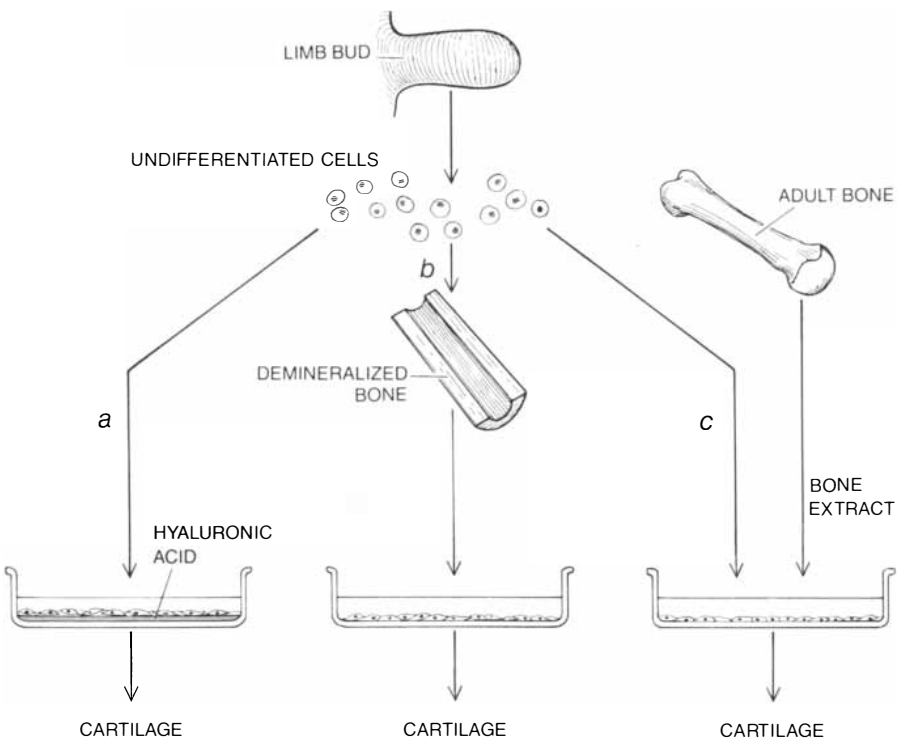
The process by which cartilage gives way to bone, allowing calcification and vascularization in the developing body, suggests a mechanism by which the body develops: it suggests that cartilage and bone include substances that profoundly affect other tissues. The idea could ultimately have major implications for the management of disease. It could also shed light on the developmental cycle of body tissues.

Take the striking circumstance that normal, healthy cartilage is avascular. In the chick embryo, cells that will express themselves as chondrocytes can be identified in the core of the developing limb because the core excludes blood vessels. In the fully formed organism the situation is much the same. Working at Rush Medical College and the Harvard Medical School, Klaus E. Kuettner, Rubin Eisenstein, Nino Sorgente and M. Judah Folkman have shown that cartilage contains a substance that specifically inhibits the incursion of blood vessels. Only when the expressional program of cartilage is played out and chondrocytes are hypertrophied and dying do blood vessels ever enter the tissue, as a preliminary to bone formation.

The identity of the substance inhibiting blood-vessel ingrowth has not yet been established, but efforts are under way. They consist of making extracts of cartilage, then testing each fraction to find ones that inhibit vascularization. The substance is well worth finding, if only because the growth of a tumor requires a rich supply of blood. Perhaps a factor derived from cartilage, applied to a growing tumor, could throttle its blood supply and thus bring on its death. In any event, the search for the substance exemplifies the activity at a number of universities and corporations. Programs at these institutions are directed at isolating and purifying from adult cartilage several substances that may turn out to affect many aspects of growth, development and disease.

One pattern emerges clearly from the study of cartilage: the development of cartilage in the embryo and the aging of cartilage in the body seem to depend on the same types of programmed changes—changes that are manifested in the extracellular matrix chondrocytes produce around themselves. Certainly the development and the aging of cartilage reflect changes in the synthesis of proteoglycans, which ultimately are controlled by the organism's genetic machinery. In my laboratory proteoglycans are now being found in embryonic muscle and even in the embryonic brain. We suspect these proteoglycans may turn out to have the general function of reserving spaces by filling them with structured water; these spaces may be reserved as sites where organs and tissues can develop their proper form, just as cartilage reserves space in which the long bones of the body take shape. The study of the development of cartilage has led, therefore, to a model that may yield insight into mechanisms controlling development in general.

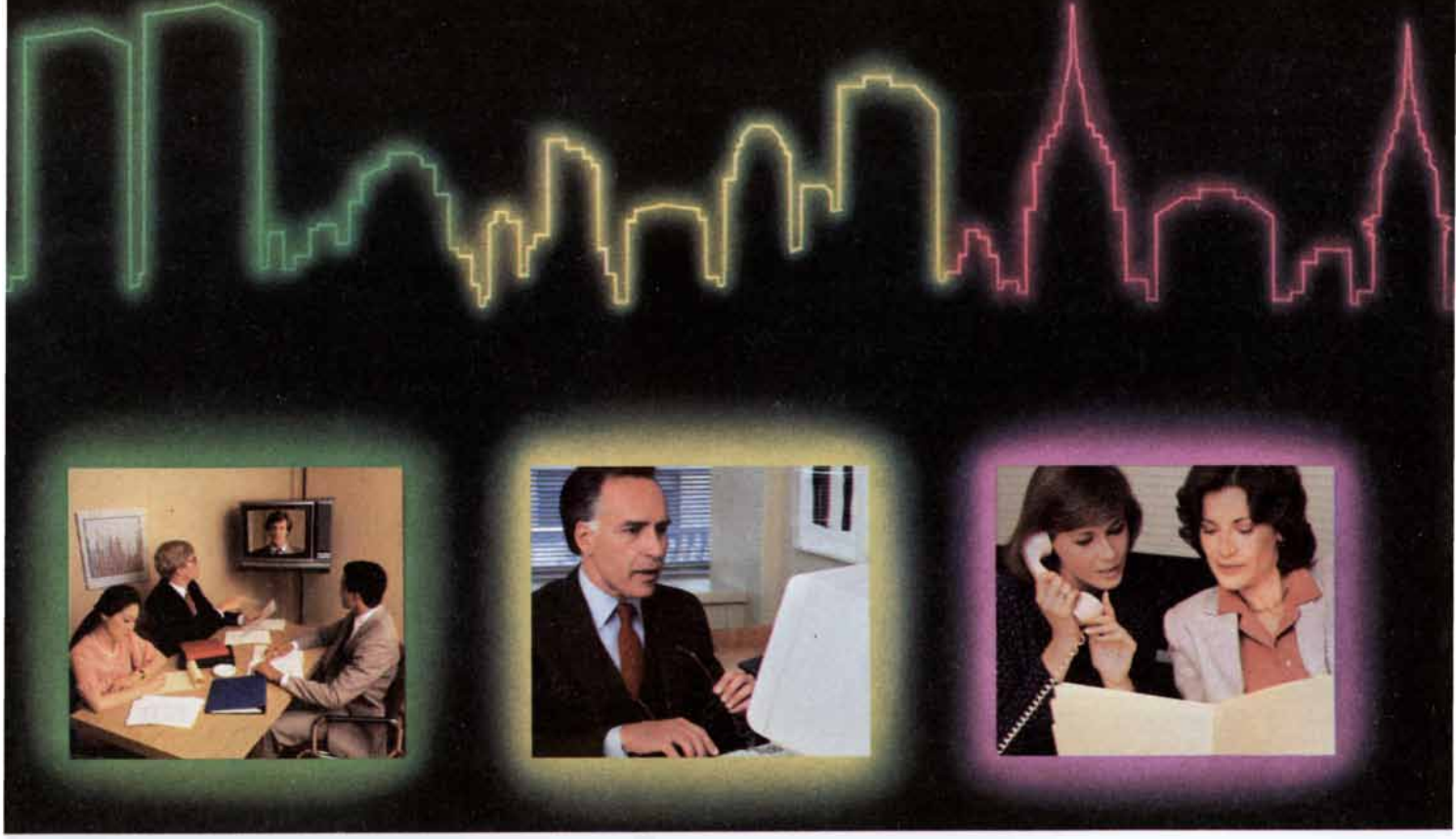
Aging too is being studied. Perhaps people who age without being victimized by problems with their joints do so because something in their genetic makeup beneficially affects the structure of their cartilage or slows the clock that governs its change. The elucidation of the structure and function of cartilage on the molecular level may thus yield the means by which the downward arc of the body's developmental program can someday be changed for the better.



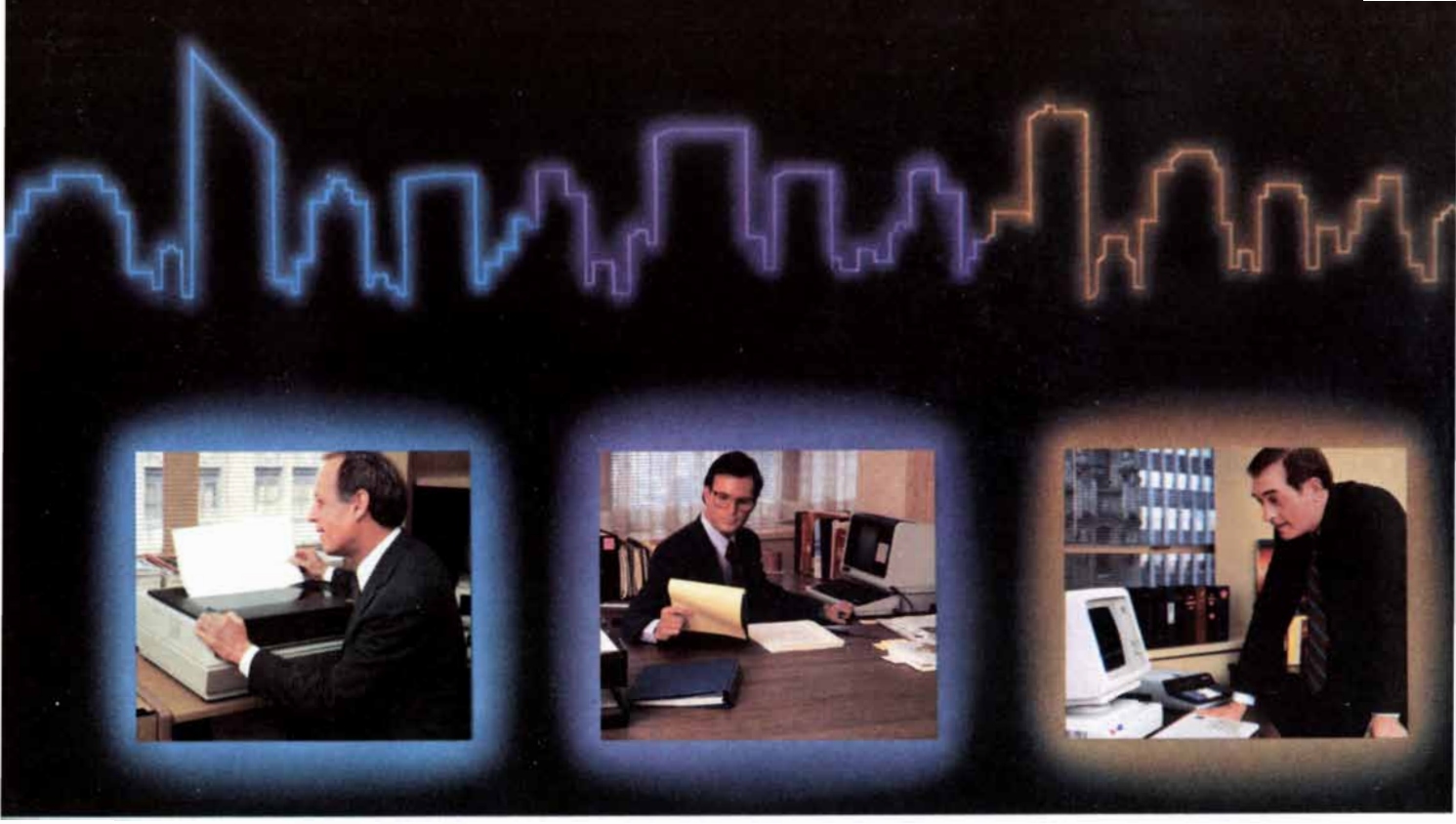
CHEMICAL INFLUENCES on the differentiation of cells into chondrocytes was studied in several experiments. For each experiment limb-bud cells were seeded into culture dishes at the intermediate seeding density. In one experiment (*a*) the dish had been "carpeted" with hyaluronic acid; the cells on this carpet became chondrocytes. In a second experiment (*b*) calcium and phosphates were leached out of adult bone, and limb-bud cells were seeded in the remaining bone matrix before they were placed in the dishes. The cells became chondrocytes. In a third experiment (*c*) the cultured cells were treated with extracts of bone. Again the cells became chondrocytes. Limb-bud cells seeded into culture dishes at the intermediate seeding density do not normally differentiate into chondrocytes. For its part, the bone extract can stimulate two distinct cellular activities: cell replication and cell differentiation into chondrocytes.

Now that you're ready for a change of pace it's time to try John Jameson.

Take a sip of John Jameson. Note the light, delicate taste. Luxurious and smooth as you would expect a premium whiskey to be. But with a distinctive character all its own. Set a new pace for yourself. Step ahead of the crowd with John Jameson, the world's largest selling Irish Whiskey.



Presenting a telecommunications network from New York Telephone that's so advanced it can handle all your business needs, no matter how complex.



The New York Telephone network. A high-tech leader in a high-tech age.

Nowhere are more big businesses attuned to putting information to work than in New York—the communications capital of the world.

And no one in New York is more tuned in to the information and communications needs of businesses than New York Telephone.

Today, New York Telephone has the experience and technology to transmit your company's voice and data traffic better, faster and more cost effectively than ever before. Behind that claim is our network of copper and lightwave cables, digital radio, high-speed switching centers and computers—all woven together to work with fantastic speed, efficiency and precision. Our **remarkable network** provides communications and data access for more than six and a half million business and residence customers, and links them with the worldwide long distance services.

The groundwork for tomorrow, today.

A big part of serving those customers is our fiber optics technology—a technology that we've been working with for years. Fiber optics are hair-thin strands that use laser light sources to transmit voice and data at amazing speeds. In fact, we have 32,000 fiber miles of cables in place today. They can transmit at speeds of up to 405 million bits of information per second and are serving many of America's largest financial and industrial companies.

We are anticipating the future.

And fiber optics aren't the only tools in our high-tech workbench.

New York Telephone is currently working with a technology called packet switching. This technology allows the transmission of many calls from many users simultaneously over the same line.

Another technology we're working with is Local Area Networks, which will allow your company to tie together all of its information devices—including mainframe and personal computers for accessing data- and text-processing resources—so that they can "talk" to each other, even if they're at different locations.

As the communications needs of New York and businesses such as yours grow in scope and sophistication, New York Telephone will have the solutions. We have the people, the ingenuity, the imagination and the technology. As we have for nearly a century, we will continue to provide you with high-quality communications service in the future.

New York lives on New York Telephone.

New York Telephone

A **NYNEX** Company

Epsilon Aurigae

Infrared and ultraviolet observations made during the latest eclipse of this binary star system suggest the eclipsing object is a hot, young star surrounded by a large cloud of dust and gas that gave it birth

by Margherita Hack

Late this month in the middle latitudes of the Northern Hemisphere the constellation Auriga can be seen on the eastern horizon, rising just to the north of Orion shortly after sunset. With the aid of a simple star chart one can quickly find the third-magnitude star labeled epsilon in Auriga, a thumb's width to the southwest of Capella, the brightest star in that region of the sky. In the fall of 1982 a careful observer might have noticed that Epsilon Aurigae was slowly beginning to darken. By December of the same year it had diminished to only half its initial brightness, and it remained a fourth-magnitude star for nearly 11 months. Then gradually it began to brighten until by the middle of last May it had regained its former magnitude.

This cycle of darkening and brightening has been repeated four times in the 20th century, at precise intervals of 27.1 years. The basic mechanism of the cycle was first explained by the German astronomer Hans Ludendorff in 1904. It is the result of an eclipse of the primary, or bright, visible star, by an unseen companion object; the visible star and its companion form a binary system held together by gravity. The length of the cycle is the longest orbital period known among eclipsing binaries, but it is the dark phase of the cycle that is hardest to explain. Its duration implies the eclipsing object must be enormous: the width of the object needed to account for nearly two years of eclipse is about 1,500 times the radius of the sun. Although each eclipse has been studied by a new generation of astronomers with new and more precise observational techniques and with increasingly sophisticated theory, the identity of the eclipsing object remains the great mystery of Epsilon Aurigae.

The eclipse of 1982–84 was the first one in which the star could be observed from satellite platforms, and the results have extended observational knowledge of Epsilon Aurigae into the far ultraviolet. At the other end of the electromagnetic spectrum, observations have been made in the deep infrared by instruments such as the three-meter Infrared Telescope Facility of the National Aeronautics and Space Administration at the summit of Mauna Kea on the island of Hawaii. The data gathered during the most recent eclipse are still being studied and will be the topic of a workshop to be held in Tucson, Ariz., next January. They confirm earlier evidence that Epsilon Aurigae is a unique object in the Milky Way, and they have already made it possible to eliminate several earlier models that were proposed to account for the observations. Nevertheless, by building on and combining some of the earlier models of the star one can now give a plausible and rather detailed interpretation of the spectroscopic and photometric code carried to us from a star 1,900 light-years away.

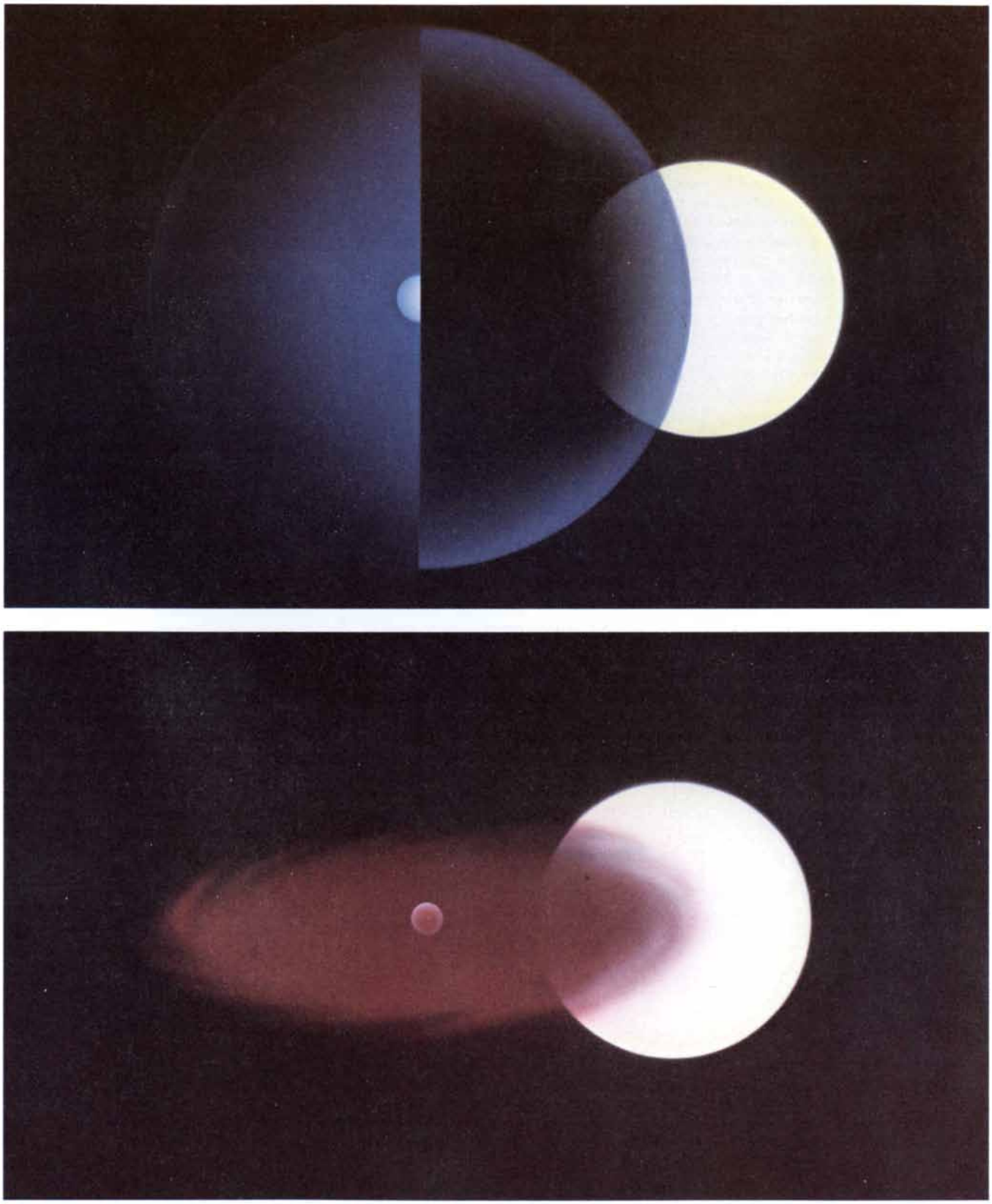
The visual observation of an eclipsing binary system is not unusual by itself. At least half of the stars in the Milky Way are found in gravitationally bound binary systems, and many of these systems are oriented in such a way that from our vantage in space the stars periodically eclipse each other. All eclipsing binary systems are too distant to be resolved into their stellar components, and so from our vantage such a system appears to be a single star. Nevertheless, an eclipse of the system can cause the apparent star to change its brightness periodically in much the same way the headlight of an oncoming car flickers briefly in the night when its path is crossed by an otherwise invisible pedestrian.

The two-year-long dimming of the light of Epsilon Aurigae was the first evidence that the star is unusual, but the spectroscopic data suggest the true dimensions of the problem. By measuring the Doppler shift in the spectral lines of the primary star one can determine the component of its velocity along the line of sight between the star and the earth. That component enables one to estimate the orbit of the primary star about its center of gravity and to give upper and lower limits for the mass of the companion. The orbital characteristics of the system suggest the mass of the companion is between four and 15 times the mass of the sun.

Given such a mass and the enormous size implied by the length of the eclipse, it is surprising the companion is not visible even with the largest telescopes. During the secondary eclipse of the companion star by the primary star it should be possible to detect a companion star up to three magnitudes fainter than the primary star, or a factor of about 16 in brightness. That sensitivity, however, can be attained only if the orbital plane of the system is nearly parallel to the line of sight of the observer; if the inclination of the system is such that the eclipsing primary star only grazes the disk of the companion, a brighter companion star could go undetected. Nevertheless, in observations with ground-based instruments made before and after eclipse, the spectrum of the companion star is completely overwhelmed by the spectrum of the primary star. The companion must therefore be at least two magnitudes darker than the primary star.

One might suppose that during the eclipse of the primary star by the companion the spectrum of the companion could be discerned. What is seen instead is one of the most puzzling aspects of the Epsilon Aurigae system and one of its most telling features. The lines in the spectrum of the system during the eclipse are almost identical with the lines seen before and after the eclipse. The observation might suggest both objects in the system belong to the same spectral class and have the same surface temperature, but that explanation is ruled out by the failure to observe the companion before and after the eclipse.

The alternative explanation is simply that the spectrum of the primary star must always be visible. For example, the disk of the primary might not be completely occluded by the companion, or it might be eclipsed by a cloud of dust or gas that allows some of the light of the



EARLY MODELS of the eclipsing binary star system Epsilon Aurigae sought to describe an object that could make an eclipse last for almost two years without ever completely cutting off the light of the eclipsed primary star. Moreover, the model object had to account for the approximately constant dimming of the light over all visible wavelengths. One of the most successful early models suggested the eclipse is caused by a shell of ionized gas and free electrons that surrounds a hot companion star, which is observable only in the ultraviolet (*upper illustration*). Here a quarter sector of the shell has been cut away to reveal half of the disk of the companion. Another early model suggested the eclipsing body is a disk made up of grains of dust whose average diameter is significantly larger than the wavelength of infrared radiation (*lower illustration*). Such a disk is similar to the disk of dust from which the solar system was evolved, and a protostar is probably embedded in the disk. Both electrons and large dust grains can scatter the light of the primary star and thereby dim it in a way that is independent of wavelength. Recent observations in the ultraviolet and infrared regions of the spectrum indicate that neither model can entirely account for the data; a new model that combines the two models, however, now seems to be quite successful.

primary star to pass through. Whatever model is adopted, it must also be reconciled with the observation that the intensity of the light emitted by the primary star is reduced almost uniformly during the eclipse, regardless of the wavelength at which the event is observed.

What can account for the unique set of observations? Ludendorff suggested that the companion is not a star at all but instead is a swarm of meteoroids in orbit about the primary star. That suggestion was revived in the 1960's in somewhat revised form, but at the time of the initial suggestion the huge mass that would have been assigned to the meteoroidal cloud would have been unacceptable to most astronomers.

In 1937 G. P. Kuiper, Otto Struve and Bengt Strömgren of the Yerkes Observatory proposed a model based on observations of the eclipse of 1928–30. According to their model, the companion is an enormous but extremely tenuous mass of gas called an *I* star, which is partially transparent to light. The bulk of gas needed to account for both the probable mass of the companion and the duration of the eclipse must be so rarefied that a mechanism had to be introduced to increase the opacity of the eclipsing *I* star.

The mechanism is analogous to the one responsible for the formation of the earth's ionosphere. The atoms in the outer layers of the star were to be ionized, or stripped of one or more of their electrons, by ultraviolet radiation emitted by the primary star. Free electrons absorb light and scatter it in a way that is effectively independent of wavelength. Hence the emissions of the primary as seen from our solar system could be uniformly attenuated by passing through the ionization layer of the *I* star. The temperature of the *I* star was assumed to be about 700 degrees Celsius, which is far too cool for detectable emissions of visible radiation and so accounts for the failure to observe the companion star.

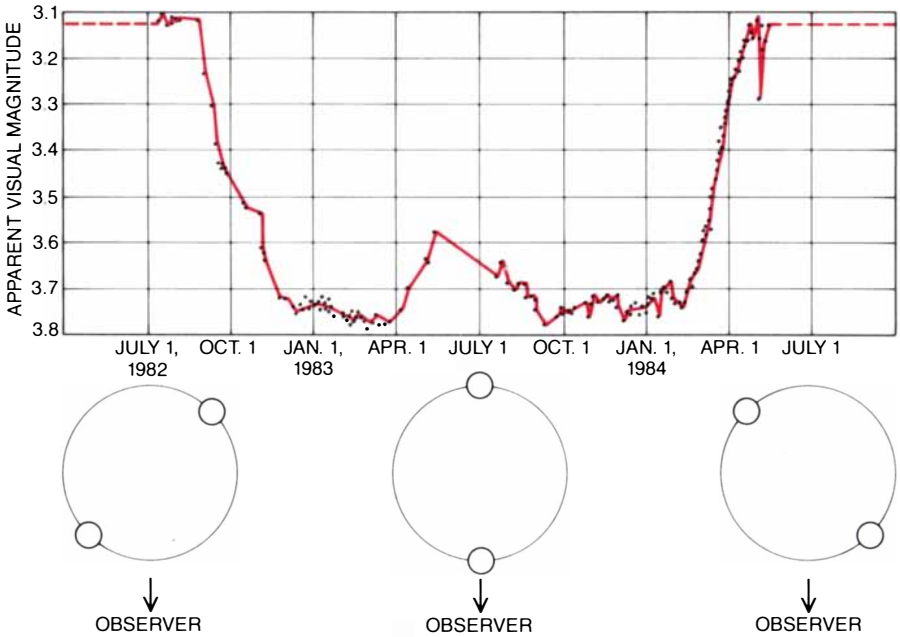
One of the main criticisms of the 1937 model was that even the free electrons would not account for the reduction in the light output during the eclipse. In 1954 Zdeněk Kopal of the University of Manchester proposed that instead the eclipsing body is a ring of solid particles or dust grains surrounding the invisible companion star. The ring is assumed to be thin enough so that when viewed roughly edge on it obscures only half of the visible hemisphere of the primary star. Kopal has since suggested the disk closely resembles a protoplanetary nebula, a rotating disk of gas and dust from which a solar system can be formed. In this model too the spectrum of the primary star is always observable, and the eclipsing body, which only absorbs the light of the primary, does not contribute appreciably to the light of the system.

Almost all astronomers agree with the basic geometric picture described by the last two models. In other words, the eclipsing body is either a semitransparent shell absorbing half of the light of the primary or it is a flat disk that covers only half of the hemisphere of the primary. There is almost no agreement, however, on the answers to two major questions that appear capable of being resolved: What is the composition of the eclipsing matter, and what is the nature of the invisible companion?

Only two kinds of material are known that can absorb light independently of wavelength. I have already mentioned that a gas of free electrons can absorb and scatter all wavelengths equally. The second candidate material is a cloud of dust grains. If the diameter of each grain is substantially larger than the wavelength of infrared and visible radiation, a cloud of dust would absorb radiation of all detectable wavelengths. According to several models of Epsilon Aurigae, which collectively have both shell and disk geometry, dust is the main cause of the eclipse.

A close study of spectrums made both in and out of eclipse, however, tends to contradict the dust-grain hypothesis. The spectrums give a subtle but clear way to distinguish the light scattered by the eclipsing body from the light that passes unimpeded from the primary star to the earth. The spectrum of the primary star is made up of a series of dark absorption lines set against a background of emitted light that forms a continuum of spectral colors. The absorption lines are caused by the excitation of atoms in the photosphere, or outer and visible layer, of the star. According to the laws of quantum mechanics, the atoms can be excited only to certain discrete energy levels, and so the energy required to excite them must take on discrete values corresponding to the allowed excitation levels. Photons that carry the discrete values of energy are thereby absorbed and the rest of the energy is transmitted; the result is an absorption spectrum characteristic of the atoms that make up the photosphere.

If some of the light emitted by the star were to pass through a shell or a disk made up of dust, the light would be scattered in all directions. The spectrum of scattered light would be identical with the spectrum of the primary star observed outside the eclipse, except for a small correction. The stars in a binary system generally rotate about their axes in the same clockwise or counterclockwise sense as they revolve about their common center of mass. If the eclipsing



LIGHT CURVE is a graph in which the intensity (apparent visual magnitude) of the light emitted by Epsilon Aurigae is plotted against time. Such a curve was the first kind of evidence that led to the realization the star is an eclipsing binary system. Below the curve the eclipse is depicted schematically: part of the light from the more distant of the two stars in the system is periodically obscured by an object along the line of sight between the distant star and the earth. The light curve shows that the total light output of Epsilon Aurigae decreases during eclipse by a magnitude of .8, or about half its brightness out of eclipse. (Stars of the first magnitude are among the brightest stars, and stars of the second, third and higher magnitudes are progressively dimmer.) The partial phases of the eclipse, both ingress and egress, last for 192 days each, and totality, or the maximum depth of the eclipse, lasts for 330 days. The data were plotted from measurements made during the eclipse of 1982–84 by observers at the Hopkins Phoenix Observatory in Phoenix, Ariz., and the Tjornisland Astronomical Observatory in Sweden.

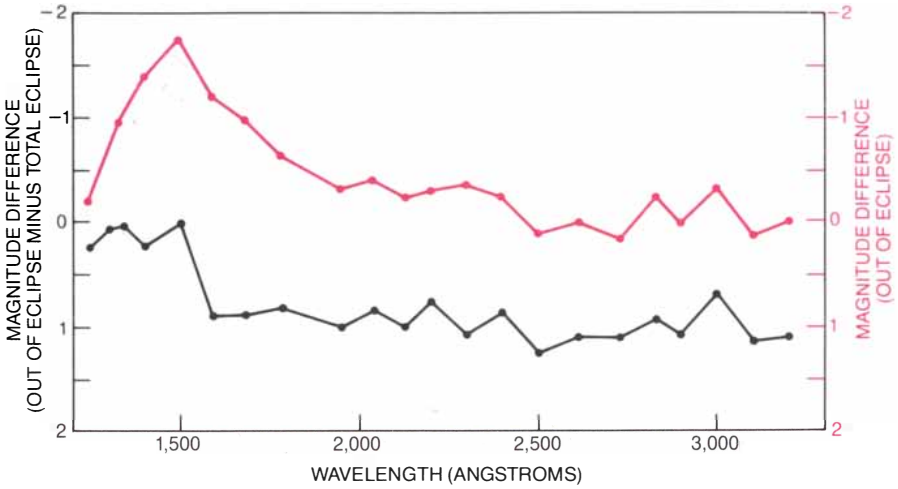
body is a shell or a disk, it too must rotate, and during the ingress phase of the eclipse, just before totality, the part of the shell or disk that eclipses the primary rotates away from the observer [*see illustration on next page*]. The scattered absorption lines are thereby shifted toward the red end of the spectrum by the Doppler effect.

During the egress phase of the eclipse, just after totality, the effect is reversed. The part of the shell or disk that eclipses the primary rotates toward the observer, and the scattered light is Doppler-shifted toward the violet. During totality the shift is zero because it is the central part of the shell or disk that eclipses the primary; the radial component of its rotational velocity is zero. The observed spectrum is the superposition of two components, namely the spectrum of scattered light and the spectrum of light that reaches the earth directly from the primary star without being scattered. Thus just before and just after totality the absorption lines of the primary star would be broadened slightly in the direction of the Doppler shift and the difference between the light intensity in the darkest part of the line and the light intensity of the surrounding continuum would diminish.

If the light from the primary star passes instead through a gas of free electrons, the observed spectrum is slightly more complex. Not only is the light scattered by the electrons but also it is transformed by its passage through the gas of ionized atoms from which the free electrons are derived. The gas and the photosphere of the primary star have approximately the same atomic composition, and the gas too absorbs photons of light whose wavelength corresponds to some excited atomic-energy level. The absorption lines in the resulting spectrum are again added by superposition to the lines in the spectrum of any light that reaches the earth directly from the primary star.

Because of the additional absorption by the gas, however, the absorption lines are enhanced during totality, or darkened with respect to the background continuum. During ingress and egress the absorption lines from the shell or disk of gas are Doppler-shifted, just as they are when the light passes through a cloud of dust. The enhancement of a line combined with the Doppler shift often splits the line into two sharp points. The double-pointed lines in the absorption spectrum provide an unmistakable signature of the shell or disk that scatters the light; the shifted component of the lines is called the shell spectrum.

Doppler-shifted lines in the spectrum of Epsilon Aurigae were noted by Ludendorff in 1901 and became widely known after the eclipse of 1928–30. The lines show a clear enhancement attributable to the shell spectrum, and many of them have two sharp points during ingress and egress. Consequently there seemed little doubt that the eclipse of the primary star is caused by gas rather than by dust.

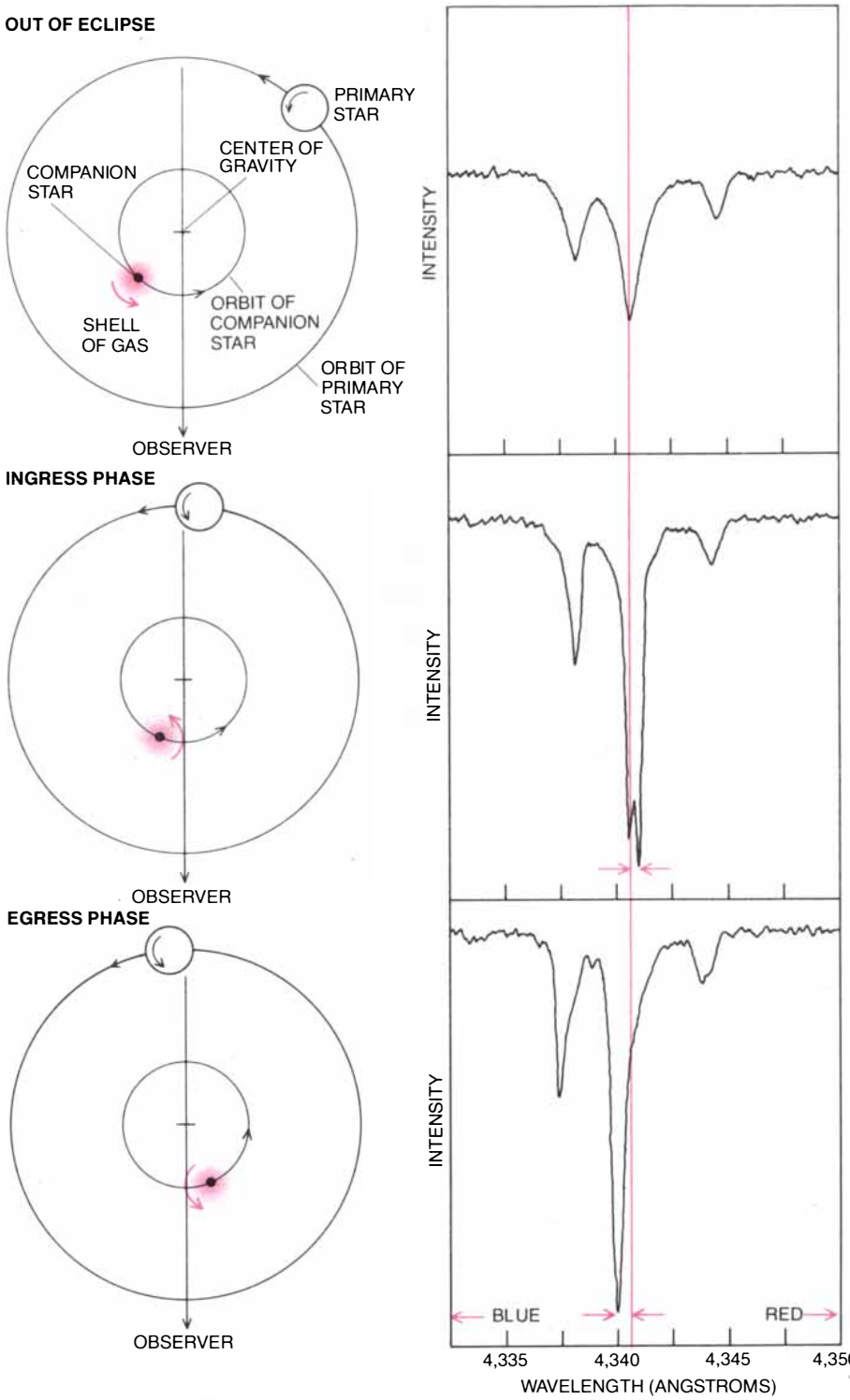


DIFFERENCE BETWEEN THE MAGNITUDE of Epsilon Aurigae before eclipse and its magnitude during eclipse is plotted in black. In the range of near-ultraviolet wavelengths between 1,600 and 3,000 angstrom units the difference oscillates between .8 and 1 magnitude, or a factor of about 2 to 2.5 in brightness. In the far ultraviolet, however, at wavelengths shorter than 1,600 angstroms, the difference is less than .2 magnitude, or a factor of about 1.2 in brightness. At such short wavelengths the radiation from the eclipsed primary star is insignificant and so the effect of the eclipse is small. What is seen instead is the light of the hot companion star, which is in the foreground during the eclipse and emits radiation primarily in the far ultraviolet. Also plotted against wavelength is the magnitude difference between observations made at two different times before the eclipse (*color*). The brightness of the system is roughly constant at wavelengths longer than about 1,700 angstroms. At shorter wavelengths, however, the magnitude varies with time, which indicates the intrinsic variability of the companion star.

The record of the eclipse of 1928–30 confirmed this conclusion in a surprising way. Struve noted that certain absorption lines, which are formed only when atoms are excited to a high energy level, are not enhanced during totality and are not split in two during the partial phases of the eclipse. For example, during eclipse there is no detectable shell component to the blue line with a wavelength of 4,481 angstrom units that is associated with singly ionized magnesium. (One angstrom unit is 10^{-10} meter.) If the eclipsing body were made up entirely of dust, the blue magnesium line would be Doppler-shifted by the shell or disk, and the change in the shape of the line would be observable in the spectrum. On the other hand, if a gas of atoms and free electrons eclipses the primary star, the blue magnesium line is easy to understand. The temperature and the density of the shell or disk of gas are presumably much lower than the temperature and the density of the stellar photosphere where the line is initially generated. Thus the number of photons and particles in the shell or disk energetic enough to cause atomic excitation at high energies is too small to bring about a measurable enhancement in the corresponding absorption lines.

In 1954 Robert P. Kraft, then a young associate of Struve's at the University of California at Berkeley, found a difficulty with the electron-scattering hypothesis. Struve had assumed the primary star was the source of the energy needed to ionize the atoms in the eclipsing shell or disk of gas and thereby create the free electrons. The energy output of the primary star, however, is known by direct observation. Its surface temperature is about 7,500 degrees C. and its radius is about 100 times the radius of the sun. Kraft calculated that the electron density in the shell or disk that could be generated by such a star would be about 100 million electrons per cubic centimeter. That density is too small, by a factor of about 1,000, to explain the depth of the eclipse. In other words, if the opacity of the shell or disk is attributed to scattering by electrons generated by the primary star, the shell or disk is still too transparent to explain the observations.

Kraft suggested the opacity of the gas is caused instead by the presence of negative hydrogen. Negative hydrogen is an ordinary hydrogen atom that has captured an extra electron. It can form from the excess electrons generated in a gas hot enough to ionize atoms such as iron, manganese and chromium but not hot enough to ionize neutral hydrogen. In the limited spectral range observable from the ground the absorption of radiation by negative hydrogen is almost independent of wavelength; consequently when Kraft introduced his idea it seemed quite promising. In the ultravio-



"SHELL SPECTRUM" of Epsilon Aurigae is most clearly manifest during the ingress and egress phases of the eclipse. At the top of the illustration the blue region of the spectrum is plotted near one of the so-called Balmer lines of hydrogen at 4,340 angstroms; the wavelength corresponds to the energy absorbed by hydrogen atoms in the photosphere of the primary star. The line is a spectral component of light that travels unimpeded from the primary star to the earth. During the ingress and egress phases the line deepens and splits into two sharply defined regions of minimum intensity; the deepening and splitting can be attributed to a new component of absorption superposed on the absorption line in the original spectrum. The additional component is introduced when some of the light from the primary star is absorbed by hydrogen atoms in a cloud of gas. During the ingress phase the rotation of the cloud is such that the gas along the line of sight between the primary star and the earth is moving away from the observer, and so the spectral absorption line is shifted toward the red by the Doppler effect (*middle*). Similarly, during the egress phase the absorption line is shifted toward the blue by the Doppler effect because the eclipsing part of the cloud is moving toward the observer (*bottom*). The spectrums were made by the author and her collaborators with the coude spectrograph of the 152-centimeter telescope of the Observatoire de Haute Provence in St.-Michel, France.

let, however, the absorption of radiation by negative hydrogen becomes much smaller, and so it would be too transparent to account for more recent observations of Epsilon Aurigae made at ultraviolet wavelengths high above the atmosphere.

I began my own observations of Epsilon Aurigae during the eclipse of 1955–57, and in 1957 I had the opportunity to study the spectrums made by Struve during that eclipse. The resolution of the new spectrums was much higher than it was for the spectrums made during the previous eclipse. I was thus able to demonstrate the presence of certain weak shell lines such as neutral, excited magnesium and calcium that had formerly seemed to be missing. By measuring the intensity of the newly detected lines and by counting the number of lines in the hydrogen spectrum I was able to make a direct estimate of the electron density in the shell or disk of gas, without regard for the source of ionizing energy. The density turned out to be about 100 billion electrons per cubic centimeter, or 1,000 times the density calculated by Kraft. Given a shell or disk with a thickness of about .7 astronomical unit, such a density would be just sufficient to explain the observed depth of the eclipse by electron scattering alone. (An astronomical unit is the mean distance between the earth and the sun, or about 93 million miles.)

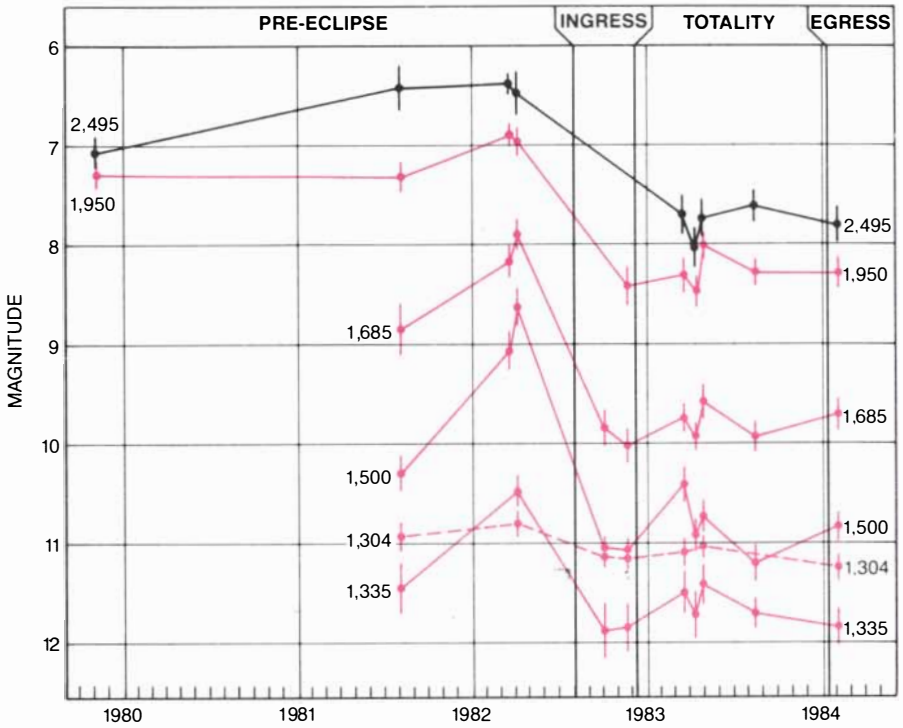
Since Kraft had shown the primary star is not hot enough to produce the necessary electron density, I still faced a major problem: What is the source of energy that ionizes the shell or disk? In 1961 I suggested the ionization is caused by the invisible companion star, which is surrounded by a thick shell of gas. The shell of gas is dragged along by the companion as it moves in its orbit, and the eclipse is observed when the shell is imposed between the earth and the primary star. I was able to calculate a range of possible temperatures and radii for the invisible companion from the characteristics of the shell spectrum. The calculations showed the companion could be a giant star with a surface temperature of 15,000 degrees C. and a radius 60 times that of the sun, or it could be a hot subdwarf with a temperature of about 100,000 degrees and a radius about equal to that of the sun.

At the time I made my proposal the detection of a hot companion star seemed unlikely. It is well known that hot stars radiate most of their energy in the ultraviolet region of the spectrum, and the radiation is almost completely absorbed by the earth's atmosphere. Confirmation of a hot star in Epsilon Aurigae had to await observations at ultraviolet wavelengths between 1,000

and 3,000 angstroms made from astronomical satellites.

The first attempts to observe such a star were made with a telescope built at Princeton University and mounted on the NASA *Copernicus* satellite, and with a spectrometer designed by astronomers from Belgium and Great Britain and mounted on the *TD-1* satellite of the European Space Agency. The Princeton telescope, built to obtain high-resolution spectrums of hot stars brighter than the fifth magnitude, gave no positive evidence for the hot star. The Belgian and British spectrometer, built to obtain low-resolution spectrums of hot stars brighter than the seventh magnitude, gave a few uncertain results. There was some evidence for ultraviolet radiation from Epsilon Aurigae in excess of that expected from the primary star alone.

In January, 1978, the International Ultraviolet Explorer satellite (*IUE*) was launched, and after a few months of testing in orbit the satellite was made available to guest observers. The *IUE,* a joint venture of NASA and the European Space Agency, was designed to give high-resolution spectrums of stars as faint as the 10th magnitude, as well as low-resolution spectrums to the 13th magnitude. On April 19, 1978, my colleague Pier Luigi Selvelli of the *IUE* European observing station near Madrid and I observed the low-resolution spectrum of Epsilon Aurigae on the screen of a video terminal. There for the first time was the spectral signature of the binary system in the far ultraviolet, at wavelengths as short as 1,300 angstroms. The presence of the previously undetected companion star was unambiguously revealed.



ULTRAVIOLET LIGHT CURVES give the temporal variation in the intensity of radiation from Epsilon Aurigae at selected wavelengths. At 2,495 angstroms in the near-ultraviolet region of the spectrum (*solid lines in black*) the intensity is roughly constant before the eclipse and falls smoothly through the ingress phase to a minimum during totality, much like the light curves in the visible spectrum. At almost all wavelengths shorter than about 1,950 angstroms, however, the intensity of the light is greater at certain times during the totality of the eclipse than it is during the ingress phase (*solid lines in color*). Most of the light at such short wavelengths can thereby be attributed to the companion star in the system, which is in the foreground and so is not undergoing eclipse. The variability in the light curves before the eclipse is caused by the intrinsic variability of the companion star. The intensity of the emission line of neutral oxygen at 1,304 angstroms (*broken lines in color*) is approximately constant both in and out of eclipse, indicating it is formed in an envelope of gas that surrounds the entire binary system. The data were obtained by Conrad Boehm, Steno Ferluga and the author, all of the Trieste Astronomical Observatory, from the International Ultraviolet Explorer satellite.

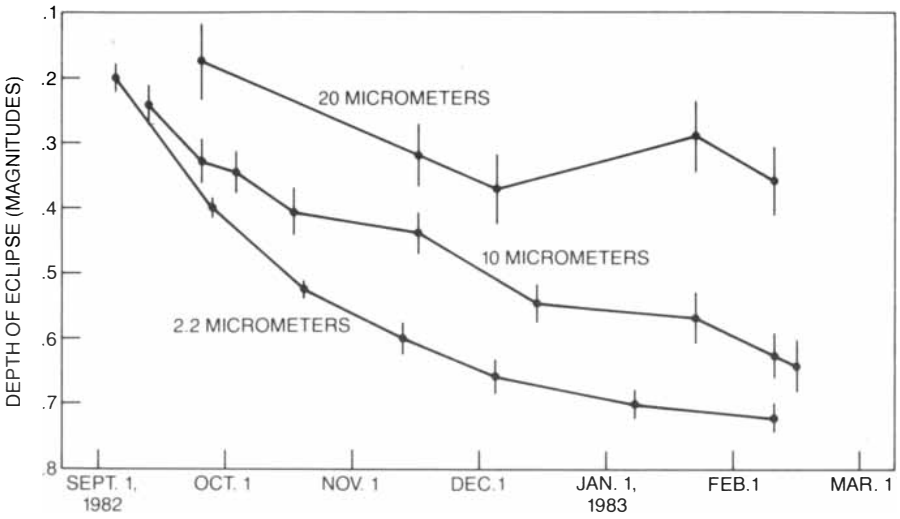
We have since made several additional observations out of eclipse, and two of my colleagues at the Trieste Astronomical Observatory, Conrad Boehm and Steno Ferluga, and I have also had the opportunity to observe Epsilon Aurigae in the ultraviolet during the most recent eclipse. The observations confirm the existence of a hot companion whose light is dominant at wavelengths shorter than 1,550 angstroms. More precisely, the dimming of the light from the system during eclipse is between .8 and 1 magnitude for all visible and ultraviolet wavelengths as short as 1,600 angstroms, whereas almost no trace of the eclipse is observable in the far ultraviolet, between 1,240 and 1,550 angstroms. Because the flux of the relatively cool primary star is negligible in the far ultraviolet, what we see in that spectral range is the flux of the companion star, which is in the foreground during eclipse.

Although the new observations have been dramatic, they have not given detailed confirmation of the model I proposed in 1961. The ultraviolet spectrum of the companion star shows it is not as hot as I had predicted it to be. Its surface temperature is about 10,000 degrees C. and its radius is between three and five times the radius of the sun.

Moreover, the infrared spectrum appears to be in complete contradiction with the data in the ultraviolet. Observations in the infrared made by Dana E. Backman, Eric E. Becklin, Dale P. Cruikshank, Theodore Simon and Alan T. Tokunaga of the University of Hawaii at Manoa and by Richard R. Joyce of the Kitt Peak National Observatory have monitored Epsilon Aurigae before and during the eclipse at wavelengths as long as 20 micrometers. The investigators have found the eclipse dims the light by about .7 magnitude at wavelengths between one micrometer and 4.8 micrometers, which is roughly equal to the depth of the eclipse for visible light. At longer wavelengths, however, the depth of the eclipse becomes smaller, and at a wavelength of 20 micrometers it is only .3 magnitude, down by a factor of about 1.3 in brightness. From these observations Backman and his colleagues concluded the eclipsing body is a cool object whose light is dominant in the far infrared. They derive a surface temperature of about 200 degrees C. and a radius of 10 astronomical units.

The infrared observations also pose a more subtle problem for my 1961 model. In a completely ionized shell of gas there is a process called free-free absorption that leads to strong emissions of radiation at infrared wavelengths. When a free electron in the gas passes near a proton or a positive ion, the electron readily gives up some of its energy in the form of an infrared photon; the greater the number of electrons and the more energy that has been absorbed by the gas from an external source, the more likely the process. The observed flux of infrared radiation could be interpreted as partial confirmation of the presence of free-free absorption in a gas surrounding the companion star. The flux, however, is much smaller than it would be if it were caused by free-free absorption in a shell as thick as the one in my early model.

How can one construct a model that



INFRARED LIGHT CURVES for emissions of Epsilon Aurigae in the near infrared, such as the one at a wavelength of 2.2 micrometers, exhibit roughly the same reduction in magnitude during the eclipse as the visible light curves do. At longer wavelengths, however, the effect is much smaller; at 20 micrometers, for example, the intensity of the emission is reduced by only about .3 magnitude, or a factor of 1.3 in brightness. The data initially seemed to contradict the interpretation of the ultraviolet light curves: they suggested the eclipsing object in Epsilon Aurigae is relatively cool instead of being hot. The infrared observations were made by Dana E. Backman, Eric E. Becklin, Dale P. Cruikshank, Theodore Simon and Alan T. Tokunaga of the University of Hawaii at Manoa and by Richard R. Joyce of the Kitt Peak National Observatory.

explains such varied and apparently conflicting observations? It is relatively straightforward to adjust the temperature of the companion star in my early model to match the observed temperature. Remember that the high temperatures in my model were necessary to account for the energy needed to ionize the shell. In the early 1960's stellar radiation was thought to be the predominant and perhaps the only ionizing agent. Since that time, however, observations in ultraviolet regions in the spectrums of several stars have repeatedly shown the presence of lines from multiply ionized elements, which cannot be generated solely by the radiation in the stellar photosphere. The phenomenon is called superionization, and it strongly suggests that other effects, of mechanical or magnetic origin, can contribute to the ionization of the gas. If superionization is the cause of the ionization of the eclipsing shell in Epsilon Aurigae, the relatively low, 10,000-degree temperature of the companion star could still give rise to the opacity of the ionized shell.

The observations in the infrared, however, suggest the need for more substantial changes in the model. The infrared radiation generated in my model by free-free absorption can be reduced to match the observed infrared flux if the thickness postulated for the shell of gas is reduced as well. In that case, however, the shell of gas alone cannot account for the observed opacity of the eclipse. The most promising solution to the dilemma appears to be a model that combines aspects of the two most popular models of the 1950's: the one that postulates a disk of dust and my model, which postulates the shell of gas.

In the new model the main cause of the eclipse is a ring of large dust particles, which absorbs half of the light of the primary star. The hot companion star is embedded in the ring and appears less luminous than it is because the ring absorbs part of its light. The ring is heated slightly by the companion star and emits some of the observed infrared flux as thermal radiation. To account for the shell spectrum, as well as for the fact that the appearance of the shell spectrum slightly precedes the absorption of light at the beginning of the eclipse, the model postulates a shell made up of gas that envelops the ring of dust. The gaseous shell is ionized by the radiation of the hot star emitted in a direction perpendicular to the plane of the ring, and it is formed of matter that could be either escaping from the hot star or transferred to it by the primary star. The shell of gas is not as thick in the new model as it is in my earlier model, but it extends to a distance of about 10 astronomical units from the hot star.

At least two additional observational features of Epsilon Aurigae can be incorporated in the new model. An extended shell of ionized gas should give rise to an emission spectrum out of eclipse as well as the observed absorption spectrum during the eclipse. The emission lines are not observed in the visible spectrum, probably because they are fainter than the continuum emission of the primary star. A few emission lines, however, are observable in the ultraviolet, namely neutral oxygen at a wavelength of 1,304 angstroms and singly ionized magnesium at wavelengths of 2,795 and 2,802 angstroms. The intensity of the lines is roughly constant both during and out of eclipse. Such lines must be formed either in a part of the eclipsing shell of gas that is not intercepting the light of the primary star or in an extended shell of gas that surrounds the entire binary system.

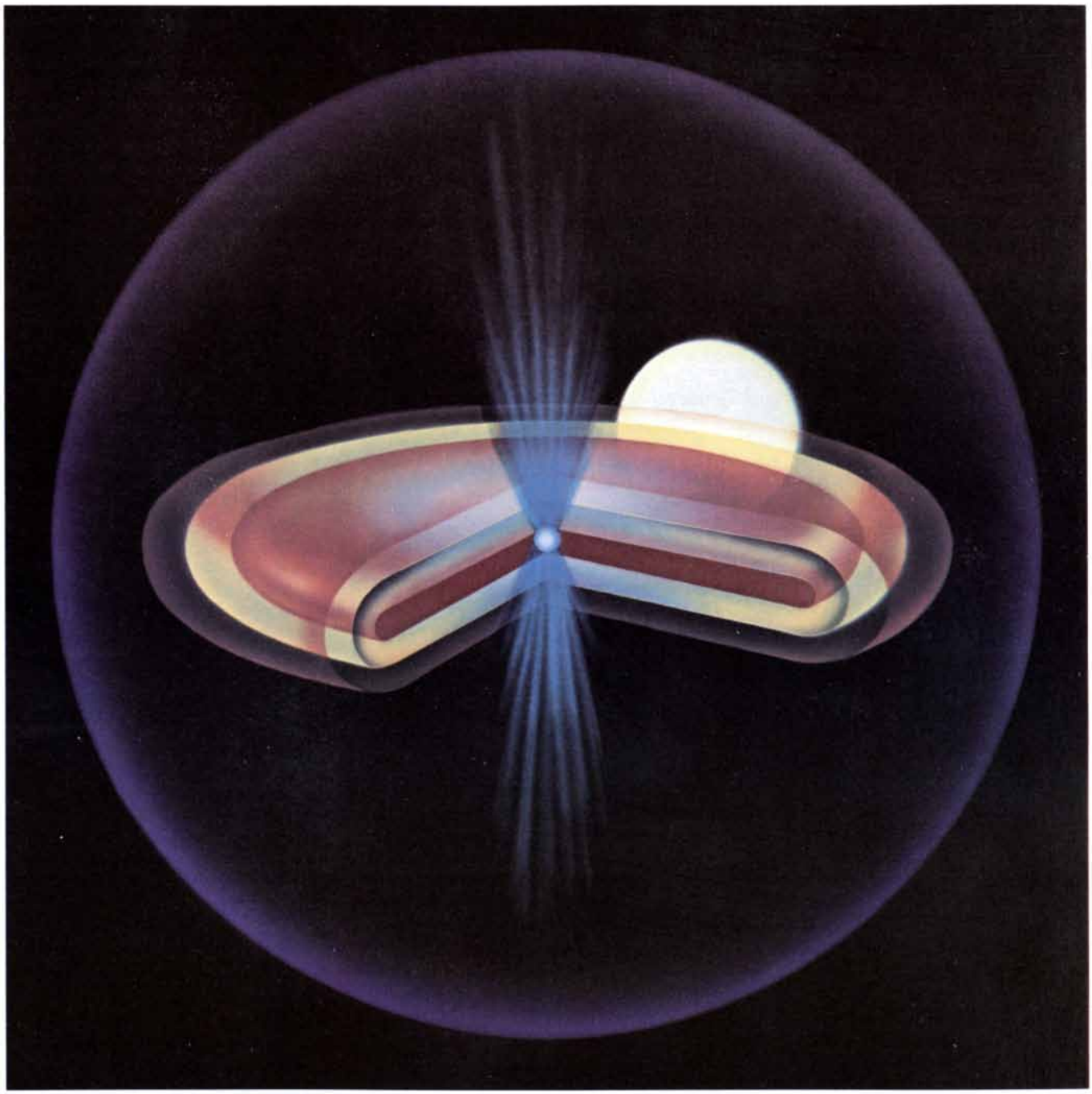
The second feature may be a clue to the evolutionary stage of the Epsilon Aurigae system. Observations in the ultraviolet show the companion is a variable star; in the far ultraviolet the variability is between magnitude 1 and 1.5, or in other words a factor of between 2.5 and 4 in brightness. The variability is similar to a rare group of stars in the same spectral category as the companion called Herbig variables, after George H. Herbig of the University of California at Santa Cruz. The Herbig variables are always found in young stellar clusters that show large and irregular variations in their emitted light. The stars themselves are thought to be still contracting from a cloud of gas and dust. Indeed, many Herbig variables are surrounded by clouds of dust whose dimensions and temperature are similar to the object postulated by Backman and his colleagues to account for the infrared observations of Epsilon Aurigae.

If the companion star in Epsilon Aurigae is a young star, one can readily account for the fact that no similar binary system has been observed. The closest relatives known are other long-period eclipsing binaries such as 31 Cygni, 32 Cygni, Zeta Aurigae and VV Cephei. In all four systems a relatively hot star is gravitationally bound to a cool giant or supergiant star. There is no evidence, however, that the hot star in any of these systems is surrounded by such an extended shell of dust as the one observed in Epsilon Aurigae. The difference, I propose, is that Epsilon Aurigae is a similar system in a different stage of its life history.

The primary star in Epsilon Aurigae is a supergiant, but it is hotter than the supergiants in the other four binary systems I have mentioned. Its surface temperature and spectral class show it is undergoing rapid evolution, and it may have recently—that is, in the past million years—completed a phase of abundant loss of mass. The companion star could be a very young star that has not yet reached its stable configuration and is still embedded in the remnant of the dust cloud from which it was formed. The grains of dust in the vicinity of the hot star have a lifetime of less than 10,000 years; if they are not replenished from the outside, the cloud will disappear in this short period of time. The short life expectancy of the cloud and the brief phase of stellar evolution now being traversed by the primary star ex-

plain why no other system with the characteristics of Epsilon Aurigae is known: the system offers a glimpse of an exceedingly fleeting stage in the evolution of a binary system.

Verification of these speculations on the nature of the companion star and the evolutionary stage of the system must await a more precise measurement of the masses of the two component stars. The measurement should become possible after 1986, when the space telescope is scheduled for launch. The resolution of its spectrograph will be far greater than that of the instrument carried by the *IUE* satellite. The ultraviolet spectrum of the companion star should exhibit a small Doppler shift, and so a high-resolution spectrum will make it possible to measure its radial velocity as seen from the solar system. With a few hours of observation a year, continuing over a period of 14 years, the radial velocity of the star over half of its orbital period can be determined. From those data the masses of both stars in the system can be derived. Hence by the year 2000 the remaining mystery of Epsilon Aurigae may be completely solved.



MODEL OF EPSILON AURIGAE that tentatively synthesizes the results of infrared, visible and ultraviolet observations gives a possible view of the eclipse. The surface of the supergiant primary star is partly occulted by a ring of dust and several flattened shells of gas that surround a blue-dwarf companion star. Here a slice has been removed from the eclipsing object to show a cross section. The innermost ring of dust is responsible for most of the reduction in the emissions of the primary star during the eclipse. Infrared emission from the ring of dust is shown in reddish brown. Surrounding shells of gas of decreasing density appear as opalescent halos, scattering the yellowish light of the primary and the blue light of the companion star. Because there is less dust around the poles of the companion star, the intensity of its light is strongest there and it is able to excite and ionize the gas. The parts of the shells of gas most distant from the companion star show the reddish glow of light emitted by the deexcitation of hydrogen. The spherical, outermost envelope of gas gives rise to a few emission lines in the ultraviolet region of the spectrum; it is shown as a translucent purple shell that surrounds the entire binary system.

A Late Ice-Age Settlement in Southern Chile

A settled community flourished in the forests of southern Chile some 13,000 years ago. The well-preserved site shows that New World Pleistocene culture was much more advanced than has been thought

by Tom D. Dillehay

The discovery of a late Pleistocene settlement at Monte Verde in the forests of southern Chile is yielding a strikingly detailed picture of how the earliest inhabitants of the New World lived. The remarkable wealth of artifacts these people left forms a rich and eloquent record of the social system, the economic strategies and the technologies through which they adapted to their postglacial forest habitat.

Radiocarbon analyses of wood, bone and charcoal from the site have yielded a series of dates between 13,000 and 12,500 B.P. (before the present). In addition to stone and bone tools, which are often found at early New World sites, Monte Verde offers a rich, unprecedented collection of wood artifacts and plant remains. Among the wood artifacts are the foundations of the earliest architecture that has yet been found in the Americas. The arrangement of these buildings suggests a community with a fairly well-defined division of labor. The organic remains show that the residents had a rich and varied diet based mainly on plants. The relatively high level of social development represented by the community at Monte Verde indicates that New World culture in the late Pleistocene was much more complex than has been thought.

The high degree of preservation of wood and vegetal materials, the social and economic patterns to which they and other materials attest, and the discovery of the site itself are all closely related to the geography and climate of southern Chile. The Monte Verde site lies in the central valley that dominates the landscape of southern Chile. The valley runs on a north-south axis and is bounded by a narrow Pacific littoral zone and a low coastal mountain range to the west and the high Andean cordillera to the east.

During the Pleistocene epoch, which ended between 10,000 and 8000 B.P., glaciers moved down the slopes of the Andes into the central valley. They carried with them sand, gravel and stone, which they deposited as the climate warmed. In addition they gouged the land, dotting the region with lakes and bogs. A network of rivers flowing west to the Pacific now drains the lakes.

The area around Monte Verde today has moderately dry summers and cold, rainy winters. The climate that prevailed in the late Pleistocene after the glaciers had receded probably resembled this pattern, although it might have been slightly cooler and more humid.

A forest made up of a mixture of deciduous and coniferous species covers the region; it abundantly supplies edible tubers, nuts, berries, fruits and soft, leafy plants throughout the year. There are also small game, freshwater mollusks and fish. The nearest point on the Pacific coast, which lies about 60 kilometers west and 25 kilometers south of the Pleistocene settlement, offers many edible species of marine organisms. All these sources of food were available to the early inhabitants.

The site of the settlement is buried in the banks of Chinchihuapi Creek, a small tributary of the Maullín River. The creek runs through a terrace with three tiers. The lowest layer consists of rough sand and gravel laid down between 45,000 and 20,000 B.P. Above the first tier is a layer of sand and small pebbles laid down between 20,000 and 14,000 B.P. Not long after the second layer was deposited the creek began to cut a narrow bed through the terrace; gravel and fluvial deposits line the bed. The two lower layers of the terrace are covered by an upper stratum consisting of more recent soils.

Mario Pino of the Southern University of Chile, the geologist of our project, studied the area around Monte Verde. He concludes that at the time the site was occupied the local topography was similar to what it is today. The two lower layers of the terrace had already been laid down. Chinchihuapi Creek was shallow and narrow, with a maximum width of four meters and a maximum depth of half a meter. Because the climate was slightly wetter and cooler, there may have been a few more bogs than currently exist.

After the settlement was abandoned the local topography changed in a way that proved to be of great benefit to archaeology. The creek began to cut a new channel in the terrace. When the creek moved, a peat bog formed on the banks of the old creek bed and also in the bed itself. A peat bog, which is composed of soil and decaying vegetable matter, prevents air from reaching buried artifacts and thus is an excellent medium for preserving organic remains. The bog ultimately filled the abandoned stream bed and covered the site of the community.

Chinchihuapi Creek ran in its new bed for thousands of years. Then, in 1976, the stream shifted again. The water cut into the filled channel, partially exposing the old banks of the creek and the artifacts buried in them. The remains of the settlement were found in the same year by a team from the Southern University of Chile in Valdivia, where I was then head of the program in anthropology. Mauricio van de Maele, director of the Museo de História y Antropológia, took part in all phases of the work at the site along with my colleagues and me.

Since 1976 the group of workers at the site has expanded to include an interdisciplinary team of 32 nonarchaeologists drawn from such fields as geology, botany and paleontology. At least one more field season will be needed to determine the full extent of the settlement and specify the details of the pattern of buried deposits.

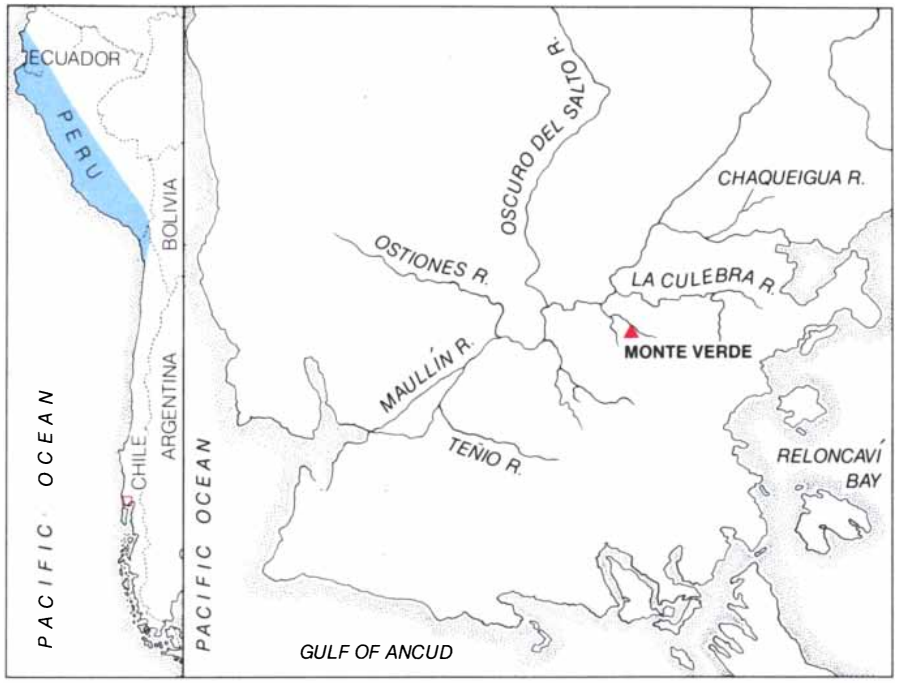
Although only part of the site has been excavated, test pits in the outlying areas suggest that the buried area of occupation has a total extent of 70 by 100 meters. Chinchihuapi Creek partitions

this area of 7,000 square meters into a northern part and a southern part. Most of the digging has taken place on the north bank of the creek but preliminary work shows there are also structures buried on the south bank. The main excavation area on the north bank covers about 750 square meters; this area has been arbitrarily divided into an eastern part and a western part.

The eastern part of the north bank contains the foundations of 12 dwellings. The arrangement of the foundation members shows that the structures were rectangular. The foundations also reveal a striking fact: the huts were joined by their walls to form two parallel rows. Test excavations farther to the west and on the south bank suggest that there are also remnants of multiunit structures buried there.

The foundations are made up of small logs and roughly cut hardwood planks held in place by stakes driven in the ground. Fallen branches and the broken stubs of vertical members show that the hut frames were constructed of local hardwoods. Saplings placed every few feet along the foundation members defined the walls. A few small fragments of what has tentatively been identified as animal skin clinging to the fallen side poles indicate that hides draped on the

FOUNDATION OF WISHBONE-SHAPED STRUCTURE uncovered at Monte Verde consists of compressed sand and gravel that probably supported a hemispherical frame of saplings covered with animal hides. The raised platform at the rear forms the short protrusion of the "wishbone." The two longer, curved legs flank the opening through which the structure was entered. The foundation is 3.9 meters long and three meters wide (the red-and-white arrow in front of the opening is 25 centimeters long). The wishbone-shaped structure may have served as the site for dressing meat, preparing hides, making stone tools or rendering communal medical care. The space in front of the foundation is outlined by branches, stones and carbonized materials. Several layers of soil of recent origin had to be removed in order to reach the remains, which lie on a geologic layer laid down between 20,000 and 14,000 B.P. (before the present).



MONTE VERDE lies in a valley in southern Chile formed by the Andes to the east and the coastal mountains to the west. It is often assumed that in South America permanent human settlements with an agricultural or maritime economy first developed around 6000 B.P. in the central Andes (*colored area on map at left*). The permanent settlement at Monte Verde is far to the south of this region, and the remains discovered at Monte Verde are at least 5,000 years older than those that have been found in the central Andes. The people of Monte Verde may have attempted to domesticate plants, although the evidence is far from conclusive. The data suggest that a settled way of life was developed independently in several regions of South America. Monte Verde is on Chinchihuapi Creek, a small tributary of the Maullín River (*right*).

poles formed the walls of the dwellings.

The dwellings vary in size from three by three meters to four by 4.5 meters, but they all have the same basic rectangular plan. The villagers evidently entered them by parting the hides that served as the outside wall. Inside each hut were found plant remains, stone tools, food stains and braziers: shallow pits lined with clay that held burning coals. The inhabitants probably used the braziers for heating the huts and warming cooked food.

That the residents of the complex cooked collectively is implied by the discovery of two large hearths in central positions outside the huts. Flotation studies, in which organic matter can be separated from soil and clay, have yielded carbon, edible seeds, nuts, fruits and berries from the braziers and the hut floors. Three rough wood mortars and several grinding stones found near the hearths were probably used to prepare plant foods for cooking.

Other artifacts uncovered on the east side of the site confirm that wood was of great significance in the life of the community. Piles of lumber, logs and other worked pieces of wood indicate that the residents stored firewood and made wood tools. Indeed, the many partially shaved or carved objects found near the huts suggest that much time was devoted to woodworking. Among the finished wood tools are three hafts with stone scrapers mounted on them, possibly used for preparing hides; a sharp-pointed tool resembling a lance that is about 1.5 meters long, and a collection of digging sticks and other small wood objects.

No human bones have yet been found in the excavation, although there may be graves in the part of the site that has not been thoroughly explored. We have, however, uncovered two pieces of indirect evidence about the anatomy and physiology of the community members. One bit of evidence is the imprint of a left foot, 16 centimeters long, that was preserved in the clay around one of the communal hearths. It is difficult to estimate a person's size and height from a footprint, but it is probable that the impression was made by a child or an adolescent.

Indirect information about the inhabitants' physiology comes from coprolites (preserved feces) that appear to be of human origin. The coprolites were recovered from small pits near the communal hearths. Coprolites often yield data about the nutrition of prehistoric people, and work is being done to identify the pollen and plant content of the ones found in the settlement.

The inhabitants of Monte Verde apparently divided their community into living areas and areas employed for other purposes. About 30 meters west of the huts we found the remains of a structure with a design quite different from that of the dwellings.

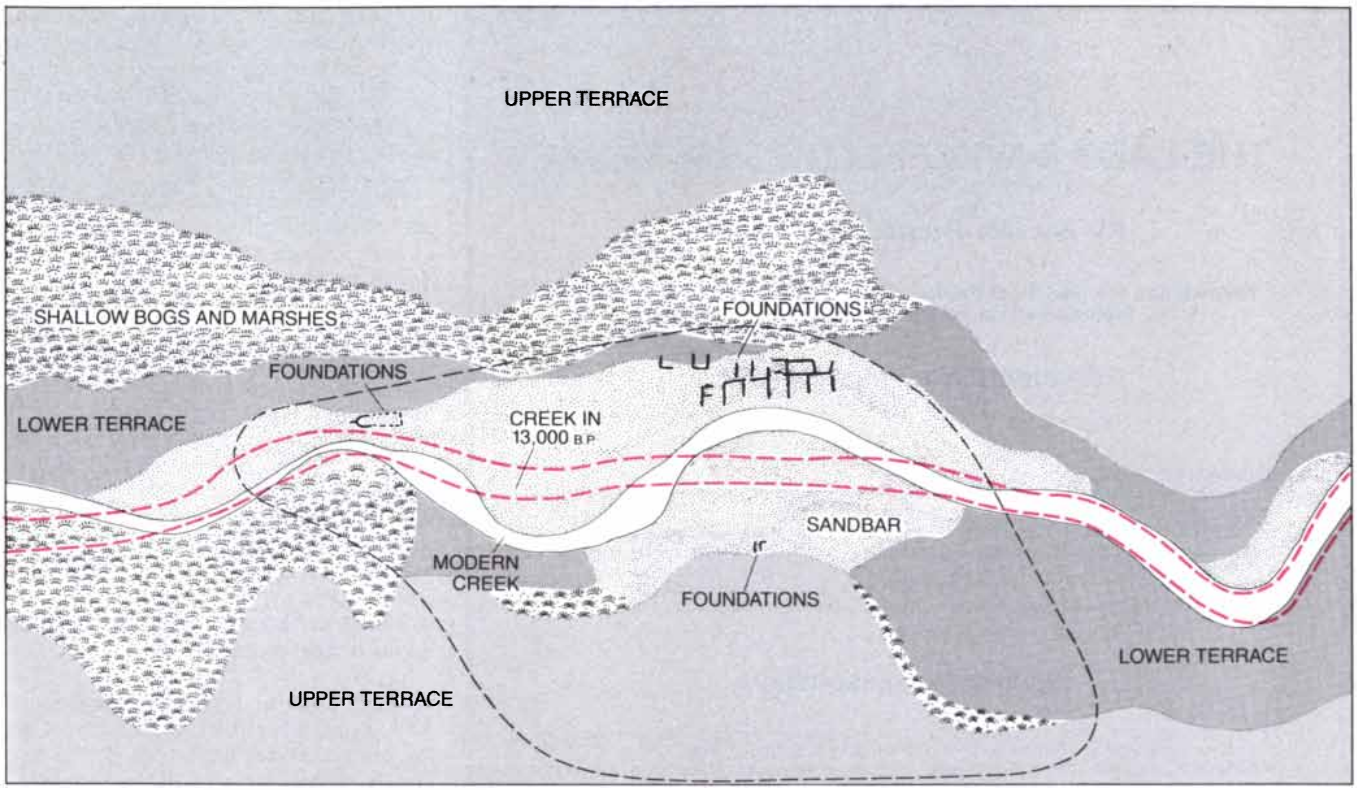
The foundation of the structure consists of sand and gravel compressed in a form that resembles the "wishbone" of a turkey. The main axis of the foundation runs east and west. A small raised platform at the west end corresponds to the small protrusion of the wishbone. Two long curved legs extend to the east and their ends define an opening opposite the platform. The entire foundation is 3.9 meters from east to west, three meters from north to south and about .6 meter high.

Along both long legs of the wishbone were found fragments of wood uprights spaced about half a meter apart. These are probably the remnants of a hemispherical frame that might have been covered with hides. The rounded structure was clearly entered through the opening between the curved legs rather than through the wall as the huts were.

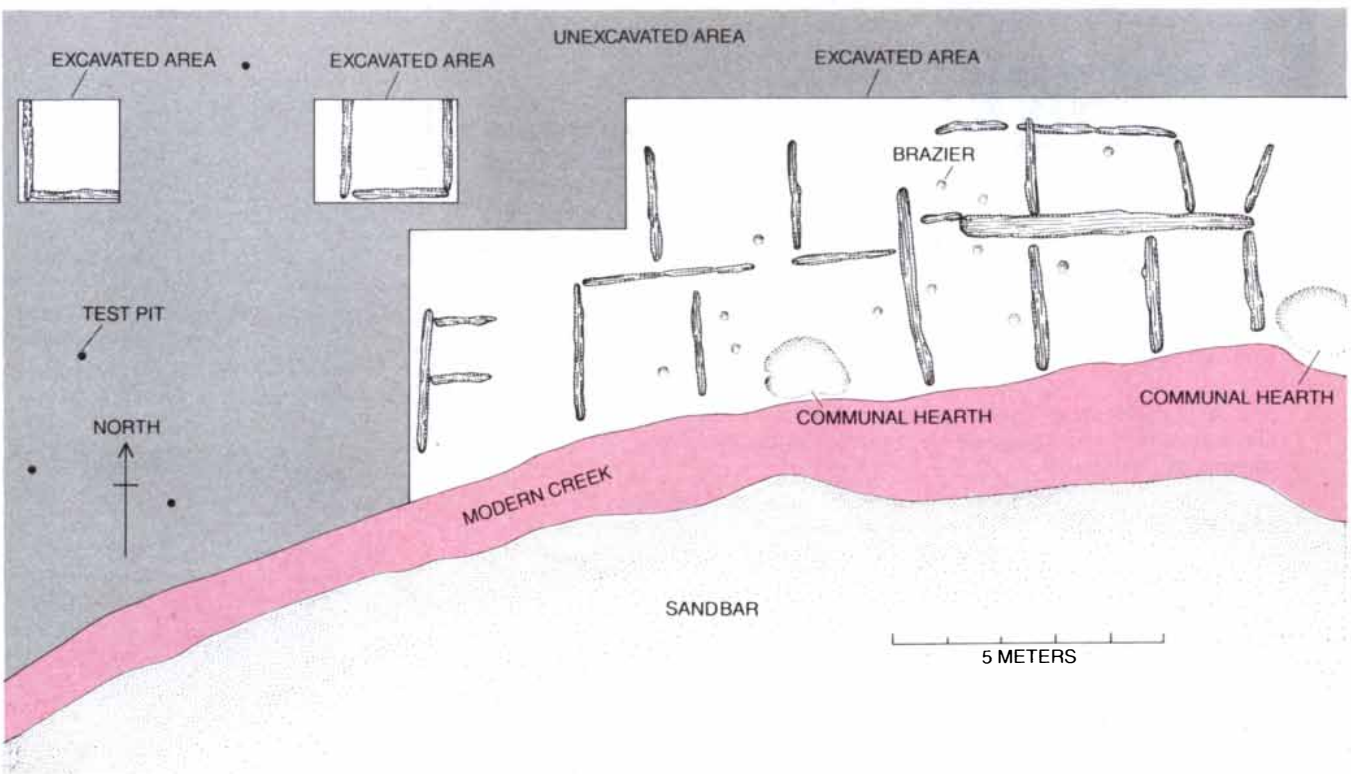
In front of the entrance to the wishbone-shaped structure is a rectangular space outlined by branches where a few stone tools were found. Braziers similar to the ones in the dwelling huts were also uncovered. Scrapings from the braziers and the raised platform at the west end of the wishbone have yielded an intriguing combination of bits of preserved animal hide, burnt seeds and stalks of the totora plant (*Scirpus californicus*) and masticated leaves of the boldo plant (*Peumus boldus*). Residents of the area around Monte Verde still brew boldo leaves into a tea that is believed to have medicinal properties.

Scattered in a large, rough circle around the structure are a group of hearths, wood piles, wood artifacts, stone tools and mastodon bones. Indeed, almost all the mastodon bones, all the tools made of stone from beyond the immediate area of Monte Verde and all the complex stone tools flaked on both sides that have been uncovered at Monte Verde come from this circular concentration. The evidence suggests the building was the site for activities including the dressing of meat from big kills, the making of sophisticated stone tools and perhaps specialized medicinal practices.

The wishbone-shaped building and the dwelling huts are components of a planned settlement that was integrated both spatially and functionally. Several observations support the proposition that the structures are the remains of a single community and not of bands that occcupied the site at different times. First, all the foundations at the site lie on the buried surface of the same geologic layer and therefore were probably constructed at the same time. Second, none of the structures overlap, as they proba-



LOCAL TOPOGRAPHY has been reconstructed to show the environment of the community at Monte Verde as it might have looked at the time of settlement. The site lies on a terrace composed of two layers of sand deposited by retreating glaciers at the end of the period of Pleistocene glaciation. When the climate warmed, Chinchihuapi Creek began to cut a channel through the terrace. Later the creek shifted its bed. The modern creek bed is shown by solid black lines and the original bed by broken colored lines. The human band probably established their settlement on the banks of the creek between 13,000 and 12,500 B.P. They occupied a total of 7,000 square meters (*broken black line*). Most of the digging has been done on the north bank of the creek. In the eastern part of the north bank the author and his colleagues found the remains of dwellings; in the western part they uncovered the foundation of the wishbone-shaped structure.

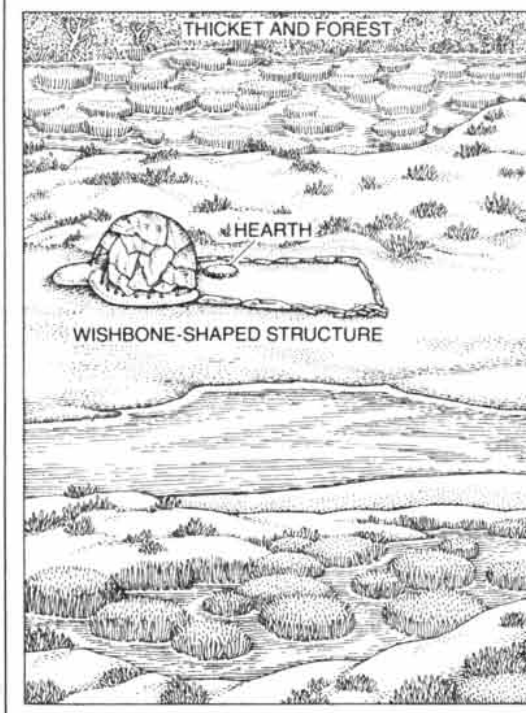


FOUNDATIONS of 12 dwellings mark the eastern part of the site on the north bank of the creek. The foundations consist of hardwood logs held in place by stakes. Ten of the dwelling huts were joined to form two parallel rows adjacent to the creek. In the floors of several huts are shallow, clay-lined pits. The pits are braziers that may have been used for warming cooked food. Two large hearths were probably employed for communal cooking. The structures uncovered at Monte Verde are the earliest architecture known in the Americas.

bly would if they had been built at intervals by different groups. Third, the same types of hearths, braziers, wood and stone tools, plant remains and animal bones are found throughout the site, implying that all the remains originated in a single cultural episode.

What did the people of Monte Verde eat? Preliminary findings by the botanist of our project, Carlos Ramirez of the Southern University of Chile, and the paleontologist, Rodolfo Casamiquela of the Fundación Ameghino in Argentina, have begun to elucidate the residents' diet. The Monte Verdeans lived mainly on wild plants supplemented by several types of meat. The plants were gathered from nearby bogs, marshes and forests and also from the delta of the Maullín River on the Pacific coast. Freshwater mollusks were also found at the site. No fishbones or marine shells have been recovered, but the inhabitants might have gathered marine organisms and left the bones or shells on the beach.

Mastodons and smaller game were hunted for their flesh. Almost all the animal bones unearthed at the site are from the carcasses of seven mastodons. Most of the bones are broken ribs but there are also a few fragments of skulls, teeth and legs. The sample of bones makes it seem probable that the residents killed or scavenged the mastodons elsewhere and carried home only meat-bearing bones. One scapula from a paleocamelid was also found. (The paleocamelids are the ancestors of the



RECONSTRUCTION of the community at Monte Verde shows the settlement included both residential and nonresidential areas. The view in the illustration is north across

modern camelid group, which includes the llama, the alpaca and the vicuña).

Monte Verde is the first site from the Pleistocene epoch in the Americas to yield an abundance of the remains of plants that served an economic function. Among these are the remains of tubers, including the wild potato. It is possible that the tubers were semidomesticated. The sample of preserved tubers is too small, however, and too little is known about the history of such plants in South America to conclude that the residents actually practiced horticulture.

What is perhaps of greater significance than the finding of the wild potato is the wide variety of edible seeds, stalks, leaves, fruits, nuts, berries and roots that have been recovered. These finds are significant because they include the remains of plants that mature during all the months of the year. The plants are from several different environmental zones, implying that the residents ranged widely in search of food. The inclusion of plants from all seasons, however, implies that the inhabitants were settled. Rather than changing their dwelling place continuously during the search (as small hunting-and-gathering bands do) they remained year round in a permanent settlement where they were sustained by a rich and dependable organic base.

Stone technology had a significant role in processing both plant and animal foods. Three different methods for making stone tools are represented among the specimens from Monte Verde: flaking, pecking-and-grinding and the selection of pebbles that were later modified in the course of use. The flaked and pecked-ground tools resemble artifacts found in the remains of later South American cultures, but the pebble technology has not been duplicated among the tools that have so far been described in the Americas.

In the flaking technique a stone is formed into a rough approximation of the desired shape; then the craftsman chips away flakes of varying size, using an implement of stone, bone or hardened wood. The flaking can be done on one face of the tool or on both faces.

Eleven flaked tools were found at Monte Verde; of these only four are bifacially flaked. Two of the bifacially flaked tools are large hand axes and two are crude chopping tools. The choppers are made of local stone; the axes are made of basalt and quartzite that do not come from Monte Verde. The four bifacially flaked tools were found among the mastodon bones on the west side of the site near the wishbone-shaped building. The tools flaked on one side only were for the most part found near the dwellings and all of them are made from local stone.

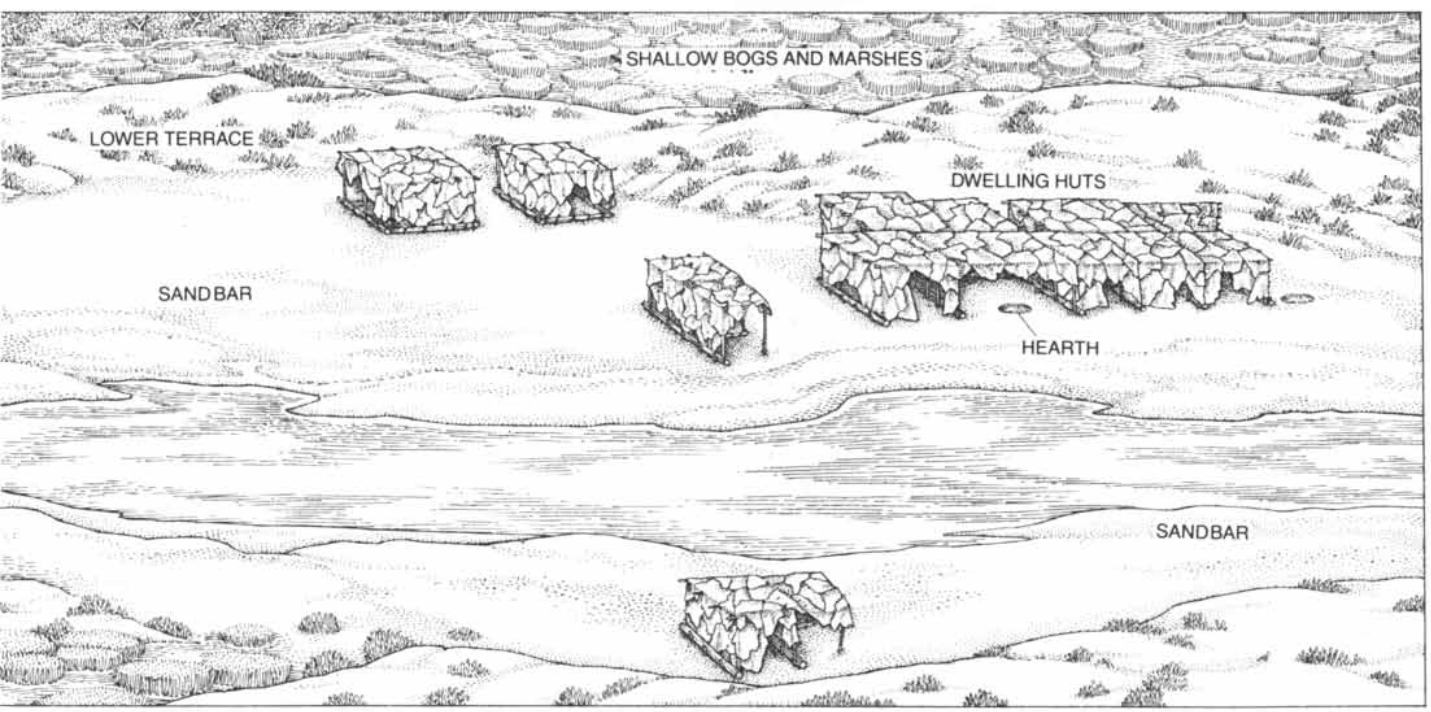
The making of the bifacially flaked stone tools requires a much higher level of skill than the making of the unifacially flaked stones, pecked-ground tools and pebbles characteristic of the eastern part of the site. It is possible that the techniques were employed by distinct groups within the community.

In the pecking-and-grinding method the craftsman gave an unfractured stone a rough, spherical shape by removing small chips. The craftsman then polished the surface by rubbing the partly finished object against other stones. Projectiles, hammers and grinding stones were made in this way.

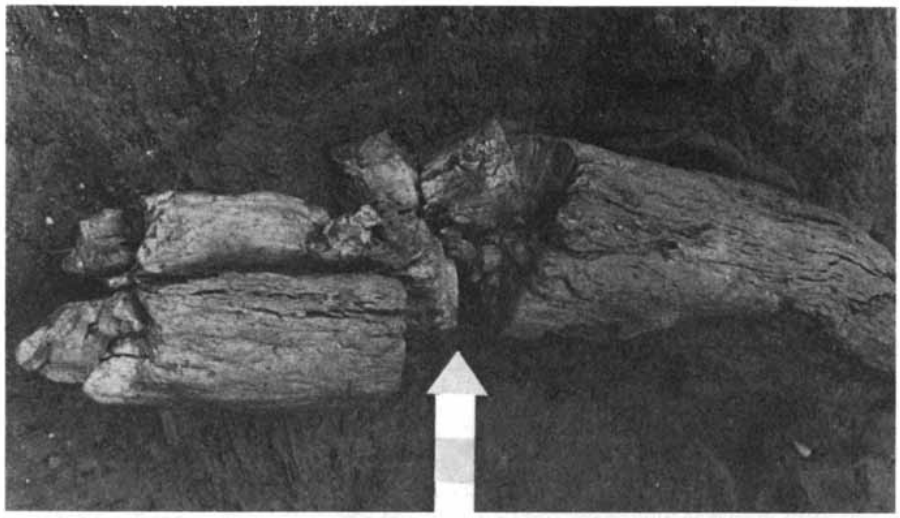
Among the pecked-ground specimens from Monte Verde are 28 almost perfect spheres made of tonalite and basalt from Chinchihuapi Creek. Two of the spheres are girdled by grooves. These were probably attached to a cord to form a bola used for harassing or stunning large game before the kill. The stones without grooves were probably hurled from a sling. We found most of the spherical stones in the west end of the site.

The pecking-and-grinding technique also had domestic applications, as is shown by hammerstones and grinding stones. We retrieved 15 hammerstones made of dense local basalt and granite from both the residential and the nonresidential areas. Ten grinding stones lay near the wood mortars and the hearths in the dwelling section. The grinding tools are made from local granites and also from a fine-grained granite found on the Pacific coast. The abundance of grinding stones and mortars confirms the crucial importance of plant foods in the diet of the settlement.

The majority of the stone tools were not flaked or ground but were simply picked from the creek bed and minimal-



Chinchihuapi Creek. The residential areas were in the eastern part of the north bank and on the south bank. Most of the lumber, wood tools and grinding stones were found near the dwellings. The nonresidential area, where the wishbone-shaped structure was found, was in the western part of the settlement on the north bank. Most of the mastodon bones and complex stone tools flaked on both sides were deposited in a rough circle around the wishbone-shaped building; they were mingled with plants that may have had medicinal uses.



CHOPPED LOG from the residential area of the site was probably cut with a stone ax at the point indicated by the arrow. The log was near one of the communal hearths, which suggests that it could have been intended for firewood or for use in constructing an additional dwelling.

SCAPULA OF PALEOCAMELID was excavated from the floor of one of the dwellings. Stone and wood artifacts surround the bone, which is at the center of the illustration. Among the artifacts are pieces of the hut frame. Paleocamelids are the ancestors of the vicuña, the alpaca and the llama. The inhabitants hunted to augment a diet based mainly on plant foods.

ly modified in use. A preliminary classification by Michael Collins of the Museum of the Southwest in Midland, Tex., and me shows that the 550 pebbles in this category make up 90 percent of the stone tools from the site.

The bed of Chinchihuapi Creek contains fragments of volcanic rock such as basalt, andesite and tonalite that were fractured when they were deposited near the site long before it was occupied. Some of the naturally fractured stones have edges suitable for cutting, planting, scraping and gouging. The residents had only to choose a pebble with an edge appropriate to the task. The stone could even be mounted on a wood handle, as two of the artifacts are. Many of these pebbles exhibit clear traces of wear on their edges.

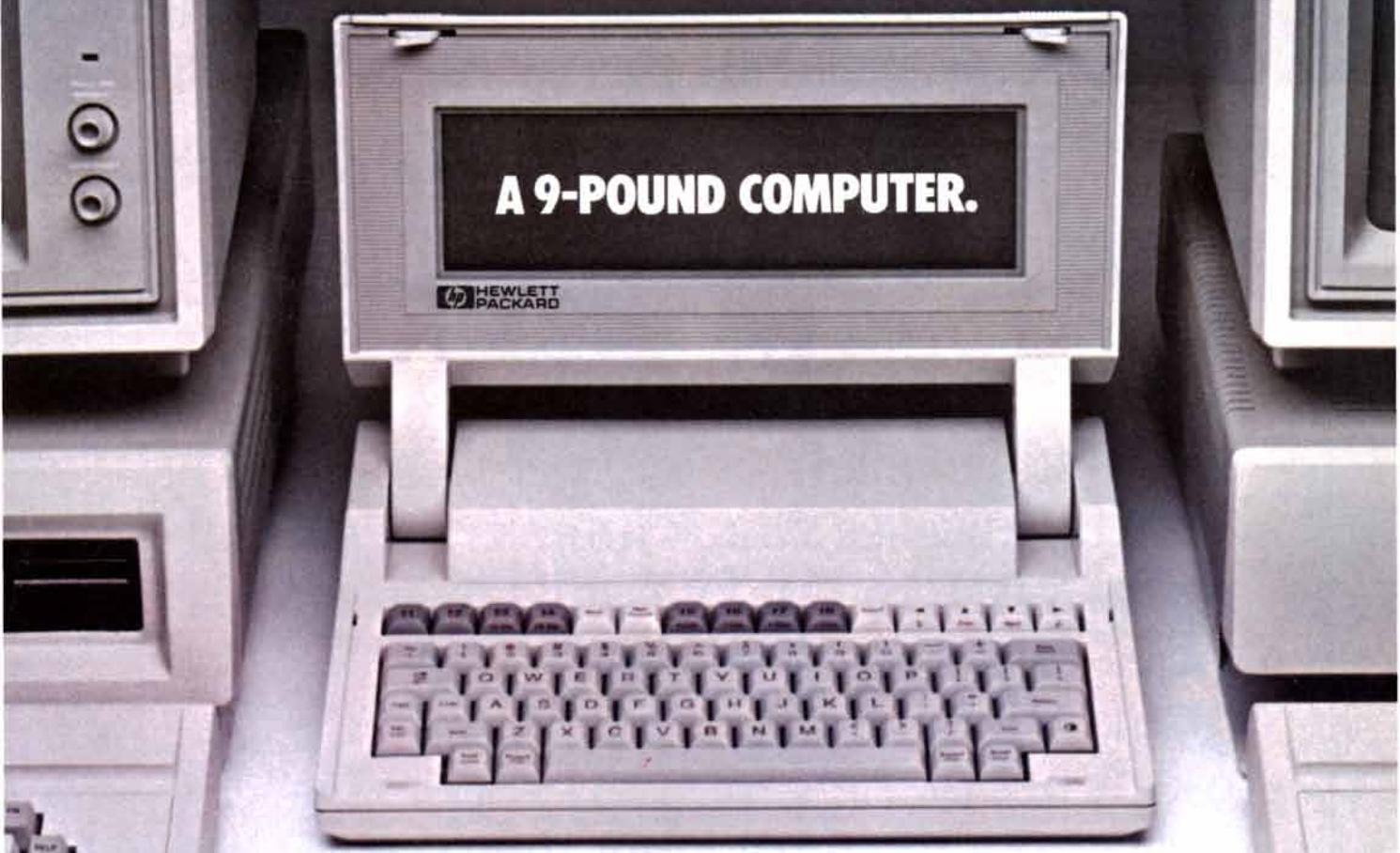
The relative absence of flaking debris tends to confirm that working in stone did not contribute as much to the community as working in wood. The presence of tools made from stones that are not indigenous confirms that the residents covered a wide territory.

In addition to the stones a few mastodon-bone tools have been retrieved from the site. One long bone fragment has small cuts and depressions similar to those found on tools for stone flaking that have been recovered from other late Pleistocene sites in South America. One fragment of bone might have served as a lance point or even a projectile point. Two fragments of mastodon tusk were polished and also show deep parallel marks of wear on their sides. Most of the bone tools come from the west end of the site near the wishbone-shaped structure.

The material retrieved from the banks of Chinchihuapi Creek is contributing to a fundamental revision of the accepted picture of early culture in the New World. Until about 20 years ago most archaeologists thought the first human bands crossed a land bridge from Asia in about 10,000 B.P., bringing with them a specialized, nomadic culture based on hunting large mammals. More recent findings, however, suggest the New World was populated in 20,000 B.P. or even earlier and that the first inhabitants had a generalized economy based as much on gathering plants as on big-game hunting.

One reason the older hypothesis was able to gain acceptance is the poor condition of the known sites dating from 10,000 B.P. or before. Disturbed soil layers, inconclusive radiocarbon dates, dubious manmade tools and the mixing of artifacts from different periods of occupation have considerably reduced the value of the older sites. Even at protected sites, artifacts made of stone or bone have resisted decomposition far better than wood tools or plant remains. As a result the hunting and butchering tech-

IT'S A SMALL MIRACLE HOW HEWLETT-PACKARD PUT 656K OF MEMORY, LOTUS 1-2-3, WORD PROCESSING, A TELECOMMUNICATIONS MODEM AND COMPLETE IBM CONNECTABILITY INTO



THE PORTABLE.

For years business people had to choose between the power of a desktop computer and the limited capabilities of the first portables. That problem was solved when Hewlett-Packard introduced The Portable.

The Portable is designed with more total memory than most leading desktop personal computers...656K in fact. That includes 272K of user memory. So, The Portable's built-in business software can work with enormous amounts of data.

1-2-3™ from Lotus,™ America's most popular spreadsheet, file management and business graphics program, is permanently built into The Portable. So is Hewlett-Packard's word processing program, MemoMaker. Just press the key and you're ready to work.

The Portable even has a built-in modem and easy-to-use telecommunications software to send or receive data using a standard telephone jack.

If you use a Hewlett-Packard Touchscreen PC, IBM® PC, XT or an IBM compatible you'll be glad to know that your desktop and The Portable can talk to each other with the simple addition of the Hewlett-Packard Portable-Desktop Link.

The Portable's rechargeable battery gives you 16 hours of continuous usage on every charge.

Finally, you can work comfortably on a full size keyboard and an easy-to-read 16-line by 80-column screen. And it all folds shut to turn The Portable into a simple nine-pound box.

The Portable. A small miracle...perhaps. But then consider where it came from.

See The Portable and the entire family of personal computers, software and peripherals at your authorized Hewlett-Packard dealer. Call (800) FOR-HPPC for the dealer nearest you.

Setting You Free

HEWLETT PACKARD

PG02412 236 B

IBM is a registered trademark of International Business Machines Corporation. 1-2-3 and Lotus are trademarks of Lotus Development Corporation.

EVERYONE SAID SMALLPOX WAS INCURABLE, EVERYONE BELIEVED TB WAS INCURABLE, EVERYONE THOUGHT POLIO WAS INCURABLE, HOW LONG WILL MULTIPLE SCLEROSIS BE INCURABLE?

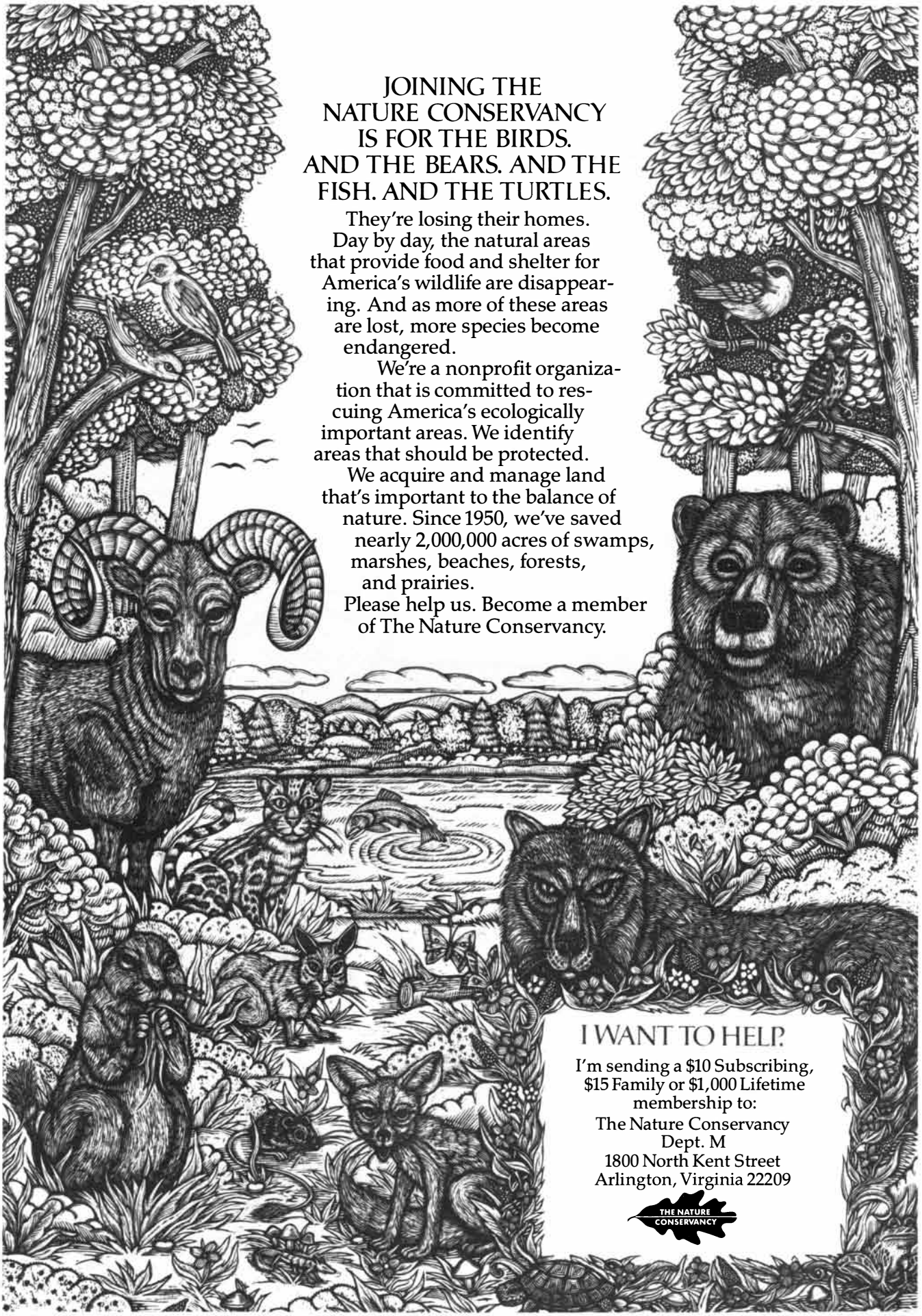
Multiple Sclerosis is a merciless, crippling disease which attacks the central nervous system. And we still haven't found the cause, let alone the cure.

How long it will take depends on the researches of the various MS Societies throughout the world. But they themselves depend almost entirely on voluntary contributions. So, ultimately, it rather depends on you.

Please help your national MS Society to bring the day of the cure a little closer.

HELP YOUR NATIONAL MS SOCIETY FOR THE WHOLE WORLD'S SAKE.

International Federation of Multiple Sclerosis Societies Stubenring 6, A-1010 Vienna, Austria.



JOINING THE NATURE CONSERVANCY IS FOR THE BIRDS. AND THE BEARS. AND THE FISH. AND THE TURTLES.

They're losing their homes. Day by day, the natural areas that provide food and shelter for America's wildlife are disappearing. And as more of these areas are lost, more species become endangered.

We're a nonprofit organization that is committed to rescuing America's ecologically important areas. We identify areas that should be protected.

We acquire and manage land that's important to the balance of nature. Since 1950, we've saved nearly 2,000,000 acres of swamps, marshes, beaches, forests, and prairies.

Please help us. Become a member of The Nature Conservancy.

I WANT TO HELP.

I'm sending a $10 Subscribing, $15 Family or $1,000 Lifetime membership to:

The Nature Conservancy
Dept. M
1800 North Kent Street
Arlington, Virginia 22209

THE NATURE CONSERVANCY

How an idea in yesterday's funny papers can become tomorrow's front page headlines.

Dick Tracy's wrist radio used to belong strictly in the Funnies.

But a revolution in electronics is moving ideas like these onto Page One.

Thanks to a semiconducting compound called gallium arsenide that's being used to make super microchips by the people at ITT.

These miniature integrated circuits work ten times faster than conventional silicon chips.

And at higher frequencies in smaller spaces.

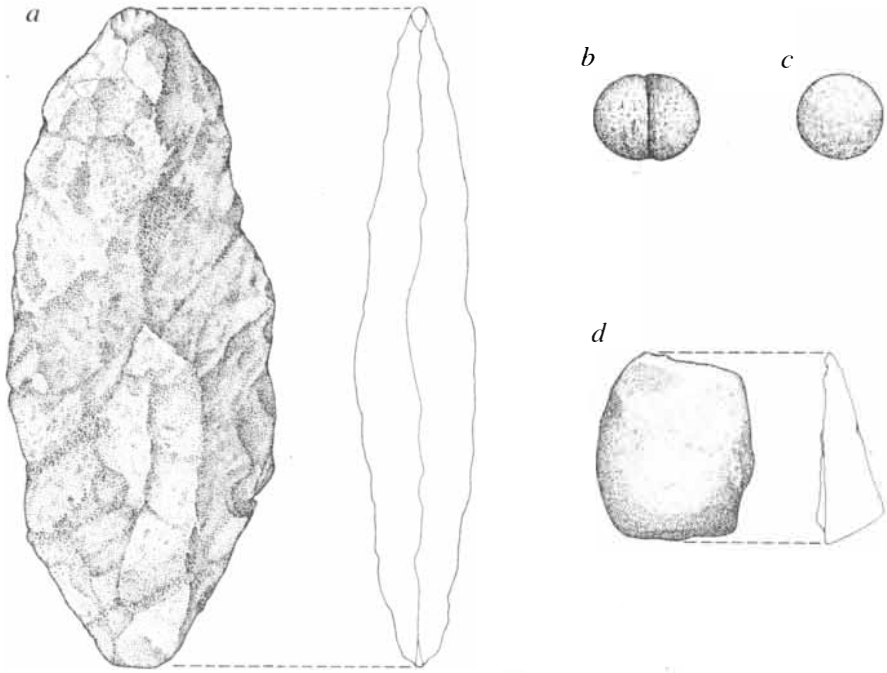
Which could make possible satellite phone calls from personal wrist phones.

And night vision devices for crime detection that are thousands of times more efficient than the human eye.

To find other ways gallium arsenide can help advance the state of the art in electronics, ITT is building a multimillion dollar research center.

There, ITT engineers will be able to use this revolutionary technology to turn the ideas of yesterday into the news of tomorrow.

The best ideas are the ideas that help people. ITT



STONE TOOLS made by three techniques have been found at Monte Verde. The most complex tools, such as the large oval chopper, were made by flaking on both faces (*a*). Round objects were made by pecking away small chips to approximate the desired shape and then grinding the article into final form. Some of the round stones are grooved. Such grooves could have been used to attach the stone to a cord, thereby forming a bola for stunning game before the kill (*b*). Round stones without grooves could have been thrown from slings (*c*). The simplest technique for acquiring a tool was to select a stone with a suitable edge from the stream bed. Among the tools of this type is a scraper that was probably used to prepare animal hides (*d*).

nologies of the early inhabitants are better understood than their domestic technology, social structure or architecture.

The archaeological work done before 1970, which was based in large part on stone and bone objects, tended to reinforce the concept of a culture that developed about 10,000 years ago and was centered on the making of specialized stone tools for hunting game.

In the 1970's findings from Pikimachay Cave in Peru and Meadowcroft Rock Shelter in Pennsylvania provided evidence for a quite different scenario. Both sites have remains that are from 15,000 to 20,000 years old. Organic remains from Meadowcroft show that the gathering of plants supplemented hunting even in very early times.

The excavation at Monte Verde significantly extends these recent findings. Because a peat bog covered the settlement, the assemblage of wood artifacts from Monte Verde is the most complete yet found in the New World. The collection of artifacts demonstrates that wood technology could have contributed as much to the development of early culture as stone technology. In addition the plant remains from Monte Verde are far more abundant than those from the other sites. This organic record gives evidence of the broad base of the economy and supports the idea of a generalized culture.

Other advantages come directly from the geography of the area around the Pleistocene settlement. Both Pikimachay and Meadowcroft are sheltered locations in which it was not necessary for the residents to erect dwellings. Because Monte Verde is in the open, the inhabitants built dwellings and nonresidential structures. The plan of these structures suggests that by the period between 13,000 and 12,500 B.P., New World culture had advanced to a far higher level than previous findings would suggest.

Moreover, it is significant that Monte Verde is set in a forest. Because wooded regions provide such a rich environment for human beings, much of the development of early culture must have taken place there. Yet most of the early sites in the Americas are either in caves or rock shelters or are on unforested plains; hence they cannot reveal much about how the forest dwellers lived.

Many questions remain about the earliest history of human beings in the New World. There is currently considerable debate about whether a generalized economy or a specialized one developed first. There is also debate about whether the two types of culture were adaptations to different physical environments or whether they coexisted in the same environment. Because of the scarcity of well-preserved early sites, resolving such questions forms a research agenda that will occupy archaeologists for decades to come.

HOW TO REACH THE PEOPLE WHO MAKE THE FUTURE HAPPEN IN GERMANY

SPEKTRUM DER WISSENSCHAFT is the German edition of SCIENTIFIC AMERICAN which attracts a young audience of affluent professionals who have come to the top because of their technical expertise. Almost half of them hold top management job titles. The fortunes of major corporations in Germany increasingly depend on technically sophisticated people. SPEKTRUM DER WISSENSCHAFT keeps these people in touch with the advances in science that drive the growth of Germany's industry.

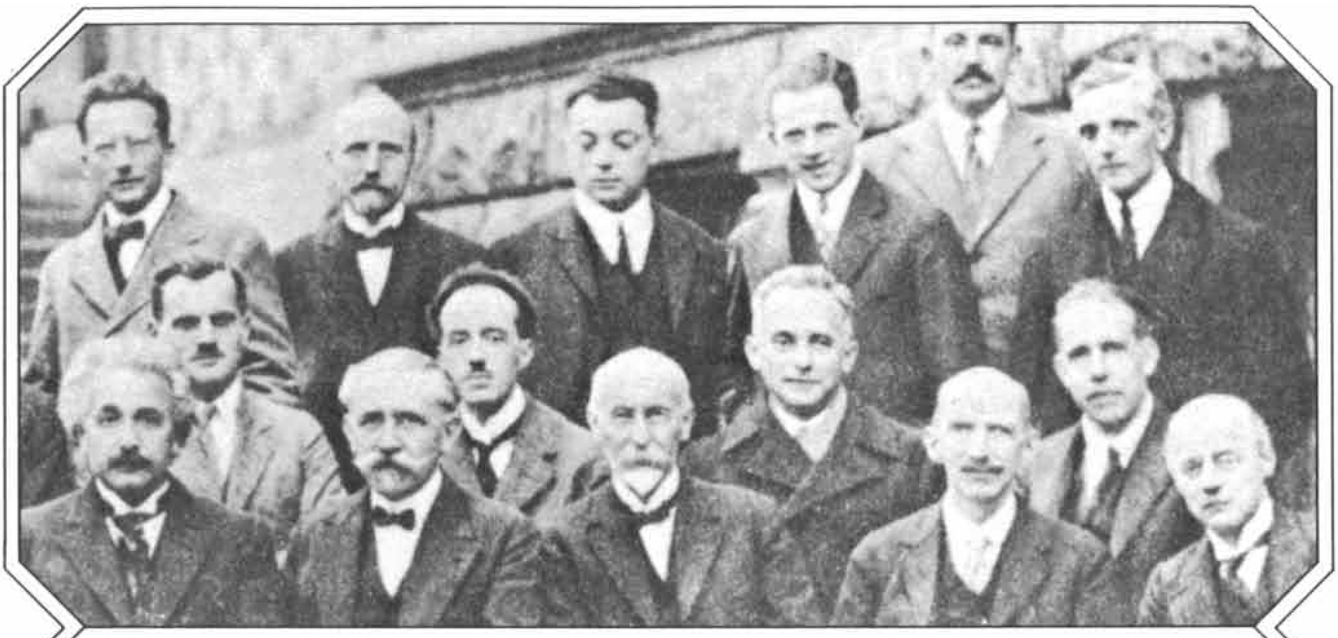
SCIENTIFIC AMERICAN speaks the languages of more than half the world's population. We are the one publication in the world today providing efficient coverage of technology-based management. We reach men and women in industry, whose qualification to make technical decisions places them in key positions in their country's government and industry.

SCIENTIFIC AMERICAN in eight languages: English, Spanish, Italian, French, German, Japanese, Chinese and now Russian has gathered in its audience three million people who make the future happen around the world.

For more information on our German-language edition contact:

Claus G. Firchow
SPEKTRUM DER WISSENSCHAFT
Moenchhofstrasse 15
6900 Heidelberg
FEDERAL REPUBLIC OF GERMANY
Telephone 011-49-6221-46037

or in New York
John Kirby
V.P./Advertising Director
SCIENTIFIC AMERICAN
415 Madison Avenue
New York, New York 10017
Telephone 212-754-0262

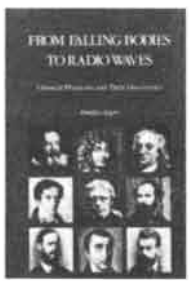


Discover the human side of physics

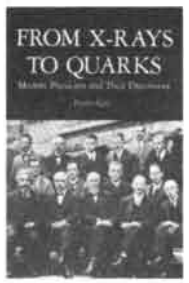
EMILIO SEGRÈ

His new book on the roots of physical thought
FROM FALLING BODIES TO RADIO WAVES

His critically acclaimed history of modern physics
FROM X-RAYS TO QUARKS



In these companion volumes, Nobel Laureate Emilio Segrè offers a warm and close-up look at the human underpinnings of scientific experiment and discovery. In examining the work of both classical and modern physicists, Segrè emphasizes the fascinating and complex personalities that are too often treated only in footnotes.



Available at fine bookstores.

Just published, **FROM FALLING BODIES TO RADIO WAVES** traces the development of scientific thought from the works of the "founding fathers" to the more recent discoveries of Maxwell, Boltzmann, and Gibbs. In Segrè's whimsical prelude, he asks us to imagine him conversing with these scientific giants, moving within the times in which they lived. We peer through Galileo's telescope, call on an inhospitable Isaac Newton, and discuss the enigma of magnetism with Michael Faraday.

In **FROM X-RAYS TO QUARKS**, Segrè describes the evolution of atomic and subatomic physics, interwoven with the personal histories of the leading researchers, and set against the political and cultural backgrounds of those various times and places. As *The American Journal of Physics* said, "Segrè illuminates this vast and cluttered field with mastery of both detail and perspective. This excellent book is one every physicist will want. It will also be a fine gift for non-physicist friends with strong scientific interests."

In his engaging and highly readable style, Segrè communicates the inspiration, creative effort, and true drama of scientific discovery.

SPECIAL LIMITED TIME OFFER Use the accompanying coupon to order the two-volume set and save more than 25%! Offer expires September 1, 1984.

☐ **YES!** I would like to discover the human side of physics. Please send me:

QUAN.	ISBN	TITLE	PRICE
	1482	From Falling Bodies to Radio Waves (paper)	$13.95
	1147	From X-Rays to Quarks (paper)	$12.95
		SPECIAL TWO-VOLUME SET PRICE	$19.95
		Postage and Handling New York, California, & Utah residents add sales tax.	$ 1.50 $
		TOTAL	$

☐ I enclose a check made payable to W. H. Freeman and Company.

☐ MasterCard ☐ VISA _____ Account #

____/____ Exp. Date _____ Signature (All credit card orders must be signed)

Name _____

Address _____

City _____ State _____ Zip _____

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

0152

The Continuous Processing of Metals in the U.S.S.R.

A research institute has developed ways to uninterruptedly convert iron and other ores into finished products so that waste is reduced, quality is improved and output is raised

by A. I. Tselikov

The traditional way of making metal has been to carry out a series of separate operations that transform a metal-bearing raw material into a finished product. Typically the material has to be moved several times. To make steel, for example, iron ore is smelted to iron in a blast furnace, the iron is transferred to a steelmaking furnace, the molten steel is moved to a casting pit or a continuous-casting facility, the cast steel goes to a rolling mill (perhaps to more than one) and the rolled steel is sent to a manufacturing plant. For other metals such as aluminum and copper similar steps of smelting, melting, casting, rolling and finishing are followed.

Such procedures consume time, materials and financial resources. It follows that important economic advantages can be gained by continuous processing, in which the materials move from the ore stage to the finished form in an uninterrupted stream. Among the results of this approach are higher yields, less waste of raw materials and of metal during the manufacturing process, products of better quality and increased productivity by workers and machines.

For many years continuous processing has been a focus of the work in the U.S.S.R. by the All-Union Scientific Research and Design Institute of Metallurgical Machinery in Moscow, of which I am the director. Our findings apply not only to steelmaking but also to the manufacture of other metals and to the shaping of such products as tubes, ball bearings, shafts, axles and gear teeth.

Fundamental to all continuous processing is the continuous casting of steel or such nonferrous metals as copper and aluminum. The main advantage of continuous casting is that it improves the homogeneity of the metal; improved homogeneity enhances the quality of semifinished forms. Those advantages in turn reduce the waste of metal when the still-plastic materials are processed into finished products and also increase the strength of the finished products.

Continuous casting as it is applied to steel is now widely employed in several countries. Molten steel is continuously poured from a ladle into a tundish, which functions like a funnel to deliver the stream into a water-cooled mold. A grip clamp, or dummy bar, introduced into the mold serves as a plug that is attached to the leading part of the stream. As the metal solidifies at the walls of the mold, the dummy bar and solidified metal are drawn out of the mold. Because molten metal is fed into the mold continuously, the solidified strand steadily grows in length.

The metal in the mold solidifies first at the walls, where it forms a thin solid crust, whereas the core remains liquid. A zone of secondary cooling is established downstream from the mold. There the strand is cooled intensively and solidifies throughout.

Traditionally the steelmaking phase ended in the pouring of the molten metal into individual ingot molds. This universal practice led to a large loss of metal due to croppage. When liquid steel in an ingot mold solidifies, it contracts, developing a concave declivity at the open end of the mold. The open space is called a shrinkage pipe, and that end of the ingot has to be cropped, or cut away for the full length of the pipe. The loss can range from 15 to 25 percent. Continuous casting substantially reduces such loss.

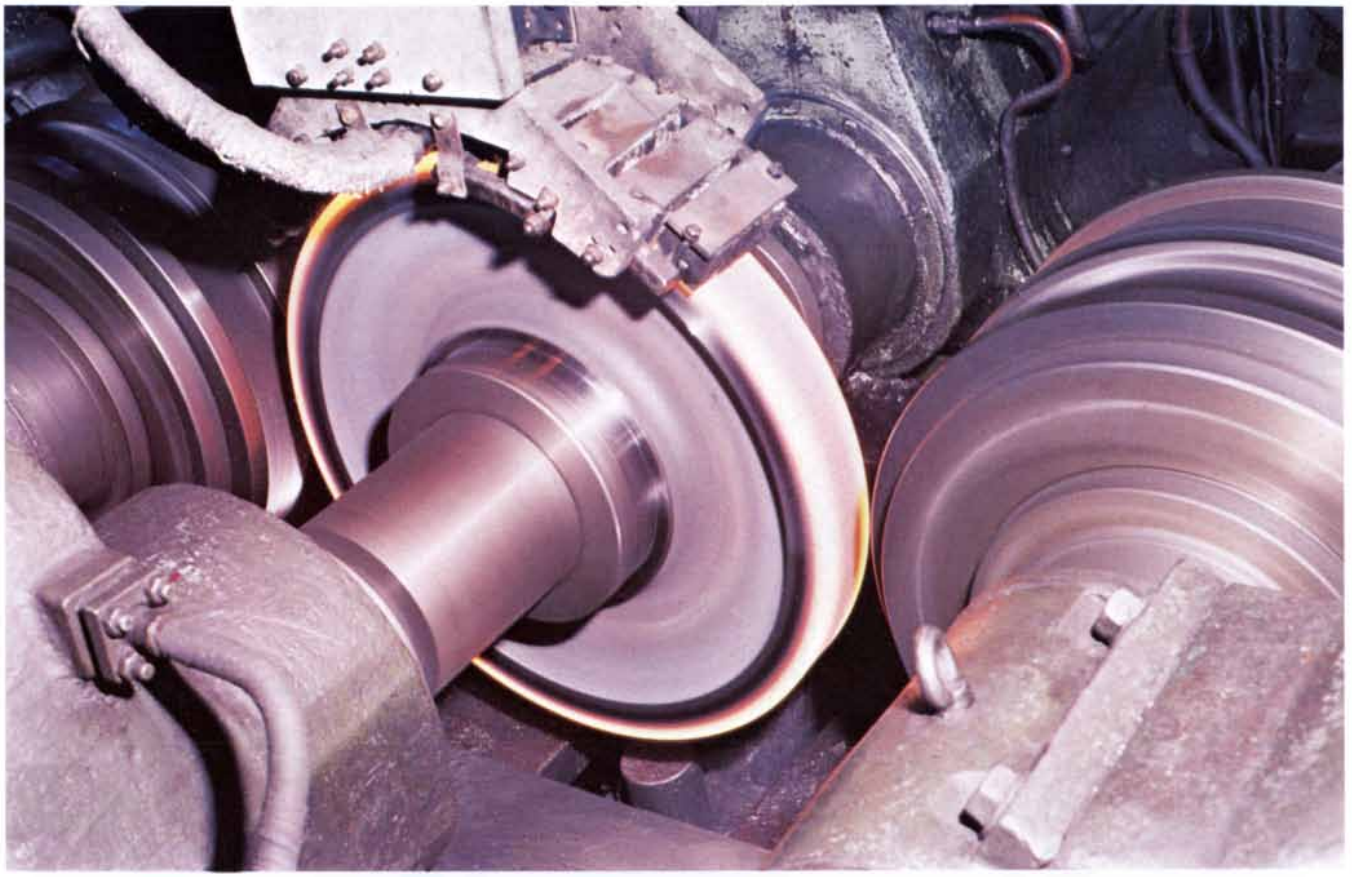
In the U.S.S.R. as in the U.S. the first continuous-casting machines were vertical. The height of the column was from about five to seven stories (15 to 25 meters), necessitating the expensive construction of towers or wells. This substantial drawback was later eliminated by arranging the area of solidification along a circular arc or a similar curve; in such a mill the emerging billet (a strand that is approximately square in cross section) is straightened as it passes onto a horizontal plane. The billet solidifies completely as it moves along that plane.

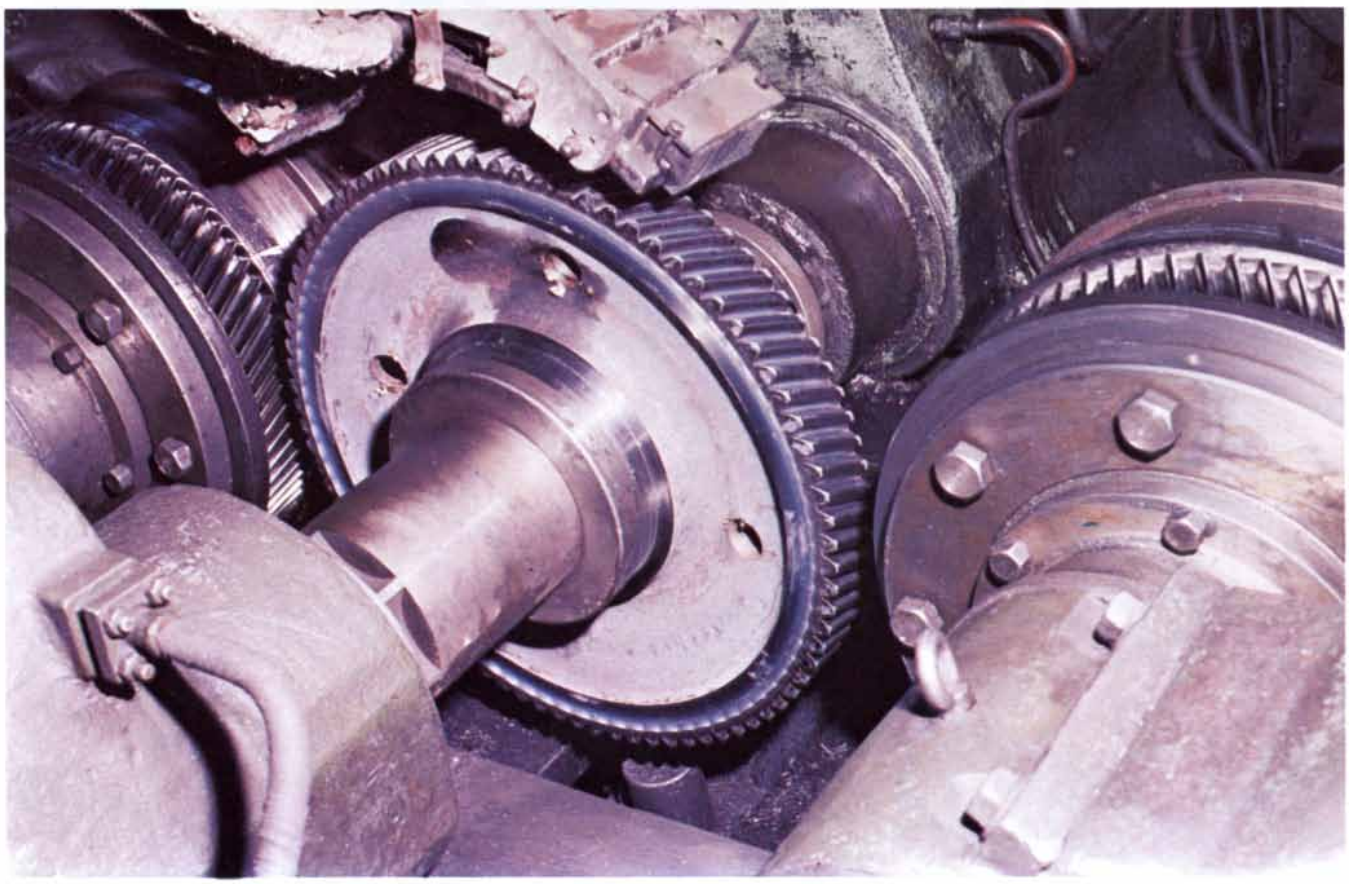
After experimental studies of the solidification of billets made in a radially curved caster workers at the institute proposed building such a machine for the continuous casting of slabs, which have a rectangular cross section with much more width than thickness and can be shaped more easily than billets into final products of a similar configuration. The first idea was to build and test a small machine, but that would considerably delay the industrial application of this obviously efficient process, and so it was decided in the 1960's to build a two-strand machine capable of casting slabs 900 by 180 millimeters (35.4 by seven inches) in cross section. (Today a continuously cast slab of that size would be regarded as small.)

Such a radial machine was designed in a fairly short time at the institute (by a team headed by N. V. Molochnikov, A. A. Tselikov and A. M. Rotenberg) and built at the institute's experimental plant. It was put in operation in 1965 at the Rustavi Iron and Steel Works.

It quickly became apparent that a radial casting machine had several advantages over a vertical machine. One was that the rate of motion of the strand could be raised to 1.2 meters per minute, which was a record at that time for such slabs. Secondly, the machine could be monitored and adjusted more easily, since the ladle was lower and the strand passed onto a horizontal roller table at floor level.

Trials of radial machines, first at the Rustavi plant and later at other plants, provided a firm basis for the industrial application of the process. For example, the Urals Heavy-Machinery Works manufactured large machines for casting slabs as wide as 1,900 millimeters; the machines were put in operation in 1975 at the Novo-Lipetsky Iron and Steel Works. Similar machines are in op-





LARGE GEAR WHEELS one meter in diameter are produced at a rolling mill near Moscow by a method called plastic working. It is a continuous process beginning with the blank wheel. In the rolling mill a high-frequency inductor (the boxlike unit above the workpiece) heats the rim of a rotating blank to a temperature of about 1,100 degrees Celsius (*top*). Next the rim is shaped first in smooth rollers and then in geared rollers. The geared rollers (*bottom right*) have been moved to one side to allow the removal of the finished workpiece.

Tomorrow's buggy whips: With the sophisticated tools shown here a communications network could be designed in a few weeks—and it could be obsolete before it got off the drawing board. Contel Information Systems' MIND* system now harnesses computer

power to get the same job done in minutes. Minutes vs. weeks—and done better, too! No wonder more than 800 network pros have become MIND converts. They get easy-to-update network designs that precisely balance high performance and minimum costs. Continental Telecom and its subsidiaries lead in this new era of network services with the equipment and design tools to make them work with increased efficiency. From telephony to satellites. Architects of telecommunication.

FOR A FREE BROCHURE ON ALL OF CONTEL'S CAPABILITIES WRITE TO:
CONTEL, DEPT. 509, 245 PERIMETER CENTER PARKWAY, ATLANTA GA 30346 © 1984 CONTINENTAL TELECOM INC

* MIND℠—MODULAR INTERACTIVE NETWORK DESIGNER

CONTEL

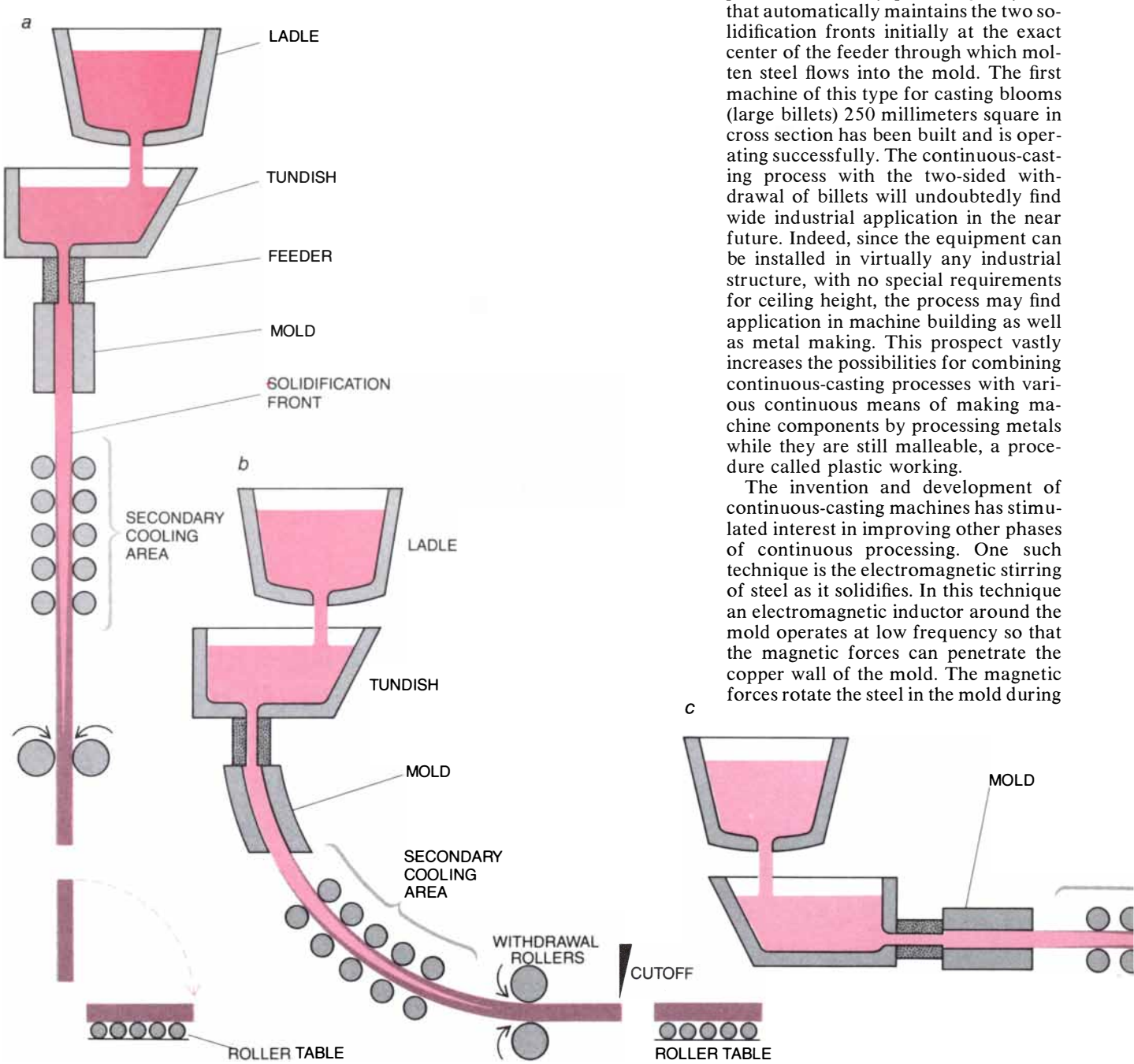
eration at other plants and have been exported to a number of countries. Such a machine can produce as much as 1.2 million tons of steel per year.

Success with radial casting encouraged workers at the institute to explore the possibility of casting steel continuously in a horizontal mold. Molten steel is fed into a slightly elevated tundish and is started moving through the mold with a horizontal dummy bar that serves as a puller. On emerging from the mold the strand enters the secondary cooling area, where it is further cooled with water. When it has fully solidified, it is cut into measured lengths. Since the machine can be installed in a building of moderate height, such as one erected for fabricating or manufacturing plants, the capital investment required for the enclosing structure is substantially less than it is for a structure housing a comparable vertical continuous-casting installation.

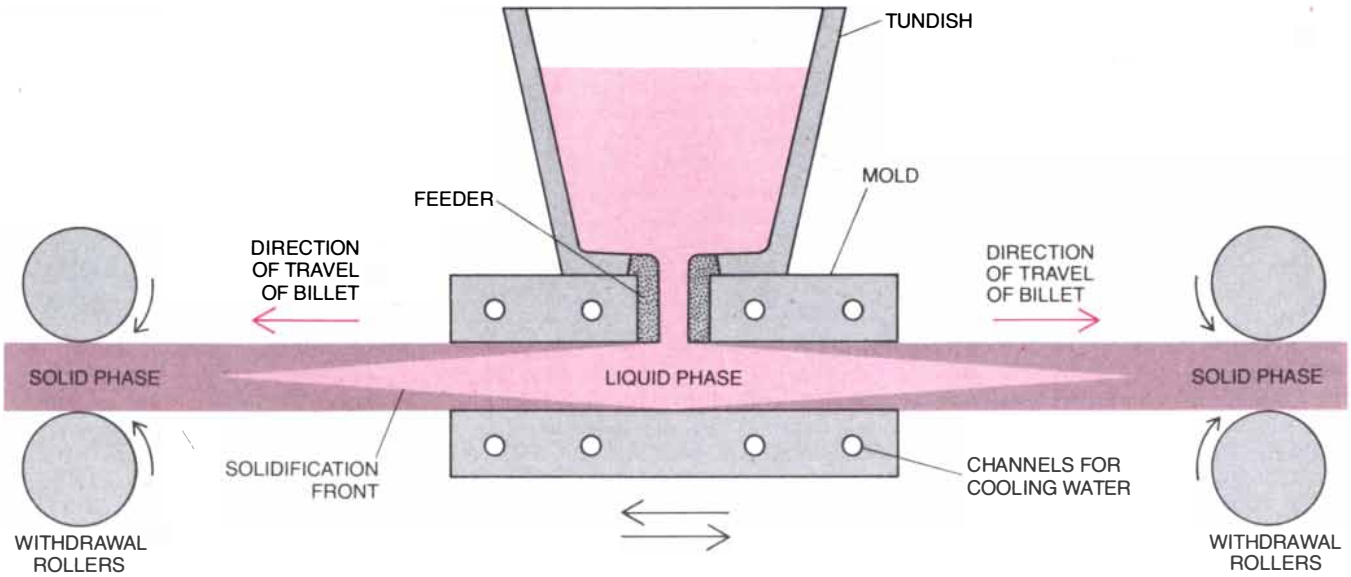
After solving the major problems presented by the development of these machines engineers at the institute addressed a minor but persistent difficulty. This was the formation of ridges on the surface of the strands caused by the reciprocal movement of the mold as a strand is withdrawn. Such ridges cannot be rolled out. To eliminate them we devised a system in which the molten metal emerges at the middle of the horizontal mold in such a way that the two strands can be withdrawn in opposite directions. The procedure goes so smoothly that it eliminates the surface and internal defects resulting from a one-sided withdrawal of the strand.

Experimental studies at the institute have made it possible to stabilize this process further by providing a system that automatically maintains the two solidification fronts initially at the exact center of the feeder through which molten steel flows into the mold. The first machine of this type for casting blooms (large billets) 250 millimeters square in cross section has been built and is operating successfully. The continuous-casting process with the two-sided withdrawal of billets will undoubtedly find wide industrial application in the near future. Indeed, since the equipment can be installed in virtually any industrial structure, with no special requirements for ceiling height, the process may find application in machine building as well as metal making. This prospect vastly increases the possibilities for combining continuous-casting processes with various continuous means of making machine components by processing metals while they are still malleable, a procedure called plastic working.

The invention and development of continuous-casting machines has stimulated interest in improving other phases of continuous processing. One such technique is the electromagnetic stirring of steel as it solidifies. In this technique an electromagnetic inductor around the mold operates at low frequency so that the magnetic forces can penetrate the copper wall of the mold. The magnetic forces rotate the steel in the mold during



CONTINUOUS CASTING of metal, typically steel, was first accomplished by machines that were essentially vertical (*a*), perhaps exceeding 40 meters (131 feet) in height. Such machines were so high that towers or wells had to be built to contain them. The strand of molten steel solidified and was cut to appropriate lengths as it descended. Efforts to decrease the height of the machine resulted in the radial caster (*b*), in which the mold and the secondary cooling area are arranged along an arc with a radius commonly equal to 30 or 40 times the width of the strand. At the end of the radial section the strand, here a billet with an approximately square cross section, passes between withdrawing rollers and enters a horizontal section where it is cut to measured lengths. A radial machine seldom stands



TWO-SIDED WITHDRAWAL OF BILLETS from a horizontal mold is a fundamentally new approach to the continuous casting of steel. Molten steel flows through a heat-resistant feeder (lined with refractory material) into the water-cooled mold and begins to solidify at two fronts. The beginning of each front is automatically maintained exactly at the center of the feeder. The mold moves reciprocally (*black arrows*) to keep the metal from sticking to its walls. The method doubles the output of the casting procedure and avoids or diminishes the surface and structural defects in the billets that are common when metal is withdrawn from only one side of a mold.

the initial stages of solidification. The procedure, which is particularly important in casting highly alloyed steel and steel with a high content of carbon, improves homogeneity. It is likely the technique will be further refined and will become widely employed to enhance the quality of the metals made with it.

Continuously cast steel is better in quality than metal cast in ingot molds but poorer than steel subjected to plastic working (rolling, for example). This fact suggested the concept of direct rolling, which combines casting and rolling in a continuous operation. One advantage of the procedure is that all the intermediate operations between the molten-metal stage and the rolling stage become unnecessary, eliminating the need to move and store ingots. Another is the heat retained by the billets as they enter the rolling stage, since the metal needs to be hot for most kinds of plastic working; in a conventional operation the billets have to be reheated before they enter the rolling mill.

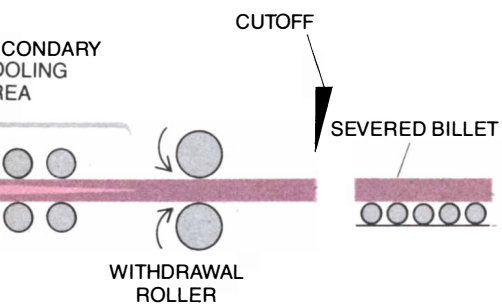
It is also significant that the rolling mill receives an essentially endless billet. Consequently the rolling process can take advantage of continuous tension. If in lengthwise rolling an appreciable tensile force is applied along the axis of the workpiece, the billet can be rolled without spreading; indeed, its width can be made to gradually diminish.

The advantage of this capability becomes obvious when rolling wire rod and shapes whose cross-sectional width is much smaller than the diameter of the roller elements. In the development of machines for direct rolling the critical point is to design the mold from which the billet emerges so that the exit speed is high enough to provide an uninterrupted feed to the rolling mill. The aim can best be achieved by a mold with walls that remain in contact with the solidifying metal of the billet as they both move together. Slippage of the billet in the mold is thus avoided, and no withdrawing force has to be applied at the emerging end of the billet.

A number of molds of this type have been designed and tested. Satisfactory results have been obtained with a rotary mold shaped like a wheel. The mold has a copper rim grooved according to the desired cross section of the billet. A steel band is wrapped over the rim in an arc of approximately 180 degrees. Molten metal fed between the rim and the band solidifies on the rotating wheel and comes out as an endless billet, which is fed into a rolling mill and rolled into wire rod. The first installation of this kind, which was for making aluminum rod, was designed jointly by our institute and the Zaporozhsky Aluminum Works under the direction of P. I. Sofinsky. Since the process was fully continuous and the initial heat of the billet was utilized in rolling, the production of rod increased more than fivefold and the plant paid for itself within a year.

Further development of this process led to the continuous manufacture of copper-wire rods from molten metal. After a series of experiments in the institute's laboratory and at the Balkhashsky integrated mining and copper-smelting works a pilot plant was designed and installed at the Almalyksky integrated works. Additional research under the direction of A. Yu. Shevchenko, an engineer with the institute, led to the production of copper-wire rod superior to that obtained by conventional technology. The success of the pilot plant demonstrated that this is the most promising technique for increasing the production of copper-wire rod in the U.S.S.R.

Direct rolling is harder with steel than it is with nonferrous metals, mainly because with steel the temperature for continuous casting must be considerably higher. The problem is that no effective methods of steel solidification exist enabling the billet to emerge at a speed high enough to be consistent with the rolling speed. A study of the problem revealed that in some cases the solution was to use a continuous caster in conjunction with a specialized planetary rolling mill. Such a mill consists of a full circle of from 18 to 30 small work rollers around each of two large back-up rollers. Because of the relatively large



higher than approximately 12 meters. A more recent development in continuous casting is the horizontal continuous caster (*c*). A continuous caster of this design is likely to be less than eight meters high, which means it can be installed in virtually any industrial building.

curvature of the small rollers, they can "bite" deep into the workpiece and achieve unusually large reductions in a single pass through the mill. Reductions in conventional rolling mills average from 20 to 30 percent, which is to say that a four-inch slab is reduced to three inches. In a planetary mill a reduction of 90 percent in a single pass is not unusual. A plant of this kind designed by the institute and now in operation at an iron and steel works comprises a radial casting machine (forming billets 80 by 60 millimeters in cross section), an induction heater and a planetary rolling mill. The steel is rolled into wire rod from eight to 12 millimeters in diameter that emerges from the rolling mill at a rate of up to 15 meters per second. The plant has proved to be particularly efficient for rolling highly alloyed steels, which have poor plastic properties and can be deformed only in a narrow range of temperature.

Research has shown that the principle of direct rolling can be applied in the manufacture of metal strips. Because solidification progresses inwardly from the outside surface and is more or less inversely proportional to the distance from the wall of the mold, the length of the mold needed for casting billets of thin cross section may be quite short. Therefore a mold for casting, say, thin strips of aluminum can be formed by the cylindrical surfaces of two water-cooled rollers rotating in a horizontal plane.

Workers at the institute have designed a machine on this principle. The metal solidifies between two rollers whose axes are horizontal. The level of molten aluminum in a bath near the rollers is in the same plane. As the rollers turn, molten aluminum is fed into a slit between them; the metal solidifies and passes out and up as a strip from three to eight millimeters thick. The strip is straightened, trimmed at the edges and coiled. Plants of this type have been designed jointly by the institute (under Sofinsky's guidance) and the Elektrostal' Heavy-Machinery Works and are operating successfully.

The effect of continuous tension is particularly important in rolling hollow, thin-walled sections such as tubes. Rolling such a section without tension increases the wall thickness because the metal spreads. Our institute has worked to make continuous tube rolling possible on an industrial scale.

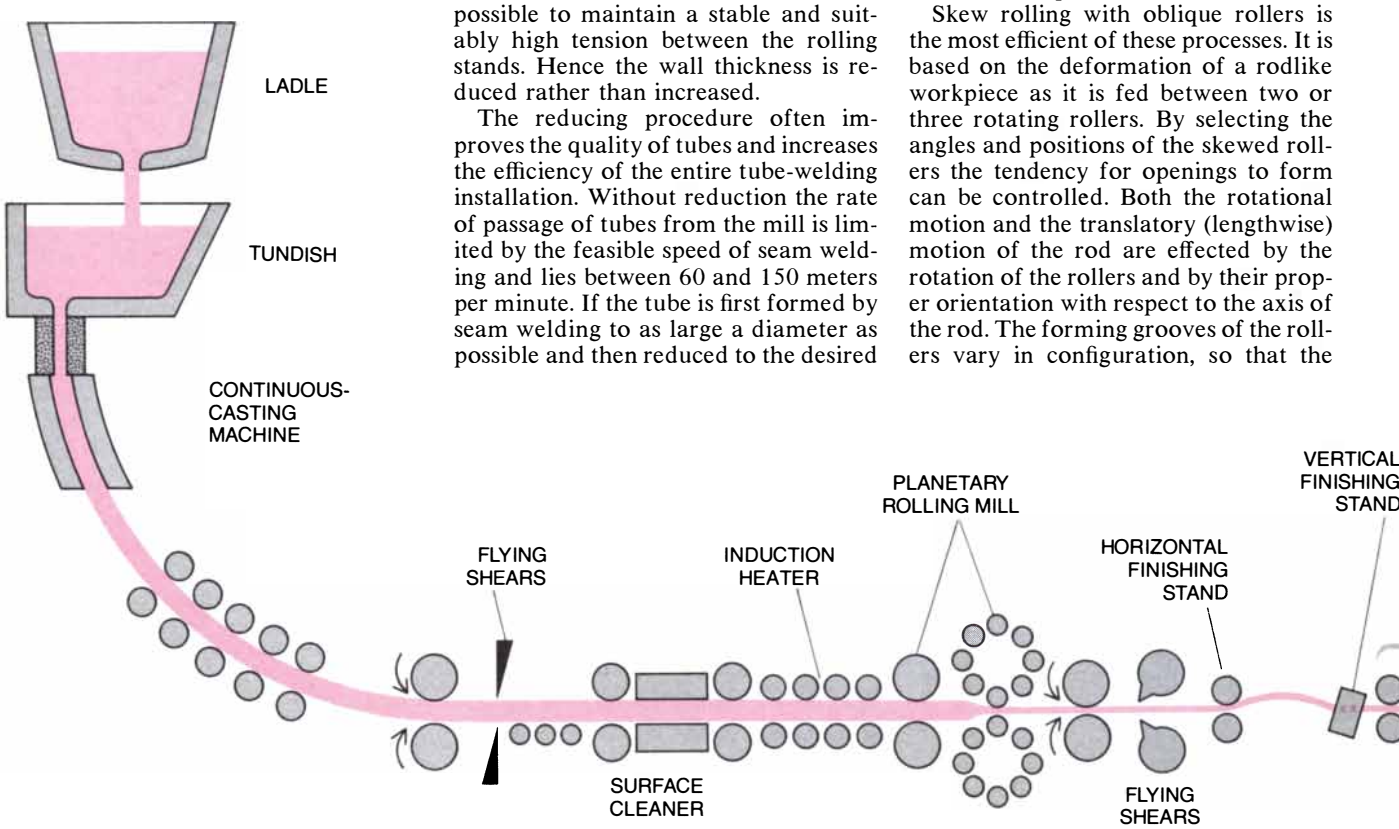
The starting workpieces are coiled strips that are welded together at the ends to form an endless band. The band is fed into a strip accumulator and then shaped continuously into a tube. The seam is welded, and the tube passes to a forming mill to be shaped to the desired diameter. Since the tube is endless, it is possible to maintain a stable and suitably high tension between the rolling stands. Hence the wall thickness is reduced rather than increased.

The reducing procedure often improves the quality of tubes and increases the efficiency of the entire tube-welding installation. Without reduction the rate of passage of tubes from the mill is limited by the feasible speed of seam welding and lies between 60 and 150 meters per minute. If the tube is first formed by seam welding to as large a diameter as possible and then reduced to the desired size, the speed of its exit from the mill can be raised appreciably.

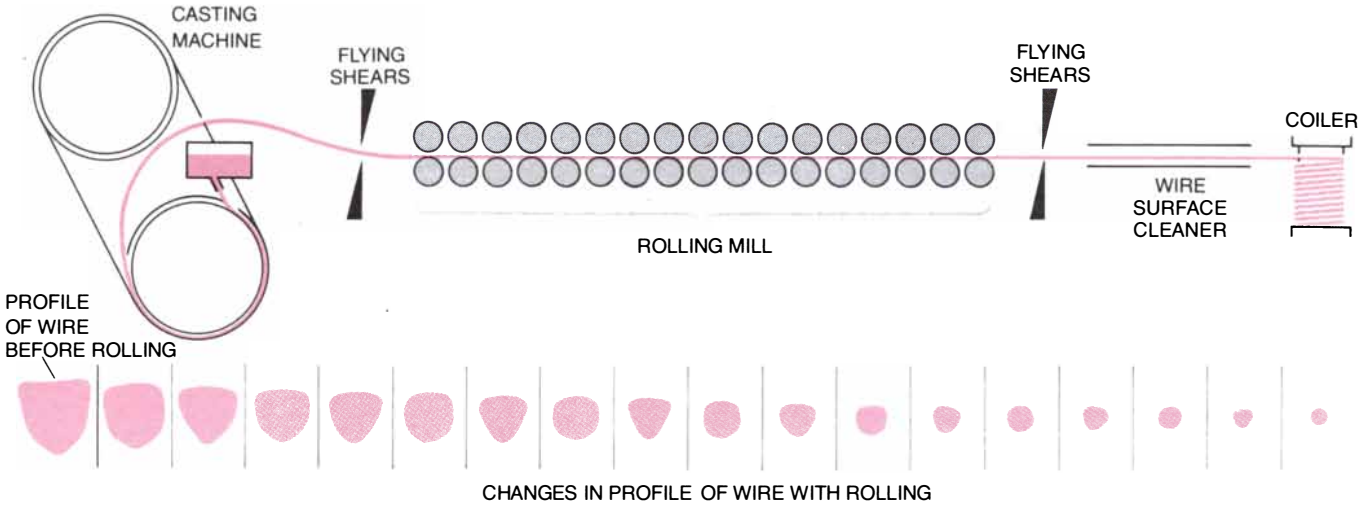
Each of the continuous processes I have described was devised primarily to produce an endless workpiece, which is subsequently cut into blanks to make various finished products by pressing and machining. At the institute we have also directed attention to procedures for the continuous rolling of discrete articles. In particular we have worked on skew, or oblique, rolling and transverse rolling [*see bottom illustration at right on page 128*]. In skew rolling the rollers are at variable, angled positions with respect to the long axis of the workpiece and therefore can form shapes that vary in diameter. Skew rolling is useful in the manufacture of seamless tubes. Transverse rolling (across the long axis of the workpiece) is useful in the manufacture of threaded screws.

Earlier investigators of skew and transverse rolling found that the rolled metal was often spoiled by internal cavities and fractures. It was believed at the time that the procedures were unsuitable for rolling solid articles, and so they were limited to the manufacture of seamless tubes. When skew and transverse rolling were studied in further detail, however, methods were found to prevent internal fractures. As a result it has been possible to devise new continuous processes for the plastic working of various solid parts.

Skew rolling with oblique rollers is the most efficient of these processes. It is based on the deformation of a rodlike workpiece as it is fed between two or three rotating rollers. By selecting the angles and positions of the skewed rollers the tendency for openings to form can be controlled. Both the rotational motion and the translatory (lengthwise) motion of the rod are effected by the rotation of the rollers and by their proper orientation with respect to the axis of the rod. The forming grooves of the rollers vary in configuration, so that the



STEEL ROD can be continuously cast and rolled. The process yields steel rod that is from eight to 12 millimeters in diameter. A single-strand radial caster forms a billet 80 by 60 millimeters in cross section. The billet feeds into a machine that cleans its surface. The flying shears ahead of the cleaner can cut the billet if something goes wrong in the rolling mill. The cleaned billet moves through an induction heater and into a planetary rolling mill, where it is deformed into a shape 20 millimeters square. Flying shears divide the emerging 20-



ROTARY MOLD makes possible the continuous manufacture of copper- or aluminum-wire rods. The continuous casting and rolling installation shown schematically here makes copper rods eight, 10 or 12 millimeters in diameter. The process begins when molten copper (*color*) flows through a tundish into the rotary mold. The emerging billet is fed into a rolling mill. Flying shears cut off the leading end (or the entire billet if the rolling process is disturbed). In the rolling mill the billet passes through 17 roller stands (*vertically paired rollers shown in gray*). The billet's changing profile, starting with a cross section of 1,600 square millimeters, is traced in the bottom drawings.

workpiece is gradually deformed to the desired shape and size. The process is highly efficient and can readily be controlled automatically.

One operation where this process serves particularly well is in forming screw threads. Skew rolling offers an advantage over the standard method of threading by transverse rolling because the workpiece is deformed in a local section moving along its axis. Hence skew rolling can form screws of unlimited length and with almost any pitch.

The output of skew-rolling mills in this application is from .3 meter to 1.2 meters per minute, 10 to 20 times the output of thread-cutting lathes. The saving of metal ranges from 10 to 15 percent. Screws made by skew rolling have more strength and surface hardness than machined screws.

Investigators at the institute have also employed the skew-rolling process for rolling steel balls. Such balls, used notably in bearings, are made by deforming a workpiece so that it resembles a string of beads on a thread. The steel thread between successive balls is broken, and the residue is worn away in a tumbling drum or is ground and polished away in an automatic machine. Carrying the idea a step further, we have designed mills for the continuous automated manufacture of steel balls, the diameters of which range from 25 to 125 millimeters. Skew-rolling mills meet the entire demand for steel balls in the U.S.S.R., which has risen in recent years to approximately 600,000 tons per year. Specialized ball-rolling mills designed jointly by the institute and the Elektrostal' Heavy-Machinery Works are employed at many plants in the U.S.S.R. and are being exported to the U.K., West Germany, Bulgaria and many other countries.

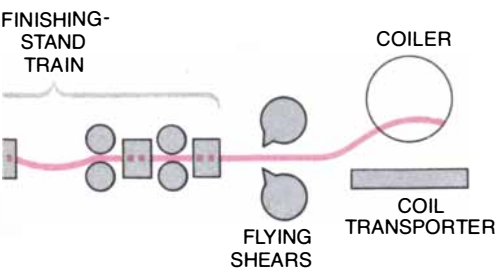
Skew-rolling mills have also proved to be effective for rolling a variety of hollow parts. The first unit in a train of skewed rollers is a piercing mill, where the starting workpiece is pierced and then fed into a shaping mill to be formed into the desired tubular shape. The rolling mill employed to make the races that hold balls functioning as bearings is an example. It can make races for conical roller bearings with an outside diameter of from 65 to 100 millimeters and an inside diameter of from 42 to 75 millimeters. The mill can manufacture from 4,000 to 7,000 pieces per hour.

Still another application of skew rolling developed at the institute serves to turn out parts such as shafts and axles, which vary in diameter along their length. Such parts are conventionally made either by forging and pressing or by removing excess metal with a lathe. Both processes require much labor and involve a large loss of metal. A pressed workpiece must be bigger than the final product to allow for subsequent machining; the shape of a cylindrical rolled blank for machining on a lathe is far from that of the final shaft or axle, which may have necks, shoulders, tapered sections and various other special configurations.

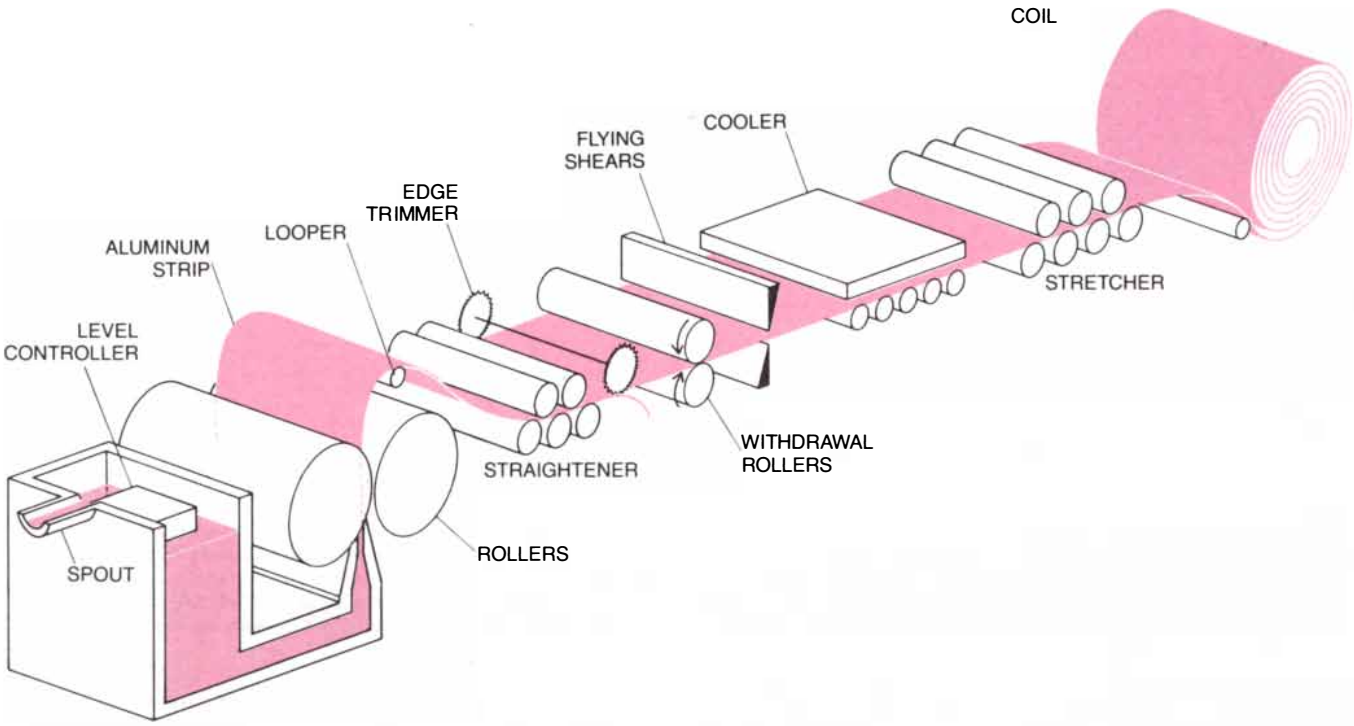
The institute's solution to these problems is a skew-rolling mill based on a continuous motion of the deformation band (the area being shaped) along the rotating workpiece. The main feature of the mill is that as the workpiece (a round rod) moves down the line its reduction can be varied by changing the roller gap according to a specified program. The shaping of the stock is independent of the diameter of the rollers, so that a change from one workpiece to another requires no replacement of rollers.

Among the first rolling mills employing this technique are installations for producing textile spindles with a diameter of up to 12 millimeters, rear half axles for automobiles that are as much as 50 millimeters in diameter, and half axles and shafts for tractors and electric motors that reach a diameter of 120 millimeters. Each part is made with from 20 to 35 percent less metal than is required by conventional methods.

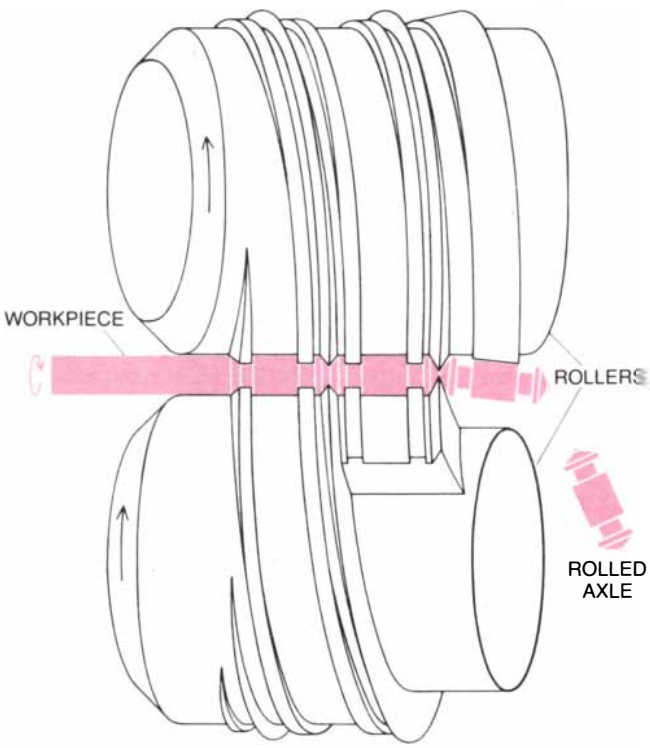
An important engineering achievement is the development, based on the skew-rolling process, of the first rolling mill for the manufacture of blanks for axles of wide-gauge railroad cars and locomotives. The mill produces some 350,000 axles per year. Making axles by continuous rolling instead of by the conventional forging offers three important advantages. First, the axles are rolled to a greater dimensional accuracy, so that there is less need for subsequent machining; this enhancement of precision



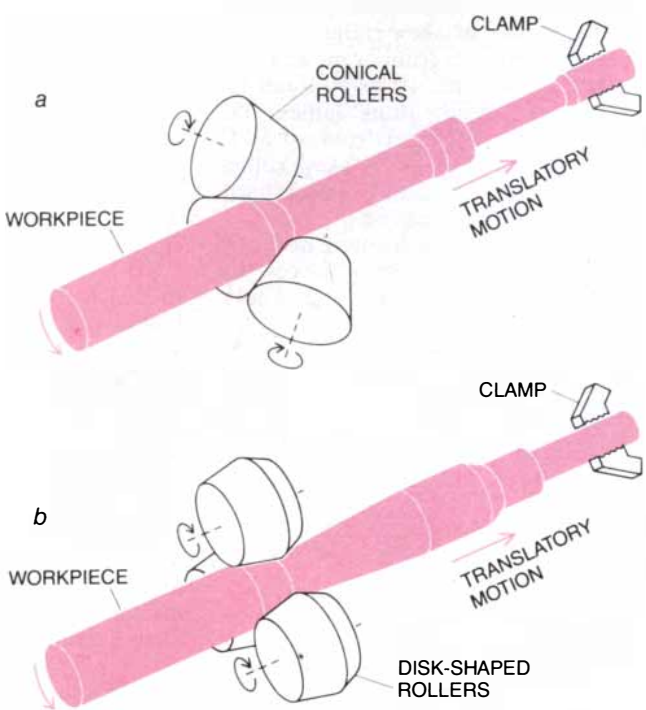
millimeter strand into predetermined lengths, which are fed into the roughing and finishing stands of a continuous rolling mill. In the rolling mill they are fashioned into rod and coiled.



ALUMINUM STRIP is cast and rolled in this continuous process. Molten metal flows from a spout into the tundish of the casting machine; the pressure of its own weight forces the metal into the gap between rotating rollers, the axes of which are in a single horizontal plane; for the process to be stable the level of molten aluminum in the bath near the rollers must be in the same plane. On contact with the rollers, which are cooled by water, the metal solidifies and emerges upward as a strip, which is from three to eight millimeters thick. The strip is bent off and put into a straightener. The edges are trimmed and the continuous strip is cooled, stretched and wound in a coil.



SHORT AXLES of varying diameter are shaped continuously by feeding a round workpiece into the gap between rotating rollers, the surfaces of which have helical grooves. With each turn of the rollers a further portion of the workpiece is gripped by the helical groove and rolled to the final shape. The axles shaped in this way have quite accurate dimensions and do not require any additional finishing.



LARGE AXLES of varying diameter are made by skew rolling, a process in which the axes of the rollers are pitched at varying angles. The rollers are conical (*a*) or disklike (*b*). The workpiece feeds into the gap between rollers and is clamped; the rollers close as the workpiece rotates while moving forward axially. The diameter of a section is set by a hydraulic system that controls the separation of the rollers.

saves from 40 to 60 kilograms of metal per axle. Second, the technology eliminates the need for a press operator (an arduous and disagreeable job). Third, the structure of the metal is improved because the shaping process increases the metal's resistance to fatigue and to severe impacts and shock. The increased strength adds to the service life of the axle and also means it and the vehicle can carry a bigger load.

Further research on skew-rolling processes will undoubtedly focus on the technology of rolling hollow axles. A method proposed at the institute is to use a piercing machine. The idea has been tested industrially for making hollow axles for wide-gauge railroad cars and automobile trailers and shafts for electric motors. The process may increase the saving of metal and reduce the mass of machines. For a vehicle such a reduction is the equivalent of an increase in load-carrying capacity.

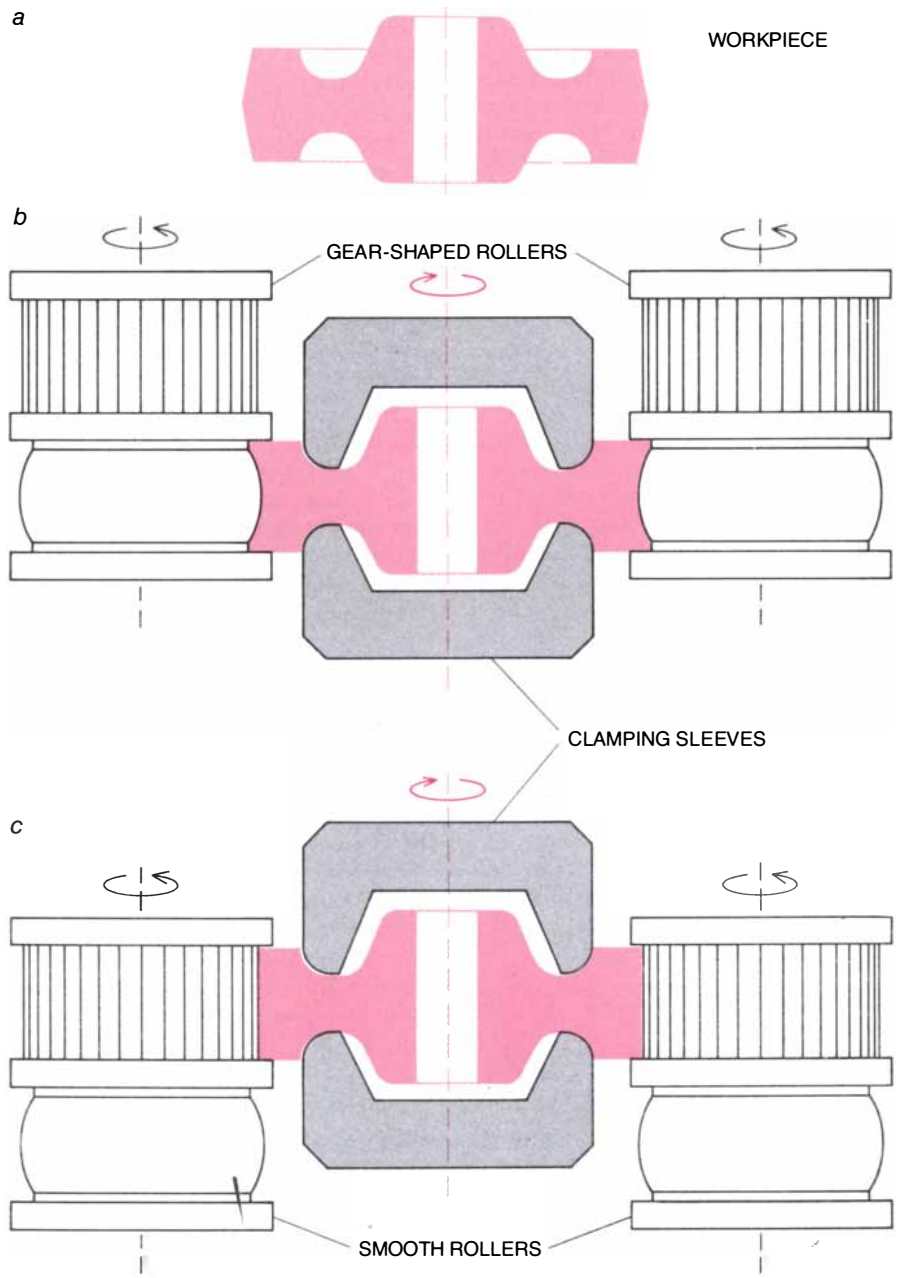
The idea of shaping the teeth of gear wheels by plastic working was put forward nearly 100 years ago. Inventors in a number of countries proposed various ways of doing so; the most effective way was rolling a gear blank in rollers mirroring the shape of the teeth. In this process a cylindrical workpiece is heated to a suitable temperature and put between two rotating, tooth-shaped rollers. The workpiece rotates at a speed corresponding to the gear ratio between the rollers and the gear wheel to be shaped. Then the rollers are moved closer together and the teeth are pressed into the workpiece.

In 1956 the institute and the Chelyabinsky Tractor Works put into operation the first gear-rolling mill that applied on an industrial scale the technique of high-frequency heating of the workpieces (by putting them in an alternating current or a magnetic field). Later a process was developed for rolling certain gear wheels directly from pressed blanks. The blank is rolled first (between smooth rollers) into a disk of the desired diameter and width and then between tooth-shaped rollers to form the teeth. The process replaces two machining operations: shaping the outer face of the wheel on a lathe and cutting the gear teeth on a milling machine to approximately the finished pattern.

The first entirely mechanized and automated gear-rolling mills (designed at the institute under the guidance of A. P. Kuz'min, M. V. Vasil'chikov, M. V. Barbarich and V. F. Mukonin) were manufactured at the Elektrostal' Heavy-Machinery Works. They have been widely adopted. Recent studies have shown that the procedure can be done with enough dimensional accuracy to eliminate the need for a finishing operation in a milling machine.

Many types of machinery have a chain-and-sprocket transmission. To fill the resulting high demand for sprockets the institute applied the principle of continuous rolling to their manufacture. The rolling mills designed for this purpose have proved to be even more efficient than the ones making gear wheels. One reason is that sprocket teeth can be rolled to a higher degree of accuracy than gear teeth, so that no milling has to be done to finish them. Second, the strength of rolled teeth increases substantially as a result of work hardening and spread, which makes it possible to diminish the thickness of the starting workpiece. For example, with a disk eight millimeters thick the width of the sprocket teeth increases to 11 millimeters. As a result the mass of sprockets has decreased by from 20 to 30 percent.

Some 40 rolling mills of this type are in operation at a number of Russian plants, mainly agricultural-machinery works. The total output is more than six million sprockets per year. The achievement indicates the considerable potential of continuous processing for improving industrial efficiency and metallurgical and mechanical engineering in a number of fields.

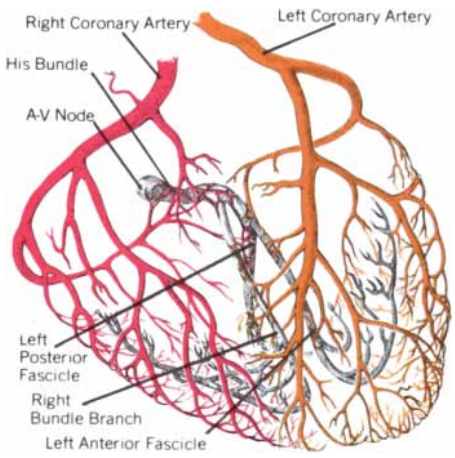


GEAR TEETH formed in the process shown in the photographs on page 121 are rolled in two stages. First, clamping sleeves hold the workpiece (*a*) and place it between two rotating smooth rollers (*b*). When the workpiece is rotating at a predetermined speed, the rollers move into contact with the rim of the wheel, imparting a concave shape so that the teeth can be accurately formed. In the second stage the blank is placed between rotating geared rollers (*c*). As the rollers close toward each other their toothlike protrusions press into the workpiece, forming teeth in it. The process replaces two machining operations: face turning and rough gear milling.

Physician, did you miss any of these significant developments in medical science?

- ***Campylobacter fetus* subsp. *jejuni* is associated with a colitis that can clinically and sigmoidoscopically resemble acute idiopathic ulcerative colitis. Stool cultures are in order for *C. fetus* before beginning nonspecific anti-inflammatory therapy.**
- **Coumarin derivatives cross the placenta. A recent study shows that the consequences for the fetus can be severe. These include embryopathy, stillbirth, and premature delivery.**
- **Nonsteroidal anti-inflammatory drugs may produce a marked reduction in glomerular filtration rate; with termination of the drug, GFR returns to normal.**
- **The first documented incident of indigenous transmission of dengue in the continental United States since 1945 has been reported in Brownsville, Texas.**

IF THESE ITEMS are familiar you must be a prodigiously energetic or prodigiously lucky reader. With 2,000 or more journals published each year, information that significantly affects patient management all too easily slips by. Textbooks are out-of-date before they are published.



Branches of the right and left coronary arteries supply blood to the A-V node and intraventricular conduction system.

SCIENTIFIC AMERICAN *Medicine* is lucidly illustrated with drawings and photographs. Some examples are seen here and on the facing page.

SCIENTIFIC AMERICAN *Medicine* is the busy clinician's answer to this problem.

Because its authors update SCIENTIFIC AMERICAN *Medicine* every month, it is always current. Because the new information appears in a single source, it is there when you need it.

This 2,500-page, innovative union of publishing and electronic technology is the work of leading scholar-practitioners from Harvard and Stanford. The editors are Edward Rubenstein, M.D., F.A.C.P., and Daniel Federman, M.D., F.A.C.P.

Each month as authors update their contributions, revisions are entered on the magnetic tape on which the text and index are stored. The tape drives high-speed phototypesetting equipment so that subscribers receive about eight new chapters and a new index every four weeks; a bulletin highlights new developments.

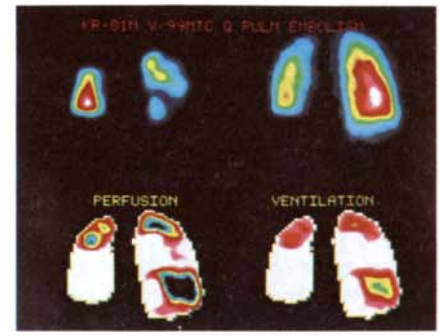
New material replaces old material in the living text, so that the information is there — up-to-date, at your fingertips.

A CME program of eight patient management problems offered over a 12-month period is available at no extra cost. As an organization accredited for continuing medical education, the Stanford University School of Medicine designates this continuing medical education activity as meeting the criteria for 32 credit hours in Category 1 for Educational Materials for the Physician's Recognition Award of the American Medical Association, provided it has been completed according to instructions. This program has been reviewed and is acceptable for 32 prescribed hours by the American Academy of Family Physicians. This program has been approved by the American College of Emergency Physicians for 32 hours of ACEP Category 1 credit.

Trial Offer

We invite you to try SCIENTIFIC AMERICAN *Medicine* — for two months at no cost. Send us the coupon and you will receive the two-volume text and two monthly updates. You may also take a CME test for credit. At the end of 60 days, if you decide to continue the subscription, we will bill you for $235 for the full 12 months (renewal is currently $170); otherwise return the two volumes.

Please mail the coupon today and let us take the hassle out of keeping up.



Computerized scintigraphy reveals pulmonary thromboembolism.

Abdominal computed tomogram reveals large renal carcinoma replacing part of right kidney.

THE DISTINGUISHED AUTHORS AND THE FIFTEEN SECTIONS OF SCIENTIFIC AMERICAN ***MEDICINE***

1. Cardiovascular Medicine

Edgar Haber, M.D., Harvard Medical School and Massachusetts General Hospital

E. William Hancock, M.D., F.A.C.P., Stanford University School of Medicine

Roman W. DeSanctis, M.D., Harvard Medical School and Massachusetts General Hospital

Adolph M. Hutter, Jr., M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Eve Elizabeth Slater, M.D., Harvard Medical School and Massachusetts General Hospital

2. Dermatology

Eugene M. Farber, M.D., Stanford University School of Medicine

Elizabeth A. Abel, M.D., Stanford University School of Medicine

3. Endocrinology

Daniel D. Federman, M.D., F.A.C.P., Harvard Medical School

4. Gastroenterology

Gary M. Gray, M.D., Stanford University School of Medicine

Peter B. Gregory, M.D., Stanford University School of Medicine

John Austin Collins, M.D., Stanford University School of Medicine

Douglas Wilmore, M.D., Harvard Medical School and Brigham and Women's Hospital

5. Hematology

Stanley L. Schrier, M.D., Stanford University School of Medicine

6. Immunology

John David, M.D., Harvard Medical School and Harvard School of Public Health

7. Infectious Disease

Thomas C. Merigan, M.D., Stanford University School of Medicine

Morton N. Swartz, M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Cyrus C. Hopkins, M.D., Harvard Medical School and Massachusetts General Hospital

Adolf W. Karchmer, M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Robert H. Rubin, M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Harvey B. Simon, M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Peter F. Weller, M.D., Harvard Medical School; Beth Israel Hospital, Boston; and Brigham and Women's Hospital

8. Intensive and Emergency Care

Edward Rubenstein, M.D., F.A.C.P., Stanford University School of Medicine

9. Metabolism

George F. Cahill, Jr., M.D., Harvard Medical School, Howard Hughes Medical Institute; and Brigham and Women's Hospital

10. Nephrology

Roy H. Maffly, M.D., Stanford University School of Medicine and Palo Alto Veterans Administration Medical Center

William Bennett, M.D., Health Sciences University, Portland, Oregon

11. Neurology

Robert W. P. Cutler, M.D., Stanford University School of Medicine

12. Oncology

Saul A. Rosenberg, M.D., F.A.C.P., Stanford University School of Medicine

13. Psychiatry

Ned H. Cassem, M.D., Harvard Medical School and Massachusetts General Hospital

14. Respiratory Medicine

Eugene D. Robin, M.D., F.A.C.P., Stanford University School of Medicine

15. Rheumatology

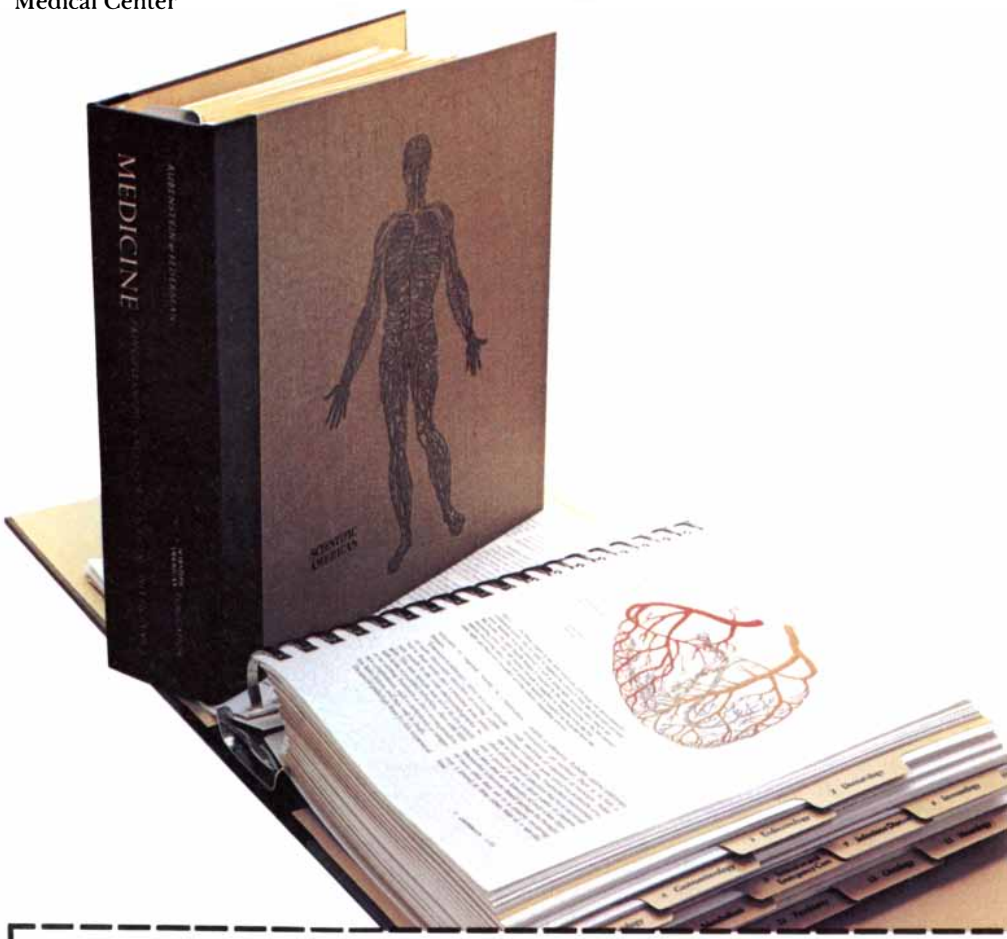
Stephen M. Krane, M.D., Harvard Medical School and Massachusetts General Hospital

Dwight R. Robinson, M.D., Harvard Medical School and Massachusetts General Hospital

Andrei Calin, M.D., M.A., M.R.C.P., Stanford University School of Medicine and Palo Alto Veterans Administration Medical Center

Order by Phone

You can order SCIENTIFIC AMERICAN *Medicine* by telephone. **Please call this toll-free number: 1-800-345-8112 (in Pennsylvania, call 1-800-662-2444);** you will be billed after your subscription begins. Toll-free calls are acceptable only for orders placed in the continental United States.



SCIENTIFIC AMERICAN MEDICINE

415 Madison Avenue, New York, N.Y. 10017

Please enroll me as a subscriber to SCIENTIFIC AMERICAN *Medicine*. On receipt of this coupon you will send me the advanced two-volume text described in your announcement and update it regularly by sending me new monthly subsections. I understand that the price of $235 for first year of service is tax deductible, as is current renewal at $170. If I am not entirely satisfied, I may cancel at any time during the first 60 days, returning all materials for a *complete refund.*

☐ Please enter my subscription for SCIENTIFIC AMERICAN *Medicine*
☐ I shall also enroll in the CME program
☐ I enclose a check made out to SCIENTIFIC AMERICAN *Medicine* for $235*
☐ Bill me ☐ VISA ☐ MasterCard

Expiration Date_____ Account Number_____
*Please add sales tax for California, Illinois, Massachusetts, Michigan and New York

Name_____

MD Specialty_____

Address_____

City_____ State_____ Zip_____

Signature_____

Please allow six to eight weeks for delivery. All payments must be in U.S. dollars. CME is available outside the U.S. and its possessed unless airmail delivery is prepaid.

7R

The Crystal Palace

Designed by a duke's gardener and built in 17 weeks, its modular construction, prefabricated elements and glass walls anticipated modern aesthetic and method

by Folke T. Kihlstedt

The Crystal Palace, a product of Victorian England, was one of the most influential buildings ever erected. Innovative in structure, completely new in its function, unusual in form and significant in the associations it embodied, it takes its place with a handful of other preeminent buildings such as the Pantheon, Hagia Sophia and Abbot Suger's St. Denis. As with these earlier constructions, the extraordinary functional demands made on the Crystal Palace stimulated a design that refined and extended the structural practices of the time, resulting in an architecture novel in its form and aesthetic.

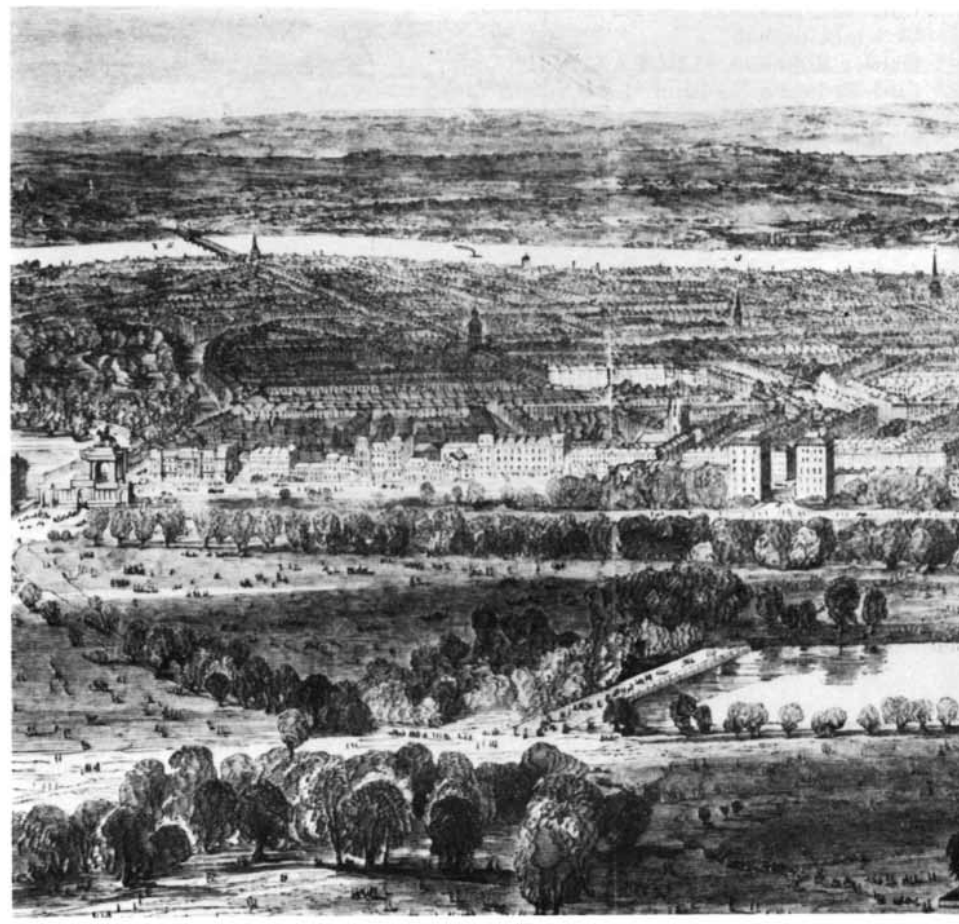
Most Victorians, however, saw the Crystal Palace not as architecture but as construction; to their eye it was styleless and lacked decoration. They were accustomed to an architecture of heavy, monumental forms in wood or masonry, which emphasized the opacity of wall surfaces.

Today architects and art historians admire the Crystal Palace for just those spare qualities, which seem to anticipate the "factory aesthetic" of the early 20th century and the formal clarity of International Style design. Ready contemporary appreciation of such formal aspects has obscured the fact that the structural innovations of the Crystal Palace heralded modern practice even more fundamentally than did its form.

The Crystal Palace was erected in London's Hyde Park in 1851 to house the first world's fair, The Great Exhibition of the Works of Industry of All Nations. It was dismantled in 1852 and reerected on the eastern slopes of Sydenham Hill, where it stood from 1854 until an accidental fire razed it late in 1936. There it had served four generations as a center for entertainment, displays of art and science, concerts and dining.

The original Hyde Park building enclosed a vast, partition-free interior space in which all the technological products of civilization were displayed. The event it housed was conceived by Henry Cole, a high civil servant and indefatigable champion of free trade and good design; it was endorsed by Prince Albert, and it was sponsored by the Royal Society of Arts. The designer of the building itself, Joseph Paxton, was a farmer's son who since 1826 had served as superintendent of gardens for the Duke of Devonshire. This position would eventually bring him into contact with Cole, Prince Albert and the Royal Society of Arts. Paxton had little education but he possessed a genius for organization. His design for the Crystal Palace would earn him knighthood and a seat in Parliament.

On March 13, 1850, scarcely a year before the opening date, the Building Commission announced a competition for the building to house the Great Exhibition. The building was to be temporary in nature, economical of materials and labor, simple in its arrangement, capable of rapid erection, dismantling and expansion, illuminated entirely from the roof, built of fire-resistant materials and



CRYSTAL PALACE originally stood in Hyde Park, where, in 1851, it housed the Great Exhibition of the Works of Industry of All Nations, the first world's fair. Designed by Joseph Paxton, gardener to the Duke of Devonshire, the structure presaged the engineering-based aesthetic of contemporary architecture, the structural innovations that made this style possible

erected over an 18-acre site, generally to a height of a single story.

Of 245 designs received by the April deadline, none was quite suitable. The commission proceeded to design its own building. It was a complicated brick structure with a large central dome of iron. This structure too failed to meet the requirements, yet it was put out to bid. At this point, with the public and Parliament in an uproar, Paxton entered the fray. On June 7 he received assurances that a new design would be considered if it accompanied a bid for the Building Commission plan. He, with his staff at Chatsworth (the Duke of Devonshire's seat) and a railway civil engineer, formulated in eight days a building of original—indeed, revolutionary—design that fulfilled all the requirements stated by the Building Commission.

Paxton also played his cards shrewdly. He won over influential members of Parliament, the Building Commission and even the public to his proposal. On July 26, 1850, the commission formally accepted Paxton's plan along with the construction tenders of the contracting firm of Fox and Henderson.

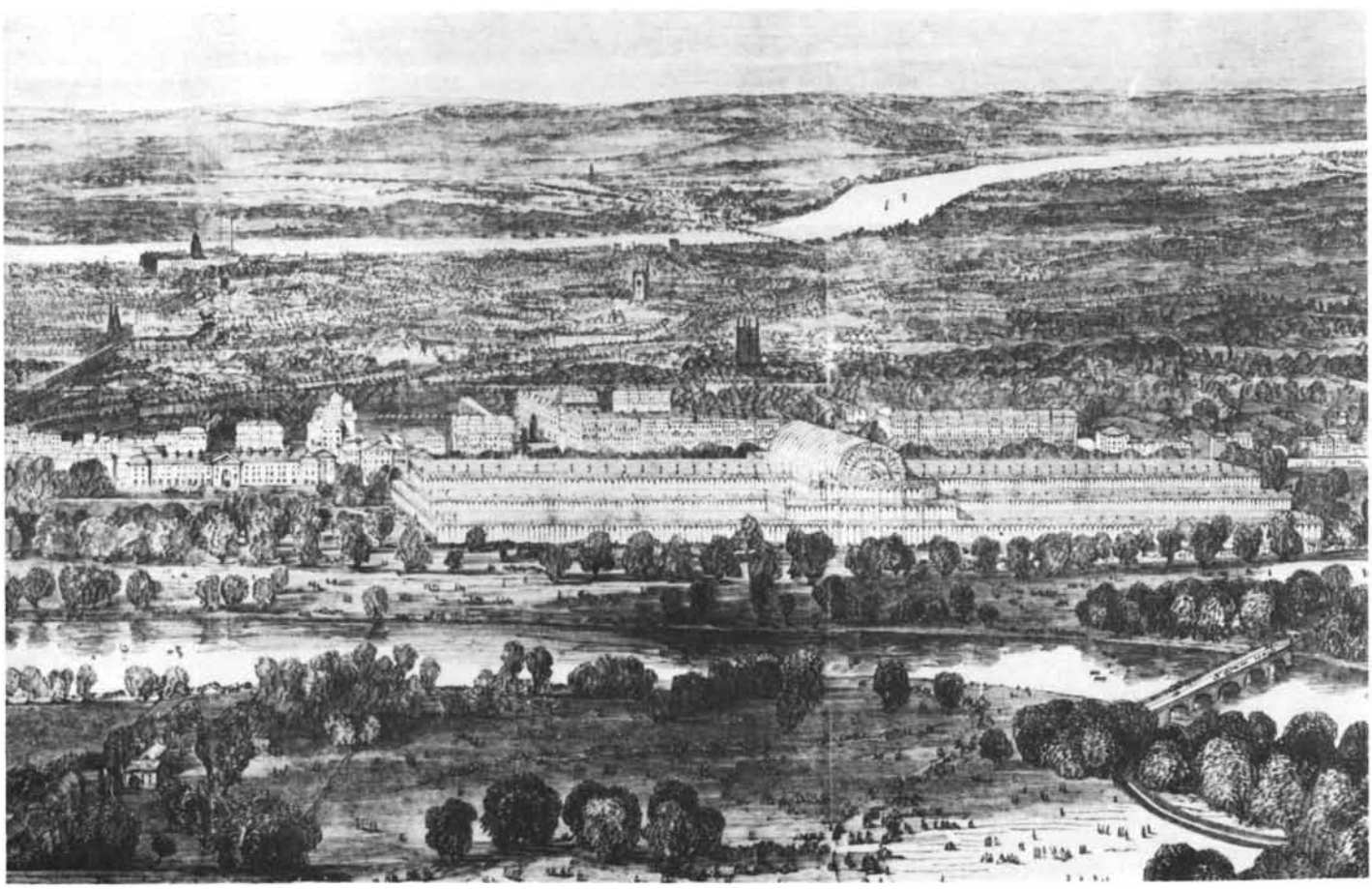
In the ensuing weeks the contractors refined the structure, determined the strength and form of every member, made mock-ups for testing and began erection. The Crystal Palace was built in the unheard-of time of 39 weeks. The Great Exhibition opened on schedule on May 1, 1851, and the building that housed it was seen by many as being the most important object on display. Its effect was totally new. Viewed from outside, thin iron columns carrying spandrel arches [*see bottom illustration on page 136*] and framing an infill of glass seemed to form an almost endless repetition of eight-foot bays. It was a marvelous constructional solution to a unique set of requirements.

In the instructions to the design-competition entrants the Building Commission had recommended that the structure have "some striking feature to exemplify the present state of the science of construction" in England. By virtue of the commission's own proposed design, which boasted an ungainly cast-iron dome 200 feet in diameter, it clearly intended that striking feature to be monumental.

The Crystal Palace possessed no truly monumental feature; it was itself monumental in a purely quantitative sense. It covered 772,824 square feet (about 19 acres) in plan. It was 1,848 feet long by 408 feet wide and had an addition on the north side measuring 936 by 48 feet. Its longitudinal central aisle, the "main avenue," was 72 by 66 feet high, and its vaulted transept was 72 by 108 feet high. Nine hundred thousand square feet of sheet glass, which would have amounted to a third of England's total glass production in 1840, formed its walls and roof. It was supported by 3,300 cast-iron columns, 2,224 principal girders and 24 miles of main gutter; 205 miles of wood sash bar held the glass roof panels in place. At a cost of about a penny per cubic foot, the Crystal Palace enclosed a volume of 33 million cubic feet.

As overwhelming as these dimensions may have been, the building itself was astoundingly simple. It basically consisted of a series of hollow cast-iron columns joined by trussed girders that supported a planar roof made of glass panes in a pleated, ridge-and-furrow configuration. In other words, the roof consisted of peaks and valleys alternating at four-foot intervals.

The components of this system included base plates for the columns, pro-



and many of the methods of construction through which it is realized. The Crystal Palace, rendered for the *Illustrated London News* of June 14, 1851, is shown as it appeared from the north, looking across the Serpentine. In the background is the Thames, crossed to the east (*far left*) by Vauxhall Bridge and to the west by Battersea Bridge. Behind the Crystal Palace are the Knightsbridge Cavalry Barracks, and farther to the east is Hyde Park Corner. Also shown are several churches, including Trinity (*center*), St. Peter's (*left*) and St. Luke's (*right*).

jecting about four inches above the floor; the columns, rising approximately 18½ feet above the plates; connecting pieces, extending a bit less than three and a half feet to provide the junction of column and girder, and six-inch-deep gutters at eight-foot centers supporting sash bars that held in place on an incline of 2.5 : 1 sheets of glass each measuring 49 by 10 inches.

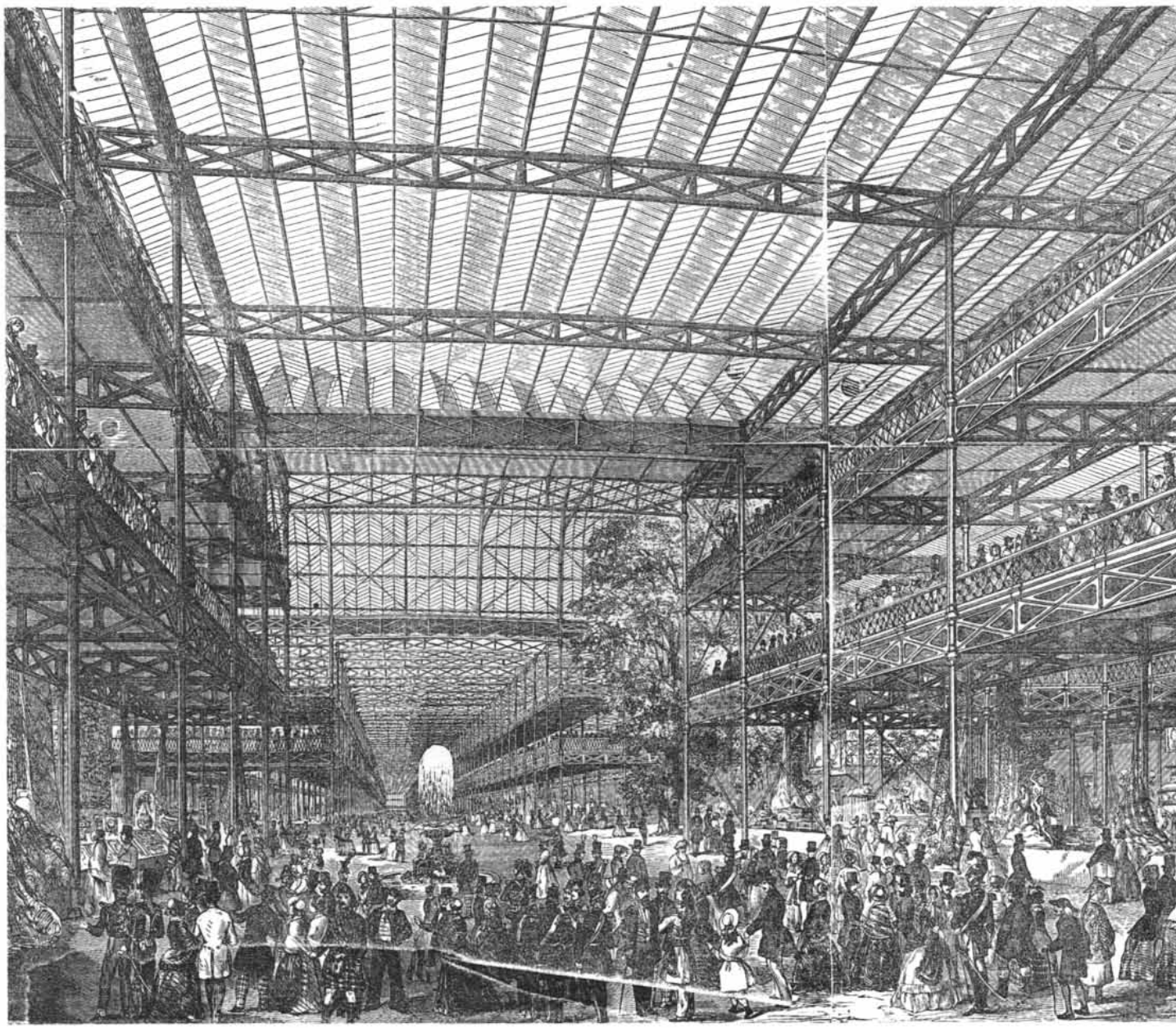
The elevation stepped up twice to the roof over the main avenue, but the only change in components was to a column approximately two feet shorter than the ground-floor column. As the longitudinal aisles and main avenue broadened, the trussed girders lengthened in a simple module from 24 to 48 to 72 feet.

The only complexities in the form of this simple post-and-lintel structure were the vaulted transept and stiffening members, which took the form of diagonally crossed wrought-iron rods [*see top illustration on page 136*]. These rods, in diameters of seven-eighths and three-quarters of an inch, were placed respectively between the columns of the outer walls and between the main ribs of the transept vault. Held in tension by an iron ring at their center, they stiffened the building frame, which had no internal walls for this function.

Filigreelike, yet visible from interior and exterior as it crossed the transparent bays of glass, this diagonal bracing presented a new aesthetic experience. Even today one must become accustomed to the interruption of the window area by diagonal bracing, as in Chicago's John Hancock Center.

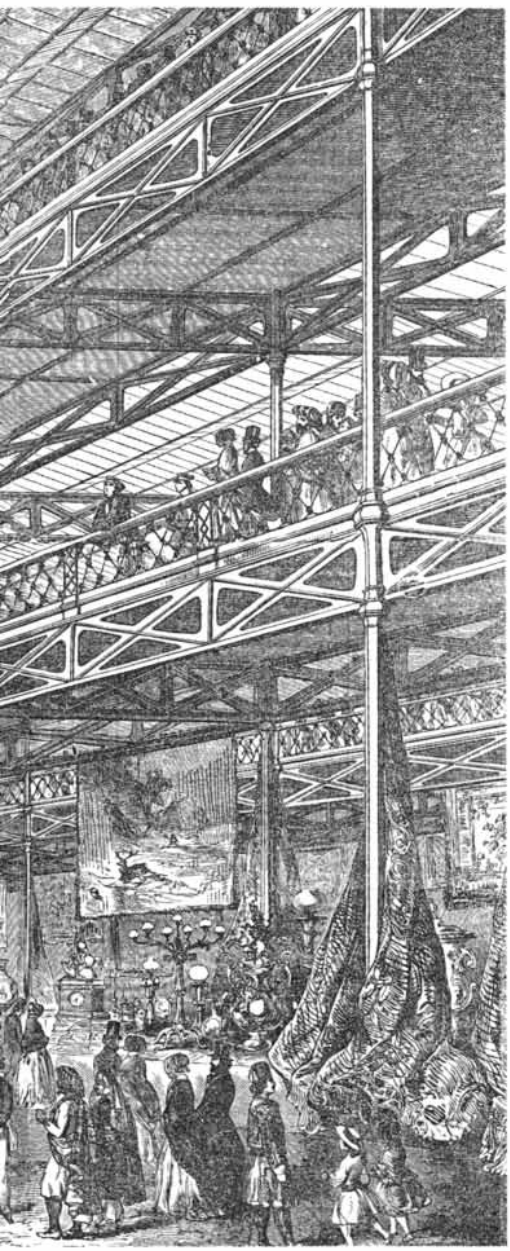
The Crystal Palace was, as a contemporary said, "a remarkable specimen of the constructive skill of this country." It certainly anticipated numerous structural features common to 20th-century architectural design, features that were not to be found in combination in any other 19th-century building. Thus it became the world's first really large, free-standing iron-frame building; the first building with what today would be called glass curtain walls; the first building to use a system of portal bracing to counteract the lateral forces of the wind, and the first large building to be made from prefabricated modular units.

The Crystal Palace could not have



MAIN HALL of the Crystal Palace is seen in this engraving from SCIENTIFIC AMERICAN for March 19, 1851. Some of the structural elements shown are girders (under the roof) and diagonal cross braces (inside the transept arch). All these elements were clearly visible, constituting a major departure from the Victorian architectural canon of hiding structural elements. Through the ridge-and-furrow roof

been built before 1845, the year Parliament removed a prohibitive excise duty on glass. Glass was essential to the functioning of the building as a display place. Into its halls, roof and vault went 400 tons of sheet glass, all made by Chance Brothers and Co. of Birmingham. Paxton demanded, and got, the largest sheet possible. Each sheet was made by blowing the glass into a hollow cylinder, cutting it longitudinally and allowing it to open up and flatten out in a kiln. These 49-inch panes, when they were arranged on Paxton's ridge-and-furrow incline, determined the eight-foot distance between the valleys (or furrows), thereby conforming with and helping to determine the 24-foot module of the entire structure.



(itself a structural innovation) can be seen the top of the vaulted central transept, a similar ridge-and-furrow construction but curved.

The idea for the ridge-and-furrow configuration of the roof belongs to John Claudius Loudon, but Paxton seems to have been the first to put it to actual use, initially in greenhouses. He refined the concept when he remodeled a greenhouse roof in 1832. He then applied the system to the arched vaults of the Great Stove at Chatsworth, a conservatory he built between 1836 and 1841 that was broader than any of the railway stations in existence at the time. Finally, as a prelude to the planar (not peaked) roof of the Crystal Palace, he designed in 1849 a similarly roofed glass box to house the Victoria Regia Lily, also at Chatsworth. Imported from British Guiana, the lily had refused to thrive until housed in the special structure, in which the water currents the plants required were reproduced.

In 1832 Paxton had championed this configuration because it would allow the morning and evening sun to pass through the glass almost perpendicularly. This, he maintained, was of great importance in the propagating and forcing of fruit. Here was the master gardener, designing empirically from closely observed experience.

Nineteen years later Paxton was designing for a different purpose. His use of the ridge-and-furrow configuration clearly anticipated what today we would call folded-plate or pleated construction. In a folded plate, used almost exclusively in reinforced concrete structures, the creased form adds stiffness. Also the loads would maximize at the valley and be carried longitudinally, with the plate form acting like a beam.

Paxton, as we shall see later, used his valley gutter precisely in this manner—as a beam. His wood sash bars, which connected gutter to ridge, probably could not effectively transform the entire roof into a series of beams as in a true folded-plate system. Nevertheless, Paxton's adherence to this system when it no longer served in what he originally saw as its capacity to admit light and retain heat demonstrates his realization that the folded planar roof afforded extra stiffness. It also suggests that Paxton was among the first modern designers to understand the structural virtues of folded-plate construction.

The same component glass sheet of the ridge-and-furrow roof was also used to form the curtain-wall exterior of the Crystal Palace. Each pane was secured by wood sills and thin mullions (slender vertical supports) framed by a system of iron columns at 24-foot centers with two intermediate columns of wood. Thus Paxton had designed the archetypal glass box in which to display the works of industry of all nations. The side walls of this "box" had sills a mere eight inches thick, a measurement that architect Matthew Digby Wyatt contrasted with the 14-foot-thick walls of St. Paul's Cathedral.

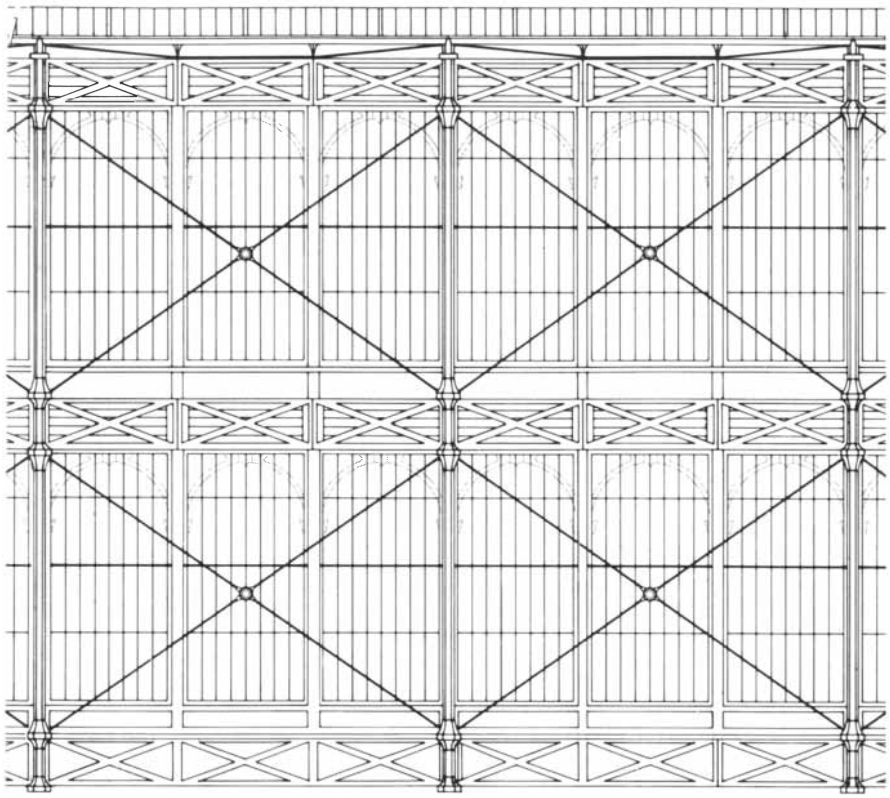
An even more critical dimension was that of the sash frames above these sills. These were only two and a half inches thick, which makes them equivalent to any curtain-wall building of the 20th century. Here is a wall that is chiefly a transparent skin to keep out the elements and let in the light. It is the essence of Mies van der Rohe's "skin and bones" architecture, envisioned in the early 1920's and realized 30 years later in New York in his Seagram Building and in its neighbor, Skidmore, Owings & Merrill's Lever House.

The true curtain wall, however, had to wait for the steel frame before Paxton's contribution could be appreciated fully. That appreciation came not with Mies's skyscraper proposals of the 1920's but with Willis Polk's Halladie Building in San Francisco. Completed in February, 1918, it boasts the first true curtain wall in a large urban structure. Its glass front was hung from floor slabs that were cantilevered three feet in front of the reinforced concrete columns [*see top illustration on page 142*]. Thus it had no supporting member on the plane of the exterior all-glass wall. Polk honored his debt to Paxton in an early perspective study for the Halladie Building, which he captioned "The Crystal Palace."

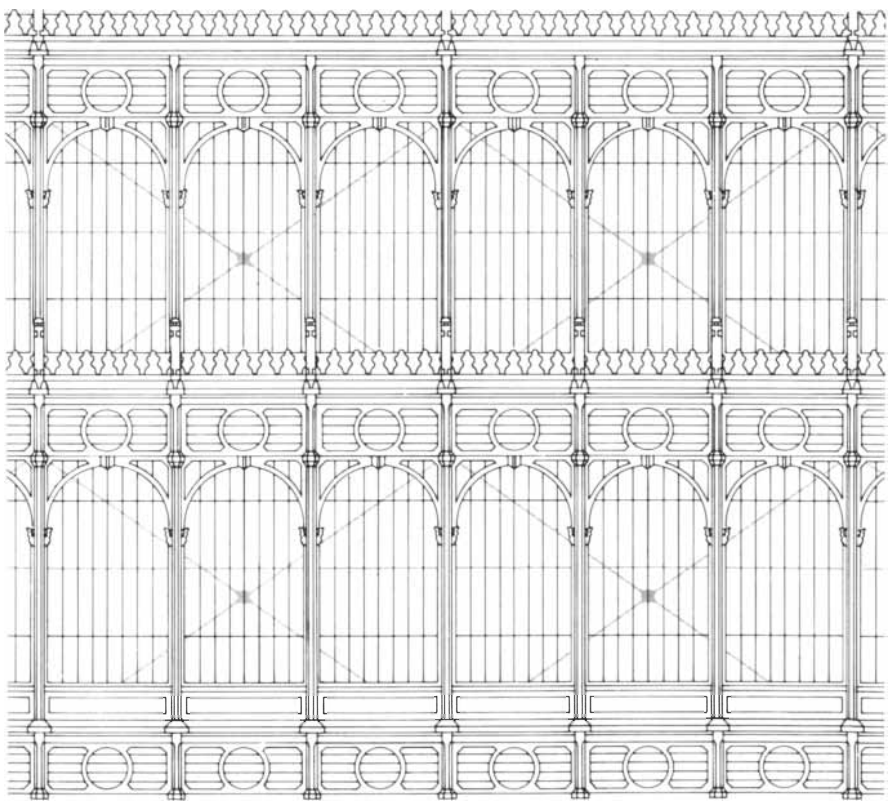
The Crystal Palace exterior walls did have supporting members in the plane of the glass, namely the exterior columns and diagonal bracing. Paxton, however, had developed a true curtain-wall system, albeit on a smaller scale in the Victoria Regia Lily House. The Lily House was a rectangular box, 60 by 46 feet, with flat ridge-and-furrow roofing carried on horizontal girders, which in turn rested on two rows of internal columns. Thus, on this manageable scale, he actually did "hang" a glass curtain wall from the girders that cantilevered beyond the supporting columns, producing a transparent outer skin.

An essential feature of Paxton's design and one that allowed him to have such thin exterior walls was the development of a type of portal bracing that provided a rigid connection between the vertical columns and the horizontal trusses. In essence a portal brace ensures a rigid connection between a beam and two columns. Consequently the rectangular frame created by these components cannot be forced into a parallelogram configuration by a horizontal force such as that exerted by the wind. The portal brace counteracts the rotation of the ends of a beam; the term derives from its use in the end, or portal, frames of a truss railroad bridge.

A portal brace also allows the depth of the beam to be much less at the center of its span than it would need to be if it simply rested on top of two columns



WROUGHT-IRON RODS provided a rigid support for the exterior walls of the Crystal Palace, which had no internal walls to stiffen it. Visible from inside and out (the interior view is shown here), these cross braces added to the building's strikingly contemporary appearance.



CURTAIN WALLS were another structural innovation introduced by Paxton and his colleagues. The 49-by-10-inch panes of glass were set behind thin spandrel (framing) arches. These panes were used throughout the walls and roof, and set the fundamental dimensions of the modular units that made up the structure. Some 400 tons of these panes were used. The Crystal Palace's transparent skin was a harbinger of contemporary glass-sheathed structures.

that were fixed in some other way to their vertical plumb position. In the case of the Crystal Palace, where the columns are bolted to base pieces that rise from point foundations, some as shallow as one foot deep, the portal bracing and the girders also served to keep the columns perpendicular.

The essential feature of Paxton's portal brace was the vertical cast-iron connecting piece above the column (designed by the contractor, Charles Fox). This connecting piece had flanged projections at its top and bottom. The flared ends of the trussed girders fitted into these projections and the connections were made rigid simply by driving wrought-iron or oak keys into corresponding grooves in the flanges and truss ends [*see illustration on page 139*]. In this way the top and bottom of the truss were connected simply and rapidly to the column to give it lateral stiffness. As Carl W. Condit of Northwestern University has pointed out, true portal bracing, in which a girder is riveted to a column throughout its depth, was first used in the Royal Navy Boat Store of 1858–60 at Sheerness [see "The Wind Bracing of Buildings," by Carl W. Condit; SCIENTIFIC AMERICAN, February, 1974].

The concept for the portal brace, as practiced in the iron-frame Crystal Palace and later in the Royal Navy Boat Store, was derived from timber construction. The big railway sheds that were being erected out of iron at the time were braced by large-scale spandrel brackets or by arched girders, both of which developed rising roofs of vertical profile. In contrast to this very different system for bracing the railway sheds, the unassuming portal bracing developed for the Crystal Palace made it the first planar-roofed, rectilinear iron building in the world.

So novel was this approach that some engineering contemporaries of Paxton and Fox did not believe the framing system could be rigid or reliable. One established engineer questioned the ability of the Crystal Palace to resist wind loads, and other engineers questioned Paxton's ingenious connecting pieces in particular as being "the weakest point of the entire structure." Even Richard Turner, an unsuccessful competitor for this building and a well-known designer of railway sheds and wrought-iron structures, clearly did not comprehend the virtues of these connectors; he singled out the column bases as too weak and predicted they would be the points where the building would be overturned.

The true rigid-joint portal frame, riveted throughout its depth, is essentially a recent form associated with the tall, steel-frame building. Its roots lie in the prophetic design of the Crystal Palace connectors. Those connecting pieces

were the basic structural elements of the Crystal Palace, as they allowed the side cladding to be of glass, which otherwise could offer no lateral stability. This general structural design finds a modern parallel in almost every steel building of trabeated construction (a structure involving horizontal beams on vertical posts) in which the glass wall is hung beyond the supporting posts.

The horizontal member of the portal frame was an openwork truss. A truss is like a deep beam but is made of individual members consisting of top and bottom chords (horizontal linear members) connected and triangulated for rigidity by one or more diagonals; the chords can be connected by vertical posts as well. Such an element can span a much greater distance than its individual members ever could if they were in the form of a solid beam.

A truss uses considerably less material than a beam of equivalent strength and can also be fabricated from small-scale components. For this reason trusses best met the demand for the approximately 2,300 horizontal framing members needed in such a short time by the builders of the Crystal Palace.

The Crystal Palace stands at an important turning point in the development of the structural mastery of metal trusses. Examples of iron roof trusses can be found in structures as early as the Soho Foundry of 1810 and the Euston Station shed of 1837. Euston Station had a roof truss made completely of wrought iron rather than a mere combination of wrought-iron tie-rods with cast-iron or timber members. This use of wrought iron was a timely response to the experiments conducted by Eaton Hodgkinson and William Fairbairn between 1826 and 1830 that demonstrated the superiority of wrought iron for any members subjected to tension.

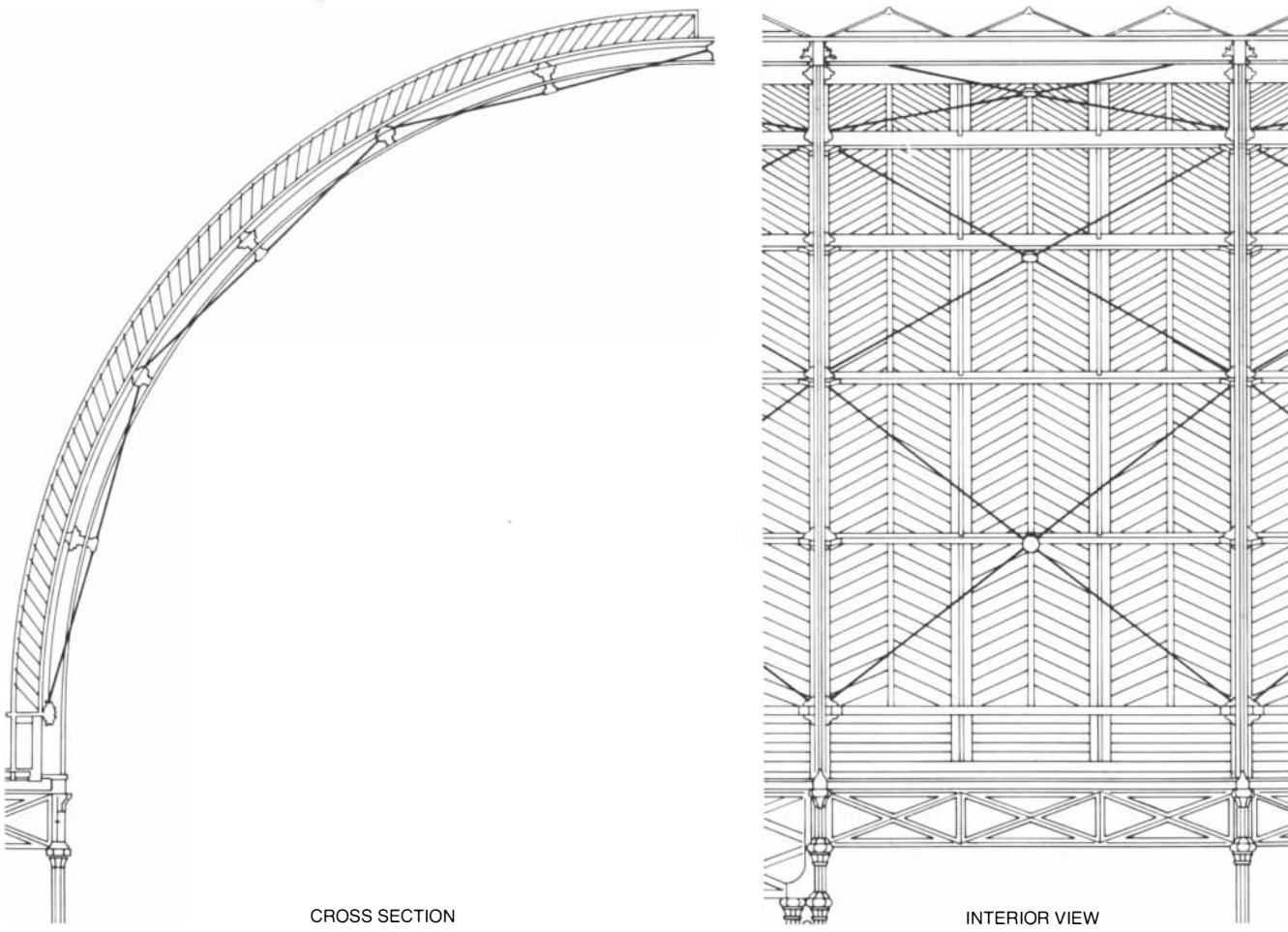
The early examples of trusses, found in Euston Station and other railroad sheds followed the contour of the roof, in either a gabled or an arched form. Since the top and bottom chords of the trusses were not parallel and horizontal, they acted principally as arches rather than as beams. Many bridge trusses, however, were designed as beams and so were a closer prototype for the Crystal Palace trusses.

The first iron-truss bridges were designed for the railways beginning in 1840. Most of these bridge trusses combined cast and wrought iron. The cast iron was often used for the top chord, which was in compression; the wrought iron was riveted together out of long flat stock to form the bottom chord, which was in tension. Depending on the particular truss design, the posts and diagonals were either cast or wrought iron.

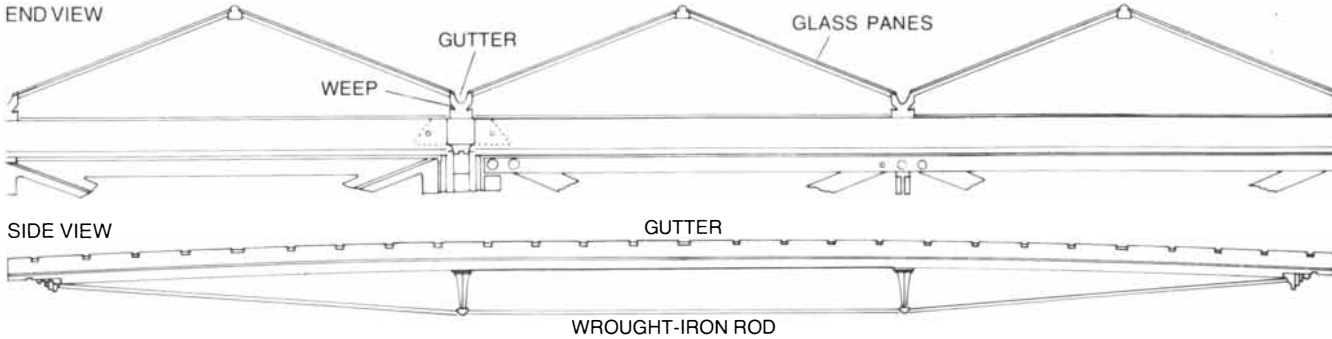
Paxton combined Fox and Henderson's expertise in the design of trusses for railway bridges with his own development of the planar roof, derived from his greenhouses. The result was a rarity for the 19th century: a major structure with a planar roof, supported by trusses inspired unabashedly by railway bridge design. Nineteenth-century buildings seldom had planar, or flat, roofs; if a wide span was required, a bridge-truss configuration of parallel chords would not have been used.

This bold approach produced an aesthetic dividend (or shock). Not until well into the 20th century, when the space frame made its appearance, would it be generally acceptable to leave a truss structure exposed to view on the interior of a building.

The basic truss in the Crystal Palace was three feet deep. There were posts every eight feet and crossed diagonal



ARCHED VAULT of the central transept appears in cross section (*left*) and in an interior view (*right*). The same ridge-and-furrow design (flat plates arranged in alternating peaks and troughs) appears in the arch and in the planar roof, which were made of identical units.



WOOD GUTTERS, which channeled water from the roof to the hollow cast-iron columns and thence to underground conduits, also acted as beams. Tightened wrought-iron rods (*bottom*) strengthened the gutters by giving them a slight upward camber. The technique constituted an evolutionary step toward today's practice of prestressing. Gutters were key components of Paxton's ridge-and-furrow roof, which is an early example of the now widely used folded-plate structure. A folded-plate roof derives its stiffness from being composed of flat plates joined at angles. Diagonal grooves visible in cross section (*top*), called weeps, collected condensation from the inside of the roof.

bracing between them. Some of these trusses were all cast iron: lighter ones spanned the 24-foot bays longitudinally and heavier ones spanned the 48 feet of the side aisles and the galleries. Others were of wrought iron and spanned transversely the avenues, which were 48 or 72 feet wide.

An interesting feature of the design of all the trusses is that their top and bottom chords were swelled out, or wider, in the central section of their span. This made them about twice the width of the end sections, where the trusses joined the columns. The cast-iron trusses were cast so that the chords gradually swelled toward their center; the wrought-iron ones were built up by riveting extra bars to the central sections.

This simple adding of area at the central sections strengthened the chords, since they took more stress in the middle of their span than they did at the ends. Deepening the trusses at their center would have been more satisfactory, but the trusses had to be of constant depth in conformity with the overall design concept of Paxton's interior.

The swelling out of the Crystal Palace trusses was in response to the development of truss theory. It was only between 1847 and 1851 that treatises, published in the U.S. and then in England, presented methods for computing truss loads and resolving the forces acting on each member of a truss. Furthermore, the methods for computing truss deflection that provided the theoretical basis for calculations still in use today were not developed until the mid-1860's. Thus it was not until 1850 that a theoretical understanding of truss design began to emerge in England.

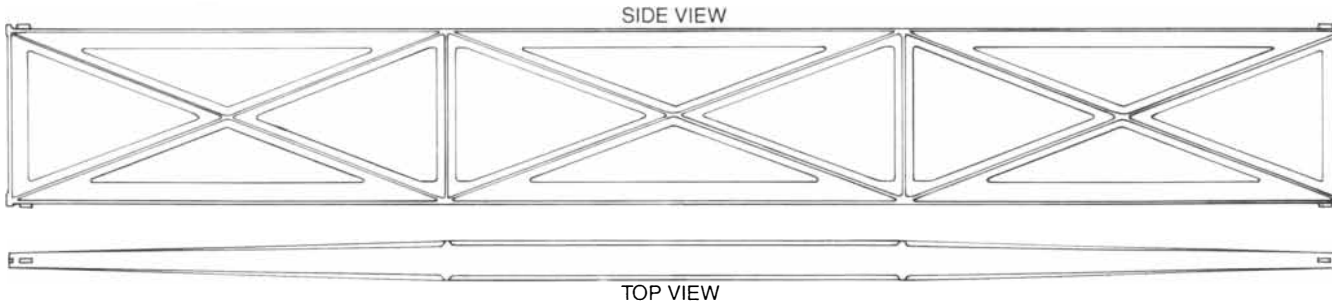
Prestressing is another contemporary construction technique prefigured in the Crystal Palace. Prestressing consists of ultimately enhancing the load-bearing ability of a structural element by loading it externally, thus modifying the state of its internal stress. A familiar example from contemporary technology is the concrete beam, containing steel rods that have been stretched and then anchored to increase the compression in the concrete.

By prestressing a structural member a designer can use methods of construction that would otherwise be impractical, given the material and form of the building. Prestressing also allows the designer greater freedom in selecting the cross sections and profiles of the major structural members without sacrificing structural efficiency.

The evolution of prestressing began in response to the problems connected with building railroad bridges using cast iron. Because wrought iron could not be rolled into a beam more than seven inches deep, cast iron was the universal structural material in 1851. It retained preeminence in spite of the fact that its tensile strength was only about one-eighth its compressive strength.

Recognizing by the middle of the 1840's the greater strength of wrought iron, engineers tried to supplement the tensile deficiencies of cast-iron beams by trussing them with wrought-iron reinforcement. They experimented with numerous ways of attaching wrought-iron bars or rods to cast-iron beams to form so-called trussed girders. Some girders were quite awkward in concept, but they often made it possible for cast iron to span increased distances, as in Robert Stephenson's 96-foot span girder for the Leopold Railway. They also led to such unfortunate collapses as the Dee Bridge failure of May, 1847, in which a passenger train carried five people to their deaths in the River Dee near Chester. Such failures quickly led to the avoidance of cast iron in any large span structure and impeded the development of prestressing.

Although designers of railway bridges generally avoided the trussed girder, Paxton and his colleagues were not discouraged. The Crystal Palace revealed two interesting examples of trussing



CAST-IRON TRUSSES illustrate the emphasis Paxton placed on speed and ease of construction. The trusses were prefabricated and shipped to Hyde Park in fixed lengths of 24, 48 and 72 feet. They were connected to columns in a way that would not allow any deviation from the perpendicular, a technique known as portal bracing. Thus any system of at least three columns and two trusses could stand free, eliminating the need for scaffolding and accelerating the construction process. For greater strength, top and bottom chords were twice as wide at the center as they were at each end (*bottom*), a technique based on advances in theoretical understanding of truss design.

girders with wrought-iron rod. These examples, more elegant than the trussed-girder bridge elements mentioned above, clearly anticipate modern concepts of prestressing.

In the Crystal Palace, Paxton and his colleagues utilized prestressing to resolve a problem that included both aesthetic and engineering elements. It was important that all the trusses be the same depth in order to facilitate rapid erection as well as to enhance the visual homogeneity of the building. Yet trusses had to be of different strengths, depending on the locations for which they were designed.

One way to achieve the required strength of the trusses without varying depth was to prestress them by fabricating them with a camber, or upward bend, in the middle. Thus when they were put in place and loaded they would deflect but not sag. The 72-foot trusses were cambered about 10 inches; the 48-foot trusses were cambered about five inches. Moreover, the truss supports for the galleries and for the lead flat roofs on each side of the transept, both of which bore live loads of moving crowds, were augmented by a wrought-iron rod. This rod ran under the bottom chord and exerted upward pressure on it by being tightened up against cast-iron shoes attached to the chord's underside. This practice of cambering and reinforcing with iron rod is equivalent to the prestressing of concrete beams with steel rod so common in today's building.

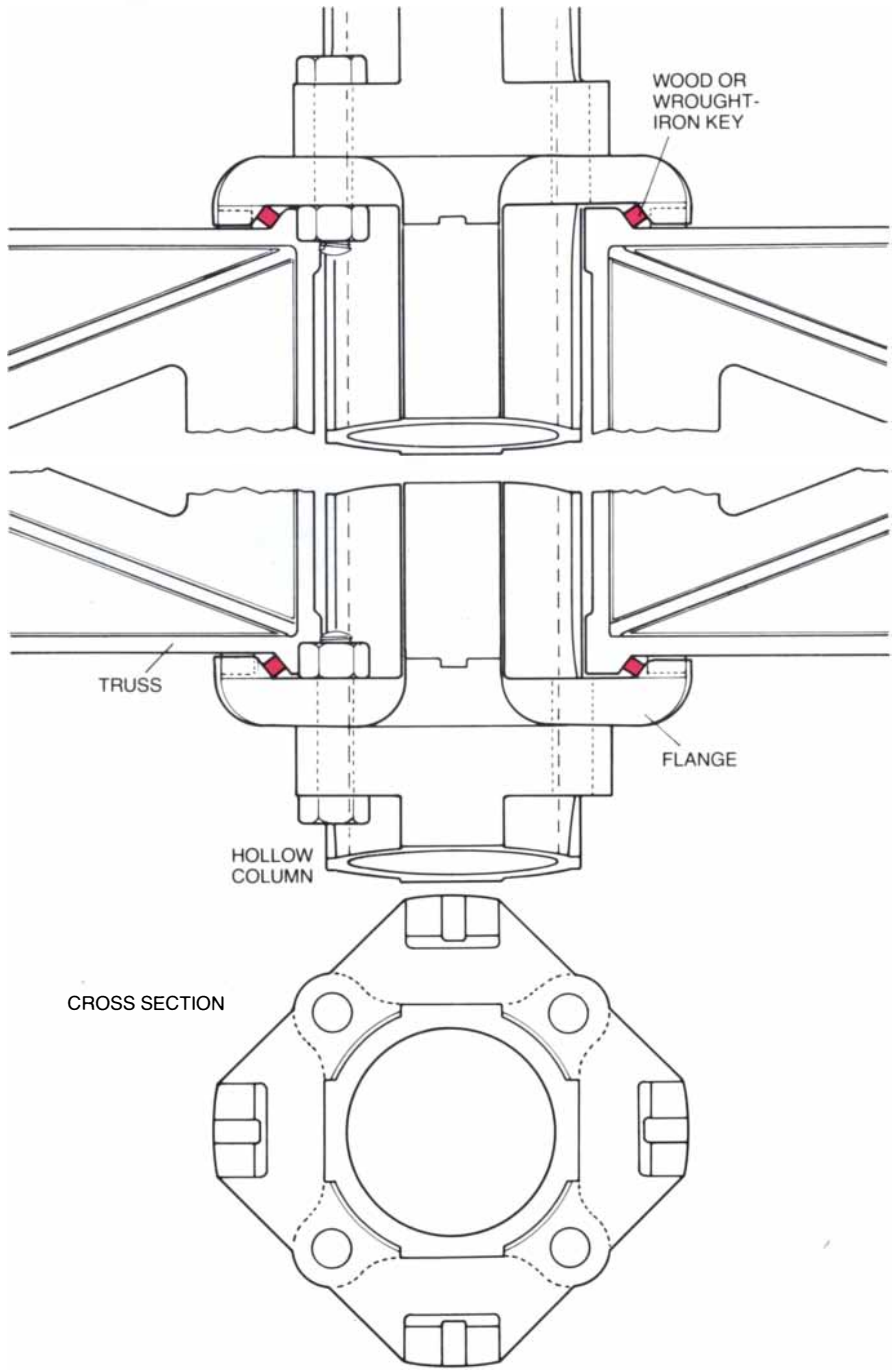
A similar system was used for Paxton's wood gutter. The gutter carried rain from the roofs to the hollow cast-iron columns, which channeled the water through the base pieces to underground pipes. The gutter was also the basic structural element that spanned the 24 feet between the girders; however, it was wood and not deep enough to be a bearing member over that distance. It was therefore cambered up two and a half inches by a wrought-iron rod that was tightened against nine-inch-deep cast-iron shoes. The camber facilitated the flow of water and made the gutter into a truss that could withstand one and a half tons before fracture.

Such prestressing could not have been practiced much earlier in the century; it depended on the development of hydraulic jacks, used to apply and control the pressure needed to test these structural members, on the superior tensile strength of wrought-iron rod and on the application of formulas to calculate the amount of deformation to be induced in each member. The full exploitation of prestressing did not occur in iron construction but in the reinforced-concrete work of such early 20th-century designers as Eugène Freyssinet.

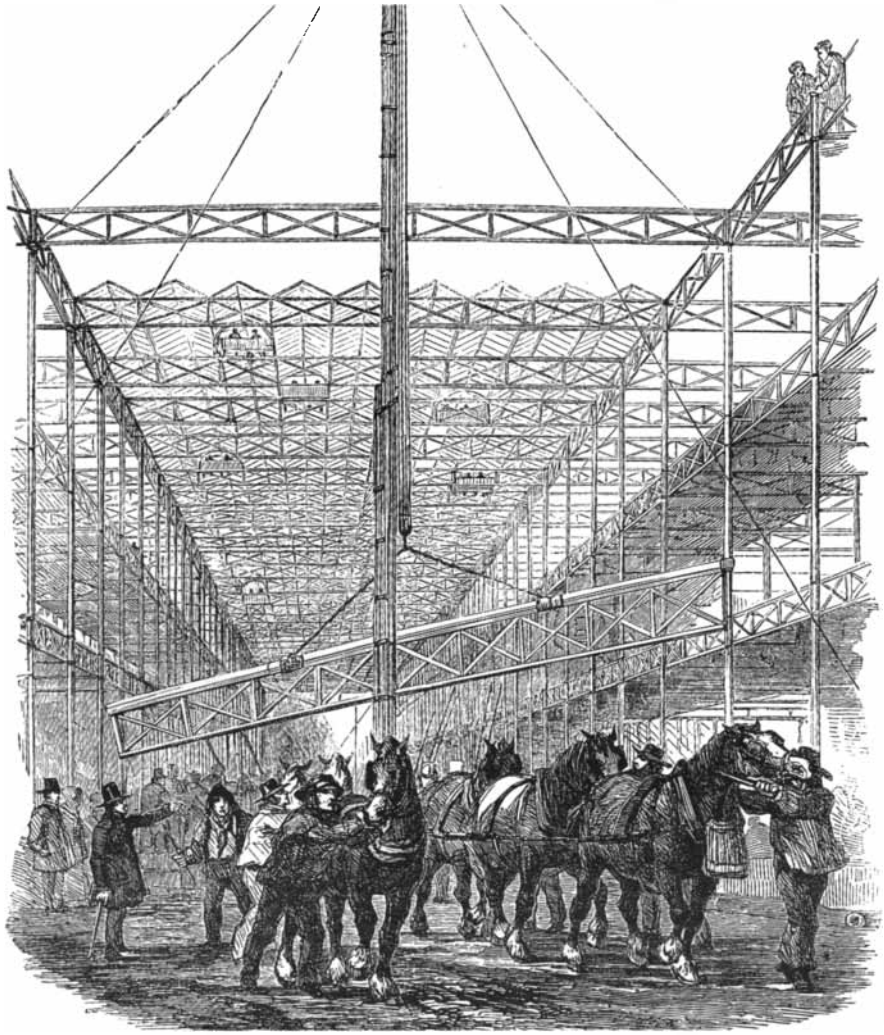
The important structural aspects of the Crystal Palace mentioned above have been overlooked by many observers simply because the enterprise itself was, and remains, overwhelming. That enterprise was a total coordination of design, industrial production, delivery and erection in 17 weeks of a building that covered some 19 acres and was ready for occupancy in six months. The fact that the Crystal Palace was a prefabricated, standardized, modular structure made this feat possible.

Some of its components forced this unusual level of coordination. In adherence to the module imposed by the glass panes, the cast-iron columns were standardized to a uniform external diameter. Moreover, the large quantities of cast iron for columns and trusses made prefabrication essential, since cast-iron members could not be cut to shape or size and fitted on the site as could other materials, including wrought iron.

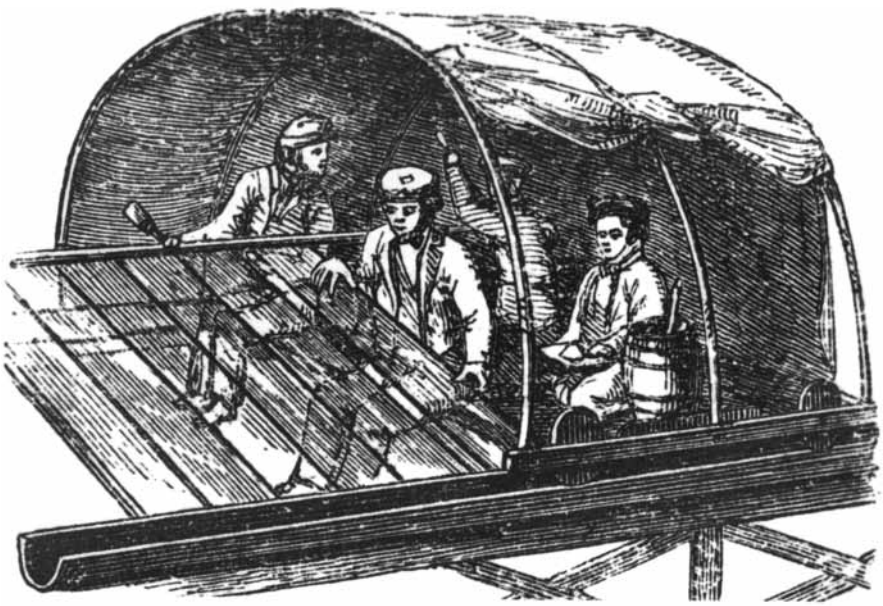
The process of assembly also aided in the structure's rapid erection. It needed no scaffolding, for example. Once four columns connecting pieces and girders were in place, the structure became self-supporting. The cast parts were designed with the easy manipulation of wrenches in mind, so that as many as 310 columns could be set up in one week



CAST-IRON CONNECTING PIECES attached trusses (*see bottom illustration on opposite page*) to columns. Joints were designed for rapid construction. Flared ends of trusses fitted into flanges and were held in place by wrought-iron or wood keys, which could be driven in quickly.



CONSTRUCTION of the Crystal Palace took only 17 weeks largely because the structure was built by highly organized crews assembling a variety of prefabricated modules. In the foreground crews lift trusses to be connected to hollow cast-iron columns. On the roof are six carts called traveling stages, which were used to install the glass panes (*see illustration below*).



TRAVELING STAGES such as this one facilitated construction of the Palace's glass roof. The custom-built carts rode along the wood gutters as the workmen installed panes and sash bars. The traveling stages enabled 80 men to install nearly 19,000 glass panes in a single week.

and the trusses quickly fixed to the connecting pieces by wrought-iron or wood keys. Traveling stages, or wagons, for glazing the roof were devised to run in the Paxton gutters; they enabled 80 men to install 18,932 panes of glass in one week. Such detailed refinements were of a completely new order.

The building site itself was as mechanized as it possibly could have been. Miles of wood sash bars and gutters were machine-prepared on the grounds from rough stock as well as painted by machine rather than by hand. Hammers, drills, hoists and derricks were powered by engines. Hydraulic testing devices made it possible for workmen to unload, weigh, test and stack the largest trusses in less than four minutes.

Paxton, Fox and Henderson made tremendous advances in erection techniques by designing the Crystal Palace for maximum interchangeability. In 1850 the interchangeability in something as simple as a machine-made nut and bolt was new and relatively uncommon; yet the degree of correlation in the erection of the Crystal Palace nearly equaled that of the contemporary skyscraper. As the historian H. R. Hitchcock said, Joseph Paxton's contribution to architecture was "of the order of Henry Ford's principal contribution to modern industrial methods."

In its economy and simplicity of fabrication the Crystal Palace surpassed even the great Victorian railway structures. Economy and simplicity, so admired today, were necessitated by the unusual demands made by the Building Commission and the May 1, 1851, deadline. Had it not been for that deadline, a different, more complicated structure surely would have been built. Not until the 20th century would structures of comparable simplicity and coordination have been erected.

Going against established taste, a few 19th-century architectural theorists had been calling for an architecture of their own time. They sought an architecture that made use of those contemporary materials iron and glass, not to imitate past styles but to evolve a brand-new style. In the Crystal Palace, Paxton and his colleagues did just that.

Transparency, immateriality and the indefinite boundaries of the Crystal Palace were also an outgrowth of the aesthetic goal of all greenhouse designers. That goal, as stated by the director of the Paris botanical gardens in 1852, was one of "hiding all material evidence that one is walking under a glass roof."

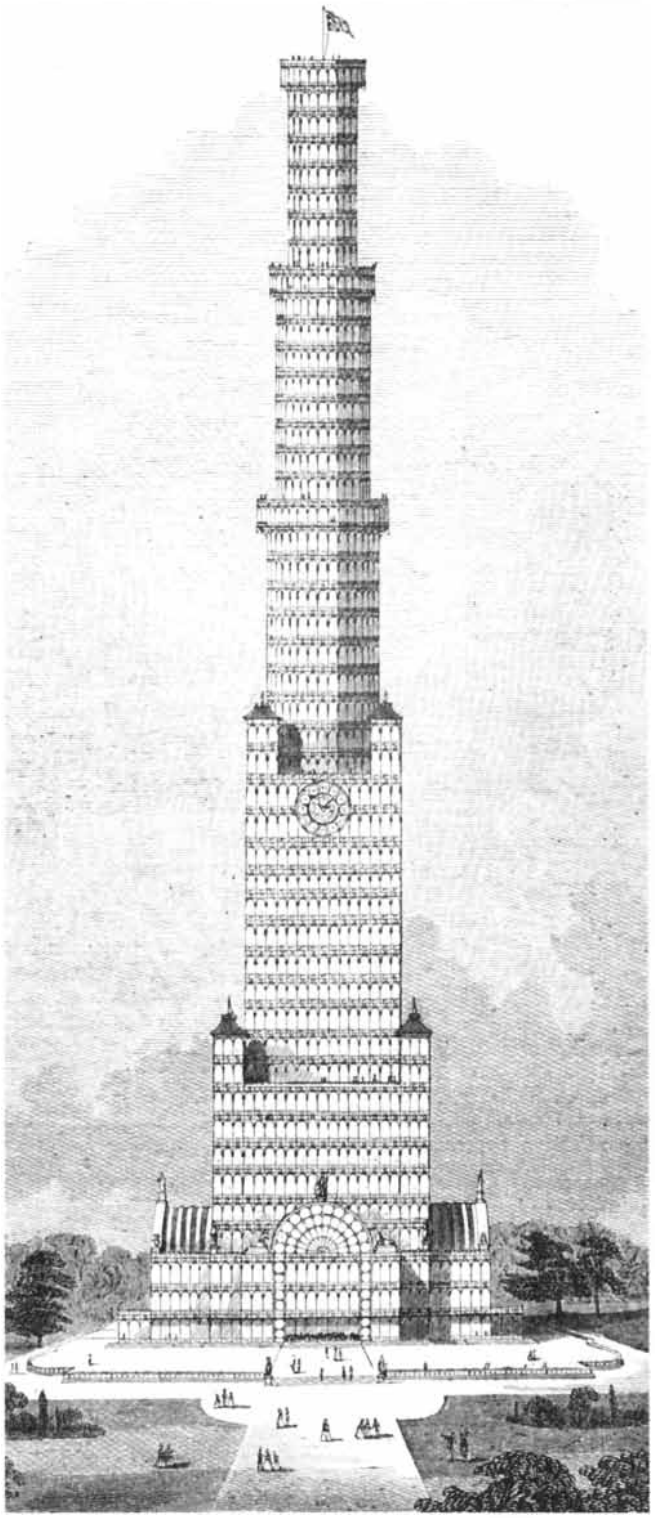
The effect was not lost on contemporary observers. William Makepeace Thackeray called it "a palace as for fairy prince," and a German visitor declared: "We cannot tell if this structure towers a hundred or a thousand feet

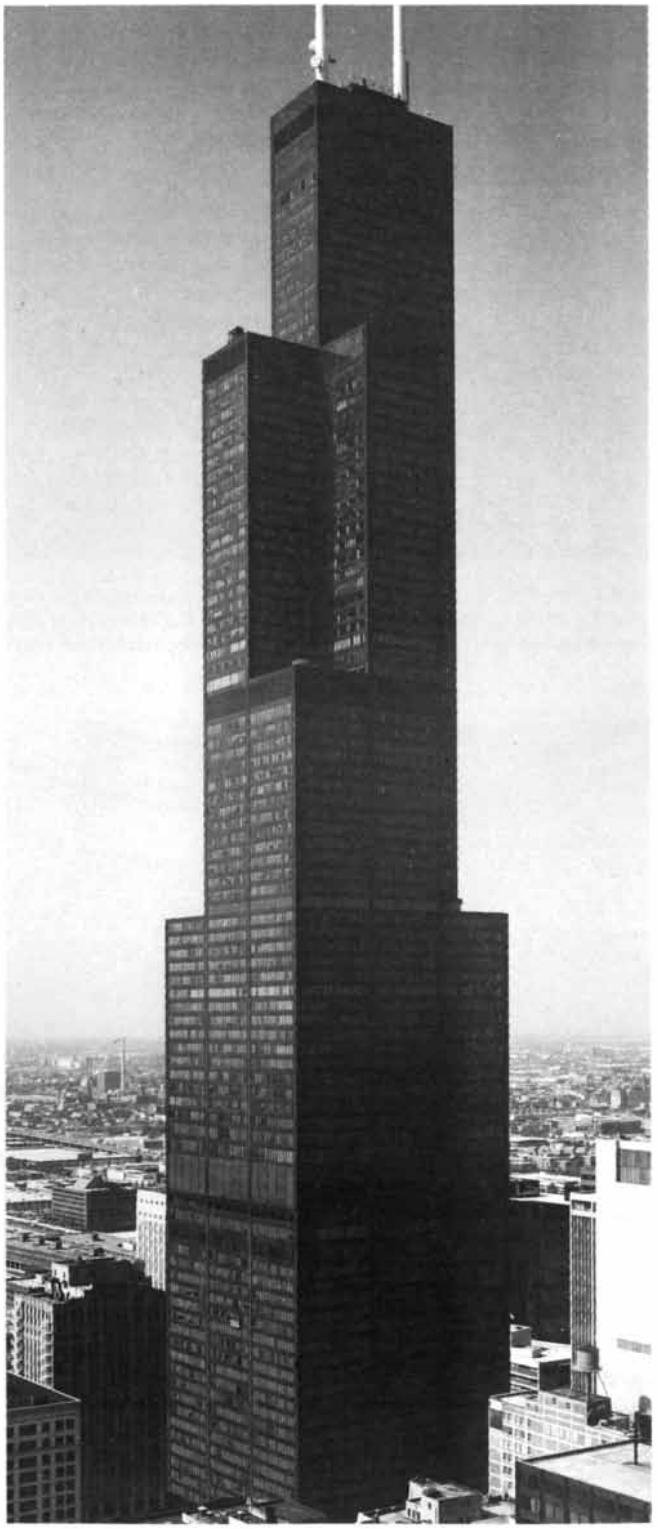
above us... all materiality is blended into the atmosphere."

Nonetheless, formal appreciation of the inherent beauties of unornamented and transparent form had to wait for the modern movements of the early 20th century. Only then could the logical extensions of the Crystal Palace aesthetic be accepted into all types of architecture, both public and private, from the UN Secretariat to the all-glass domestic houses of Philip Johnson and Mies van der Rohe.

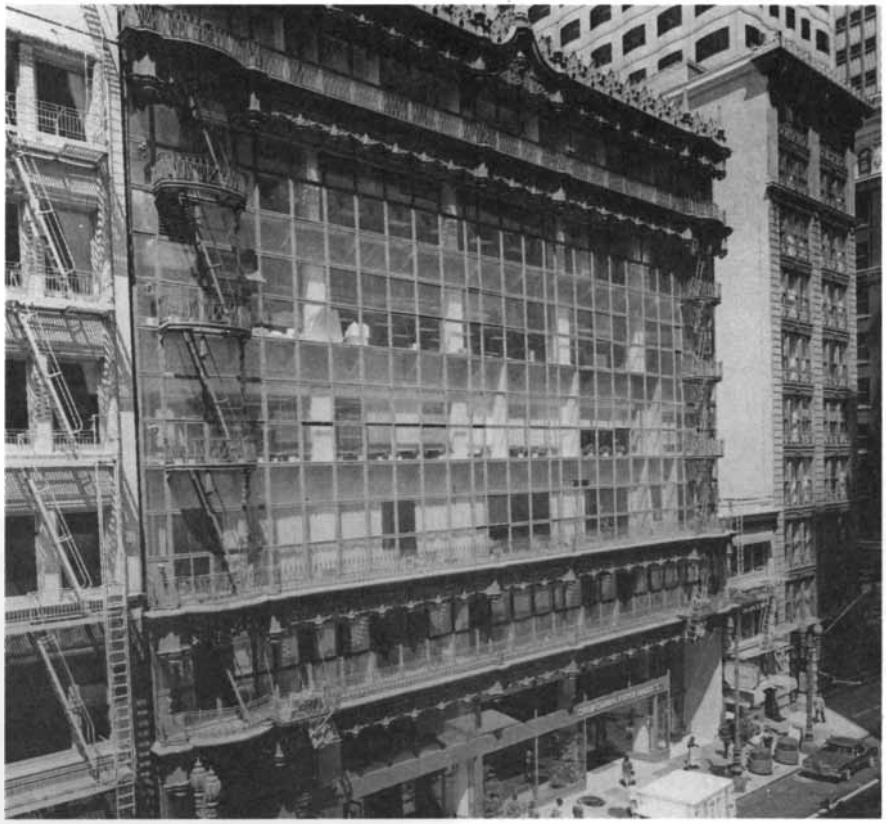
Whatever they thought of it as architecture, the Victorians admired the prefabricated, standardized and modular traits of the Crystal Palace. They saw the virtue of being able to put up buildings like it in any country.

The virtues of the Crystal Palace's modular structure even inspired an ar-





MODULAR CONSTRUCTION of the Crystal Palace prompted a contemporary of Paxton's to suggest that the modular units of the building be rearranged to form a 1,000-foot tower (*left*). A vertical Crystal Palace would have been too heavy for its cast-iron columns; now steel beams make such buildings possible. At right is Skidmore, Owings & Merrill's Sears Tower, built out of stacked modular units.



CURTAIN-WALL DESIGN, a version of which appeared in the Crystal Palace, is seen in the Halladie Building, designed by Willis Polk in 1915. The structure still stands in San Francisco. Glass windows are hung several feet in front of the reinforced-concrete supporting columns.

GALLERIA VITTORIO EMANUELE was inspired by the Crystal Palace. As in the Palace, broad interior spaces and large expanses of glass contribute to a sense of openness and light.

chitect by the name of Burton to propose in 1852 that the Crystal Palace be reerected in the form of a tower 1,000 feet high. Burton's tower could not have been built out of the components of the Hyde Park building. The structure would have been too heavy for its cast-iron columns, and it would not have been stiff enough to resist wind forces. But Burton's recognition that prefabrication and modular construction would be essential to tall buildings anticipated by 30 years the application of steel in Chicago's skyscrapers. Curiously, Burton's all-glass walls and configuration of setbacks is also prophetic in form and height of later Chicago architecture, specifically the Sears Tower by Skidmore, Owings & Merrill.

Its organic, flexible nature makes the Crystal Palace the precursor of a radical movement in Japanese architecture of the 1960's called "Metabolism." The Metabolists accept change as an unavoidable part of modern life and so design buildings that can expand or contract as functions change. Their buildings do not resemble the Crystal Palace in form, but they embody the flexibility and impermanence pioneered by that 19th-century building.

The effect of the Crystal Palace on contemporary building was immediate. It gave rise to a functionally new type, the large exposition building. It was designed to exhibit the largest machines, to display vast quantities of goods and to handle enormous crowds. A series of such buildings followed closely on the heels of Paxton's prototypical Palace.

Dublin held an exhibition in 1852 for which it built a crystal palace, mainly of wood and glass, designed by John Benson. The following year New York housed an exhibition in a crystal palace designed by the architects Georg J. B. Carstensen and Karl Gildemeister. Then Munich built a Glas Palast in 1854, designed by the engineer August Voit, who closely copied the Hyde Park building not only in form but also in structural details.

Some of the most significant functional extensions of the Crystal Palace were shopping arcades or galleries. Most large cities had a glass covered street or shopping arcade by the 1880's. The glorious Galleria Vittorio Emanuele in Milan, to mention only one of hundreds, is still the magnetic hub of that city. These glazed arcades were typically referred to as being of "crystal-palace construction." The real connection, however, was functional rather than structural. They all used large areas of glass for the purpose of lighting to enhance the display of goods and products, and all were designed with interior streets to handle the flow of crowds.

The Crystal Palace, moreover, led di-

rectly to the French and German department stores of glass and iron that were built at the end of the 19th century. Prophetically, a visitor to the original Hyde Park building had called it "neither crystal nor a palace [but] a bazaar, admirably constructed for its purpose." The department store, in fact, first emerged in the year following the Great Exhibition, when Aristide Boucicault opened his Bon Marché in Paris. An English visitor to the Crystal Palace, William Whiteley, was so inspired by the structure that he began to plan large retail stores with plate-glass fronts. He eventually became known as the "Universal Provider," the founder of England's department stores.

Paxton himself went far beyond department stores in proposing ways the glazed structure of the Crystal Palace could serve new functions. The most fantastic was his Great Victorian Way, a project that presaged R. Buckminster Fuller's air-conditioned dome for midtown Manhattan. Paxton presented this plan (the drawings have been lost) to the Select Committee on Metropolitan Communications in June, 1855. The Great Victorian Way was to have incorporated a railway and boulevard more than 11 miles long, girdling London, uniting all its railways and providing rapid transportation. It was to have been a glazed enclosure 72 feet wide with a vaulted roof rising to a height of 108 feet. Buildings were to flank it so that the boulevard shoppers would have direct access to stores as well as to apartments and offices on its upper floors. At the same time they would be protected from the London weather. Ever the gardener, Paxton emphasized the healthy aspects of such glass buildings and proposed an air-conditioning system for this mini-environment.

Already the vision of Paxton's building standing in Hyde Park prompted *Punch* to quip, "We shall be disappointed if the next generation of London children are not brought up like cucumbers under a glass." After all, if the greenhouse was for the cultivation of nature and was one source for the Crystal Palace, why not adapt it to the cultivation of society as well?

Although both the Great Victorian Way and the New York Fuller dome have yet to be realized, the Crystal Palace remains amazingly prophetic of 20th-century structural, architectural and aesthetic developments. It stood, along with the railway sheds and bridges, at the forefront of the structural understanding of its time. It rose just at the time when engineering was throwing off the mantle of intuitive craft to take on that of applied science. It was one of the world's great triumphs of intuitive engineering, and as with all great engineering, it was also great architecture.

THE AMATEUR SCIENTIST

The troublesome teapot effect, or why a poured liquid clings to the container

by Jearl Walker

The exasperating tendency of a poured liquid to cling to the outside of the container is known as the teapot effect. What causes it? One immediately thinks of surface tension, but it turns out that another factor is more important: the pressure is higher at the outer surface of the fluid than it is at the inner surface, so that the liquid is pushed against the container.

At times, such as when one is pouring acid, one needs to forestall the effect so that the liquid will not run down the side of the container and onto the counter. The technique is to put a glass stirring rod across the top of the container. The liquid then runs along the rod and arrives unspilled in the container one is trying to fill.

A similar technique works for pouring cream or milk into coffee. You tend to pour slowly in order not to generate splashes, and so you inadvertently precipitate the teapot effect. You can avoid it by putting the blade of a table knife across the top of the dispenser.

The teapot effect can also be seen in fountains where the water runs slowly from a pool or a horizontal pipe. The water clings to the structure instead of arcing away from it. Some indoor water gardens rely on this effect. Water flows down a wall built with many projecting stones. Moss grows on the wall, which is kept wet by the teapot effect.

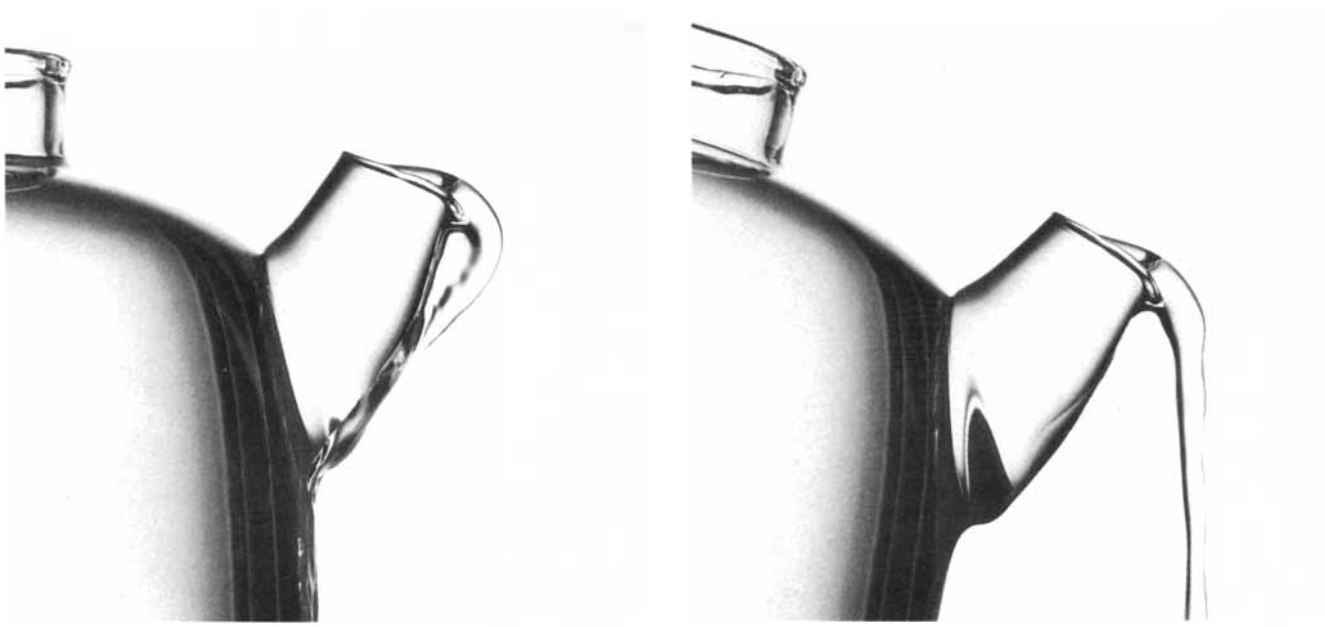
More often the effect is unwanted. When water runs slowly off the sill of a window, it can round the bottom edge of the sill and run into the juncture of the sill and the house wall. If the water seeps into the juncture, it can damage the wall. To avoid the problem a rectangular groove is cut into the underside of the sill parallel to the wall. The vertical sides of the groove are too steep for the stream to flow over, and so the water falls directly from the sill.

The teapot effect was named and investigated in 1956 by Markus Reiner of the Israel Technion–Institute of Technology. Initially he was curious about a seemingly unrelated phenomenon. If a cube of salt is put in fresh water with its top surface horizontal, it dissolves in an odd way. It shrinks in height, but it does not change in width or depth. The flow of the dissolved salt can be monitored by placing a small amount of potassium permanganate on the top surface to color the salty water. Since salt water is denser than fresh water, the flow should be across and off the top of the crystal. Instead the flow rounds the top edges and travels down the sides. Apparently the flow of salty water along the sides protects them from dissolution.

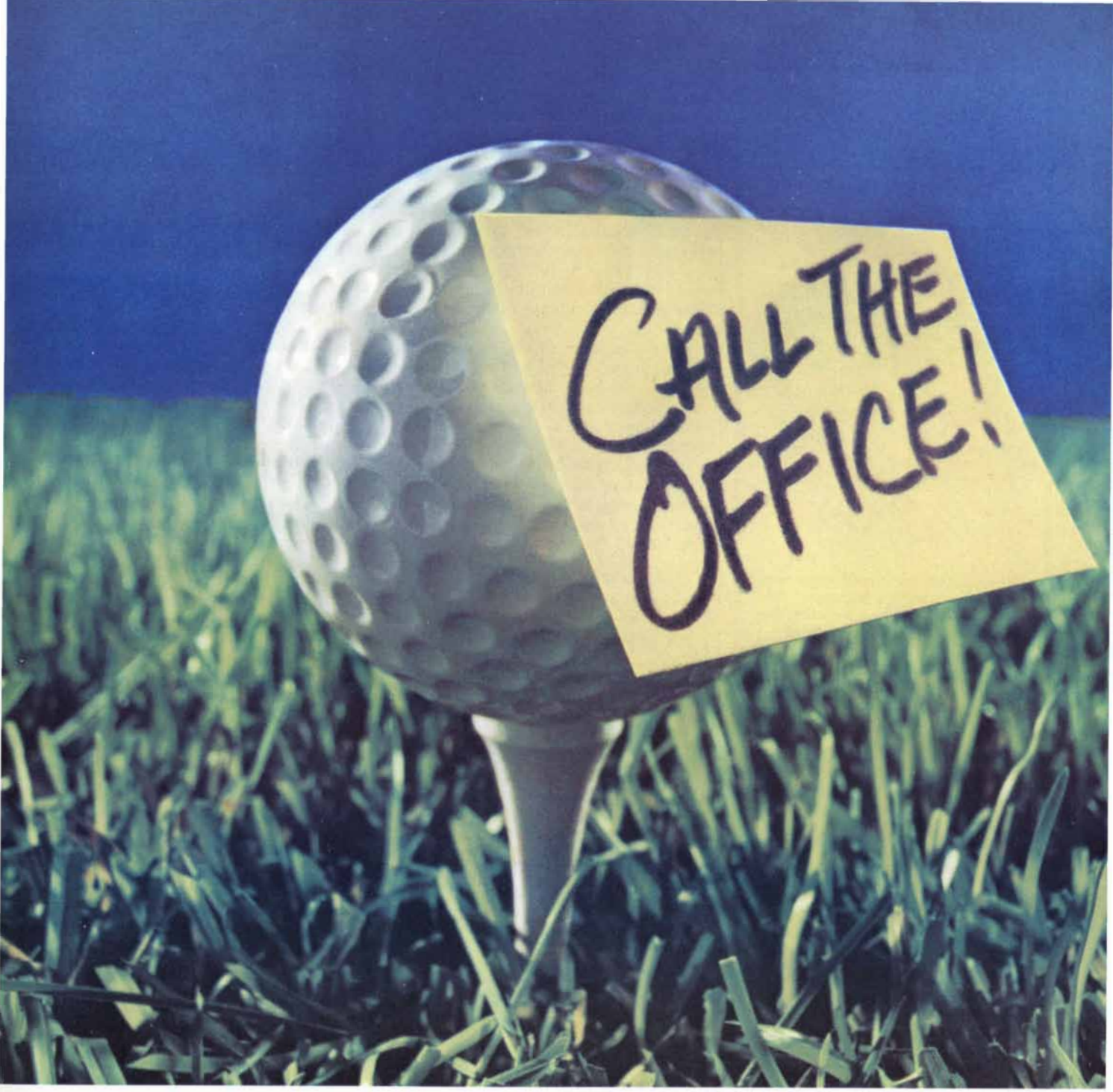
In another experiment an Erlenmeyer flask was submerged and inverted in a large container of fresh water. A stream of concentrated salty water was directed onto the bottom of the flask past a piece of permanganate. The stream flowed along the bottom, turned the edge and then continued down the sloping side instead of arcing away from the flask. After several centimeters it became unsteady, detached from the flask and proceeded downward.

In the next experiment the flask was placed right side up in a container of concentrated salty water. A stream of fresh water was directed at the bottom of the flask. The stream rounded and then flowed along the sloping side of the flask instead of arcing upward off the bottom. After traveling for a few centimeters it oscillated and broke free, rising to the surface of the salty water.

Reiner experimented with other solids. He altered the slopes, produced streams of sinking cold water in a container of warm water and replaced the smoothly curving edge of the glass flask with a sharp edge of metal. Whatever he did, the streams still rounded the edge and adhered to the solid for at least a short distance.



A poured liquid displaying the teapot effect **(left)** *and flowing freely* **(right)**



Now it's easier to leave word anywhere.

Press on a 3M Post-it® Note anywhere you want it noticed. It stays put without tape, tacks, paper clips or staples. The secret is a unique 3M adhesive that holds on almost any surface, then peels away easily.

The idea was born of a 3M employee's need to mark pages in a choir book. But it's become a communications medium that's creating a minor revolution among note-leavers everywhere.

By listening to people's needs, 3M has pioneered over 700 products for office, training and business problems. We've developed everything from correction tape for typists to microfilm systems that file 20,000 lines a minute and laser systems that find any of 108,000 computer data frames in just 3 seconds.

And it all began by listening.

3M hears you...

For your free 3M Products and Technologies Brochure, write: Dept. 060710/3M, P.O. Box 22002, Robbinsdale, MN 55422.

Name_____

Address_____

City_____ State____ Zip_____

Or call toll-free: **1-800-336-4327.**

3M

Reiner also examined the flow of water poured from a teapot. Again one fluid (the water) flowed through another (the air) in the presence of a solid surface. If the flow was fast, the stream left the spout in an arc with no dribbling. If the flow was slow, the stream sometimes went along the spout and then broke off at the bottom of the pot. Sometimes the flow left the lip of the spout in a rearward arc. Flowing at an intermediate speed, the stream split in two; one part moved in an arc and the other ran down the spout.

An explanation of the teapot effect was published in 1957 by Joseph B. Keller of New York University. To follow the flow one constructs lines called streamlines that are everywhere tangent to the velocity vector of the water. If the water flows over a horizontal surface, the streamlines are all horizontal, indicating a uniform flow. If the water flows over and past a shallow obstacle, the lines crowd together, indicating an increase in the rate of flow.

The upper illustration on page 148 shows streamlines for flows over an edge of a square corner and of a thin plate. In both cases the lines crowd together, indicating an increase in the speed of the water. The lines nearest the solid surface crowd together the most, indicating the greatest increase in speed.

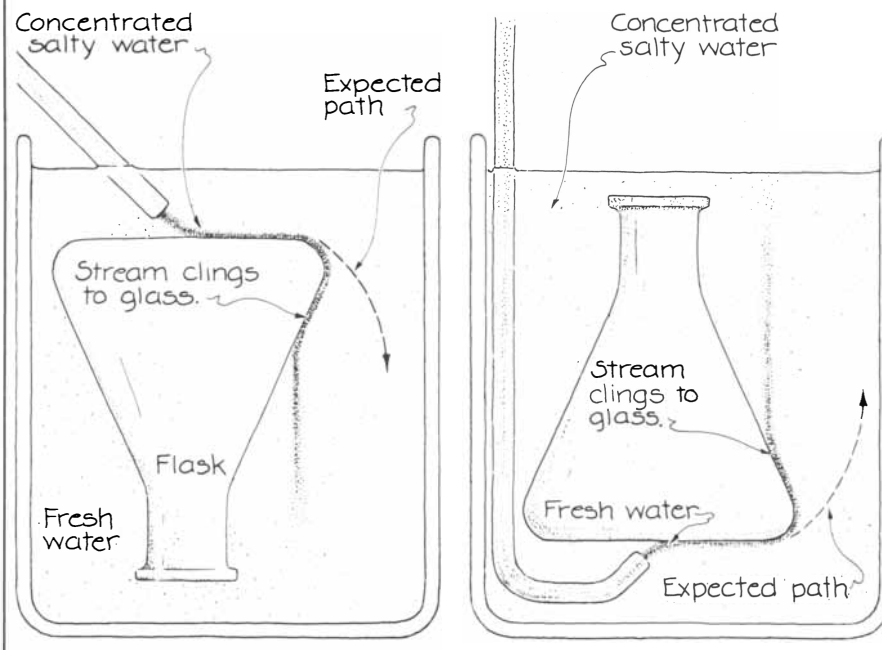
When a portion of a stream of water increases in speed, its kinetic energy increases also. Suppose the flow is not significantly downward, so that one can rule out gravity as the cause of the increase in energy. Then the energy must come from the pressure in the liquid, which can be regarded as a form of potential energy. When the speed of a portion increases, the pressure at that point decreases. The total energy of the portion, that is, the sum of its kinetic energy and the energy associated with the pressure, remains constant.

Since the total energy of a streamline is constant, the increase in speed of the portion at an edge necessitates a decrease in the pressure there. The pressure of the water on the free surface of the stream is kept constant because of the continuous push by the air. Hence when the stream reaches the edge, a difference in pressure exists across its depth, with atmospheric pressure on the free surface and a reduced pressure at the edge. It is this pressure difference that can force the stream to flow over the edge.

A stream of water carries momentum because it has both speed and mass. When you pour water rapidly, the momentum is too great for the pressure difference at the edge of the container to make the stream flow over the edge. The stream arcs from the container in the normal way. With less momentum (slower flow) the pressure difference is sufficient to make the stream turn.

Consider the stream after it has rounded the edge of a plate. Keller discovered that the stream can flow along the underside of the plate. Again the streamlines are all horizontal, indicating that the velocity vectors of the water elements across the depth are parallel and the same size. There is still a pressure difference across the depth. At the bottom of the stream (on the free surface) atmospheric pressure prevails; at the top of the stream (next to the plate) the pressure is lower.

To understand this pressure difference consider a stationary pool of water. Atmospheric pressure pushes on the



Markus Reiner's experiments with salty water and fresh water

surface. Below the free surface the pressure must be higher because of the weight of the water. The additional pressure increases with depth.

When a stream flows along the underside of a horizontal plate, the pressure must be higher at the bottom of the stream than it is at the top because of the weight of the water between the two points. Since the bottom is at atmospheric pressure, the top must be at less than atmospheric pressure. The stream is forced against the underside of the plate by the pressure difference.

The stream is not stable, so that it cannot flow indefinitely under the plate. Small perturbations quickly grow large enough to detach the stream from the plate. Keller found the length of travel to be several centimeters.

If the plate slants downward, the stream runs farther before falling away. Since the pull of gravity is then not perpendicular to the stream, its force is weaker. Moreover, the oscillations brought about by small perturbations to the stream do not grow as quickly. This is the type of flow one sees along the spout of a teapot in the teapot effect.

The stream might also be able to flow upward along a slightly slanted surface. Within a few millimeters, however, it will be slowed by gravity. Then it either collects in drops that fall or oscillates enough to detach from the surface.

The secret to the proper design of a milk dispenser lies in the shape of the spout's edge. If the milk can round the edge and then flow horizontally or downward, the teapot effect will result. If the edge is shaped so that the milk rounding it would have to flow upward at a sharp angle, the dispenser will be free of the teapot effect even at the lowest rates of pouring. With an improperly designed spout one's only solution is to position a knife across the lip. The milk flows along the vertical knife more readily than it would flow along the slanted or horizontal spout.

A related example of clinging flow is the Coanda effect, named after Henri Coanda, the Romanian engineer who discovered it. Sometimes the name is applied to the teapot effect, but usually it serves to describe two separate phenomena. When a jet of fluid (either a gas or a liquid) emerges from a slot and passes a shallow step, it can attach itself to the wall at the base of the step. The attachment results from an entrainment of the ambient fluid between the jet and the step. I do not believe this phenomenon is a factor in the teapot effect.

Sometimes the term Coanda effect refers to the attachment of a fluid stream to a continuously curved surface. Suppose a horizontal jet of water is directed at the side of a beach ball. As the surface of the ball begins to curve away from the jet it creates a partial vacuum at the adjacent water surface. Since the

free surface of the jet remains at atmospheric pressure, the pressure difference across the jet forces the jet against the ball. This phenomenon is quite similar to the teapot effect except for the partial vacuum.

I did an experiment similar to the one Reiner had done with salt. In a container of warm water I put an ice cube tilted so that as water melted from the top surface it would run down that surface. (The freshly melted water sinks, because it is denser than the ambient warm water.) On the top surface I deposited a drop of food coloring so that I could monitor the flow. The stream rounded the downward edge of the cube and ran down the side for a few millimeters before it detached and fell to the bottom of the container.

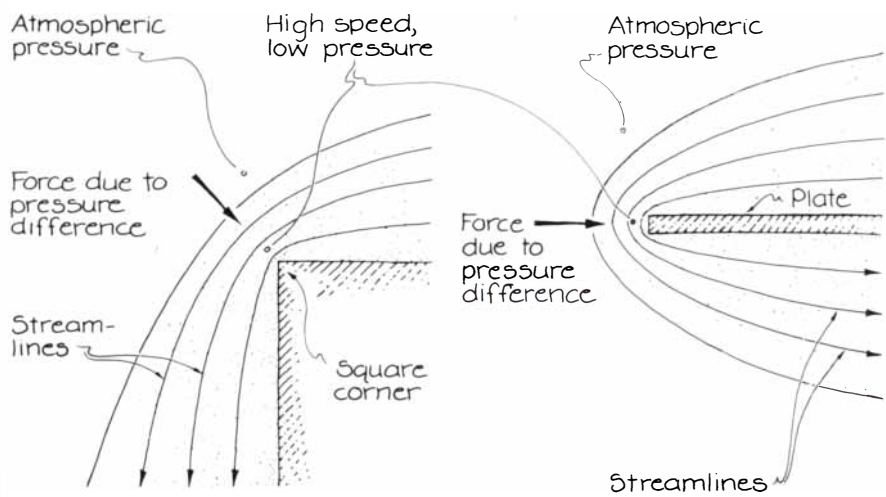
I next did several experiments on the flow of water from a clear plastic container used in the preparation of gravy. Normally after cooking meat I pour the drippings and grease from the pan into the container, wait until the grease separates into a layer on top of the drippings and then pour off the drippings. For this purpose the container has a pouring spout extending from the bottom so that the bottom fluid rather than the top fluid is poured.

In my experiments I clamped the container to a ring stand mounted on the basin of my kitchen sink. Water ran from the faucet into the container and then through the pouring spout. I could adjust the angle of the container. In addition I could vary the rate at which water entered the container and thus the rate at which it left the spout.

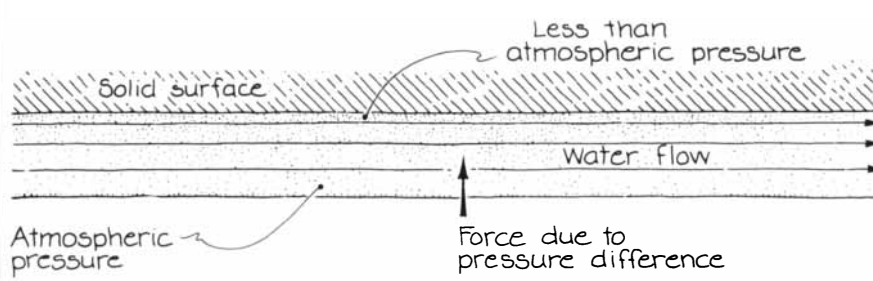
I first tilted the container 45 degrees and set the water flowing at a moderately high rate. The water arced forward from the spout in the familiar way. As I decreased the flow rate the stream receded toward the container. Soon the rate was low enough so that the water rounded the edge of the spout slightly and arced rearward toward the container. When I decreased the flow more, the stream rounded the edge and flowed down the surface of the spout, eventually falling at the bottom. The attached stream was quite stable.

I squirted a few drops of food coloring into the container in order to follow the flow. Most of the flow was near the upper part of the spout. The water near the lower part hardly moved. When the water poured over the edge, it mixed too fast for me to distinguish speeds there or farther along the stream.

When the teapot effect was established, the water flowing along the underside of the spout changed shapes. Near the edge it had a narrow, deep hump. Several millimeters down the side of the spout it suddenly changed to a wide, shallow flow that was maintained until the stream separated from the container. As I slowly increased the



Streamlines around two types of edge



Joseph B. Keller's explanation of inverted flow

flow the hump grew larger and began to oscillate. Drops were frequently thrown clear of the spout, but most of the water descended along the underside of the spout. I probed the flow with a needle at the edge of the spout. The hump and the rest of the inverted flow of water remained stable.

Then I increased the flow rate slightly. The hump grew into a loop that was nearly circular in cross section and extended a centimeter or more down the underside of the spout, changing then into a wide, shallow stream. Between the loop and the spout was a thin film of water. When I probed the film with a needle, I could detect no changes in the loop or the shallow stream below it. The water in the film seemed to be stagnant. At the lower end of the loop were stationary ripples. As I increased the flow rate again slightly these ripples grew more pronounced and the loop suddenly pulled away from the spout, forming the normal curved arc.

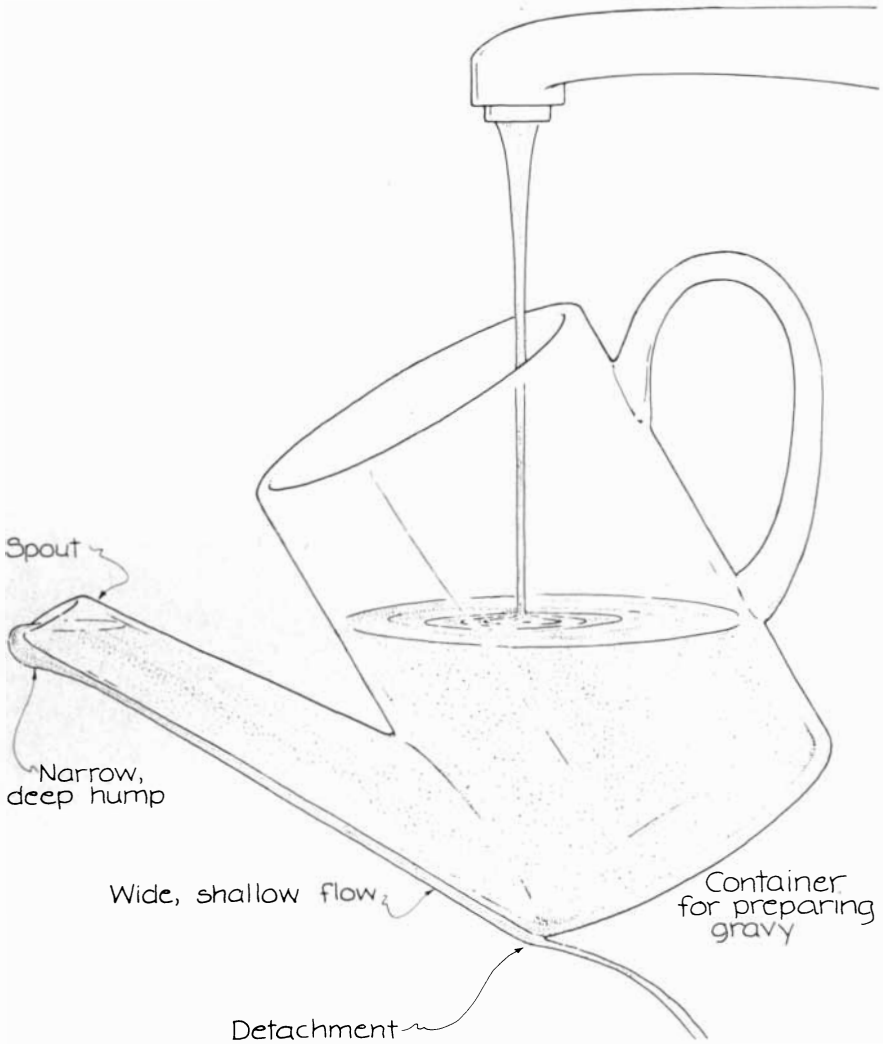
Similar stationary ripples can be seen in another thin stream of water. Hold a flat surface in a narrow stream from a faucet. Raise the surface toward the faucet. When it is close enough, the ripples appear. When it is closer still, the ripples become more pronounced, creating large oscillations in the diameter of the stream.

Lord Rayleigh once investigated such waves. They are stationary because they travel up the stream as quickly as the stream falls. Apparently the same kind of stationary wave is created in the thin stream forming the loop along the spout of a container. The impact of the loop on the spout creates the waves.

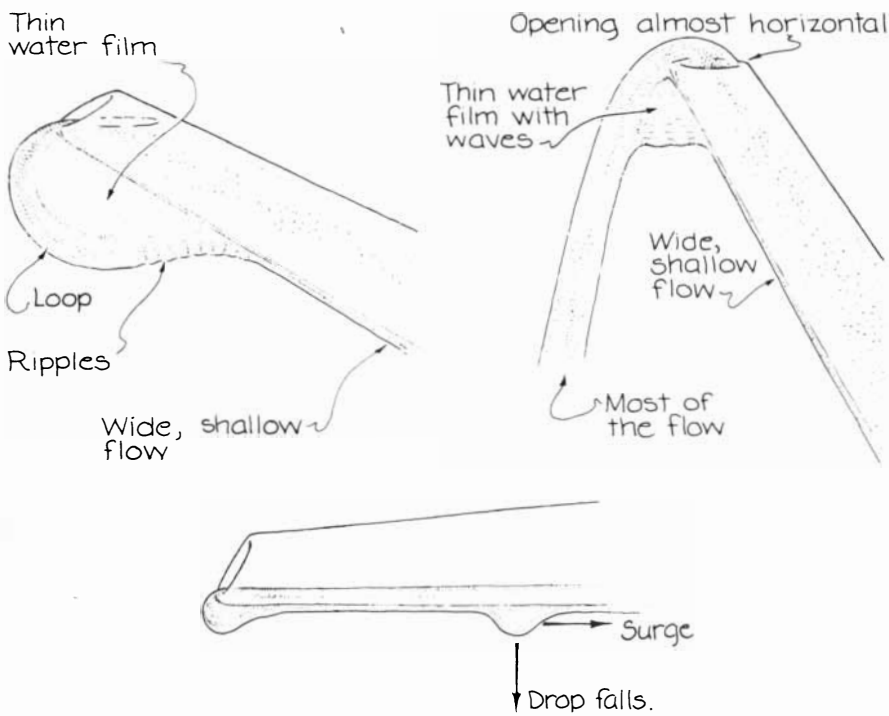
When the flow rate was low enough to create a hump rather than a loop, the stream occasionally oscillated enough to become detached from the spout. By rubbing my finger through the upper end of the spout I could easily make the stream reattach itself to the spout. If the flow rate was high enough to create a loop, this trick did not work.

The hump seems similar to one I can produce in a horizontal flow of water. When the stream meets an obstacle that is shallower than the stream, the water develops a hump over the obstacle or just downstream from it. The obstruction squeezes together the streamlines of the flow as the speed of flow increases. The speed can become supercritical, meaning that the water flows faster than waves can travel over water. At the downstream side of the hump the flow becomes subcritical. Often turbulence develops there.

Something comparable seems to happen in the water rounding the edge of a spout. The edge is an obstacle that squeezes the streamlines together. The free surface of the stream develops a hump that appears just below the edge of the spout. At the low side of the hump



An arrangement for investigating the teapot effect



Flow patterns seen with the effect

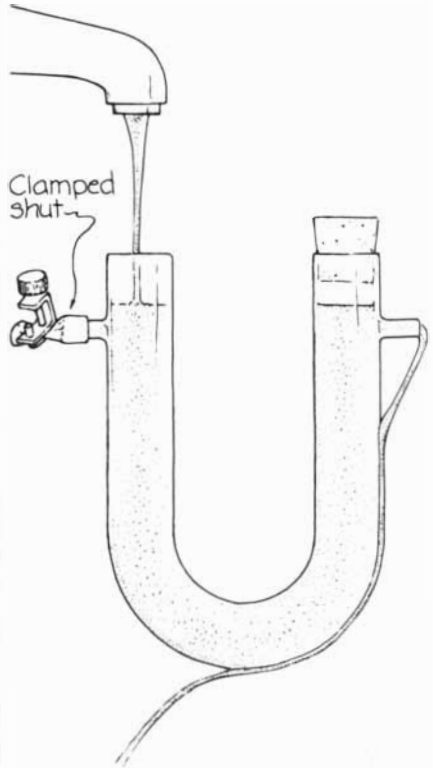
the water slows, spreading out in a shallow layer that continues down the underside of the spout. If the flow is fast enough, the low side of the hump is turbulent, throwing off drops. The rest of the stream responds to the pressure difference described by Keller and is forced against the spout.

When the flow is only slightly faster, the pressure difference at the edge is insufficient to force the stream fully over the edge. The water, which is almost thrown free, is pulled together by surface tension to form a stream with an approximately circular cross section. The stream runs downward and forms the loop, but it is not thrown free of the spout because of a thin film of water that remains between it and the spout. The pull of this film brings the loop back to the spout, where the stream then slows and spreads out. The loop is quite unstable. Once it breaks free I cannot reestablish it with my finger because I cannot create the thin film needed to reattach it. As soon as I remove my finger the vigor of the flow drives the loop forward into the familiar curved arc.

Some people think that coating the underside of a teapot spout with butter overcomes the teapot effect. I tried this trick (using margarine) without success. After removing the margarine I adjusted the container so that the spout was approximately horizontal. If the flow was slow, the stream rounded the edge of the spout and traveled along its underside for one or two centimeters. Then it detached as a rearward arc. Just below the edge of the spout the stream again had a narrow hump that extended for a few millimeters. Thereafter the stream was wide and shallow.

When the flow rate was very low, the water on the underside of the spout would surge. In the region of detachment the water slowed enough to form hanging drops. With each surge enough water was added to a drop to make it fall. The surge first pushed the drop along the spout; after the drop detached, the remaining water retreated toward the open end of the spout until the next surge drove it forward again.

I also set the container so that the open end of the spout was almost horizontal. Low flow rates enabled the water to round the edge of the spout, form a hump and then become a wide, shallow stream. As I increased the flow a loop developed, grew large and then broke free of the spout. Not all the water followed the loop. Part of it still flowed down the side of the spout. Between the free stream and the attached one was a short, narrow film on which I could see stationary waves of small wavelength. The water that breaks free of the spout must have enough momentum to counteract the pressure difference across the depth of the stream at the edge of the spout. Apparently some of the water



The U-tube phenomenon

loses much of its momentum at the edge. It can then flow down the underside of the spout in the manner described by Keller.

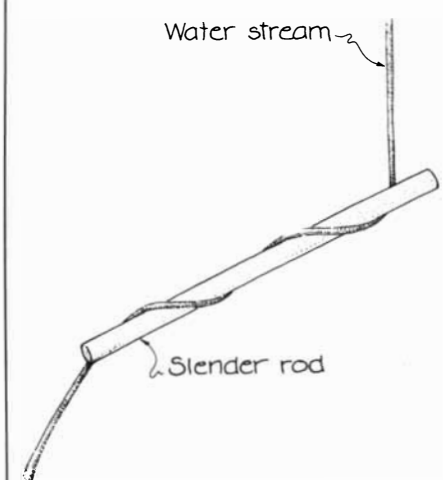
I also experimented with a U tube made with side arms. I stoppered one of the open ends of the device and put the other end in a slow flow from the faucet. Water emerged from the side arm on the stoppered side. With the flow adjusted carefully the stream left the side arm in a rearward arc and then reattached itself at a low point on the U. Usually this arrangement was unstable because the stream would immediately pull up and cling to the full underside of the side arm and the tube.

I was curious about how far an inverted stream of water would flow if the surface of an object was tilted at a moderate angle. In my backyard I set up a seven-foot length of flat metal tilted from the vertical by about 45 degrees. At the upper end of its underside I attached a garden hose. As I adjusted the flow of water I produced an inverted stream that traveled the length of the metal. Sometimes the stream rounded the side edge of the metal and then flowed along its upper side. Sometimes the stream (either inverted or on top of the metal) slowly oscillated like a snake.

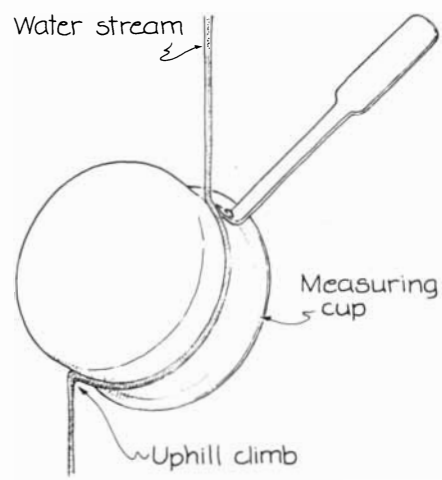
I did similar experiments with slender rods of metal, glass and plastic held at a moderate angle. I adjusted the stream from the faucet until it was narrower than the rod. When the stream struck the rod off center, it traveled along the underside of the rod and then spiraled around it several times. It never traveled upward in relation to the basin, but at some points it did travel horizontally.

The spiraling results from the curvature of the rod. Since the stream is forced to follow the curvature because of the pressure difference across its depth, it initially circles the underside of the rod and then the top side. This motion continues down the length of the rod. The spiraling disappears when I replace the narrow rod with a wider one held at the same angle. Then the stream attempts to spiral but ends up looping back to the underside of the rod, whence it descends as an inverted flow.

In a final experiment I placed a round measuring cup in the stream from the faucet. The circular cross section was in a vertical plane, so that the stream struck on one side. The stream clung to the side, following the curve until it was below the cup. Then it began to travel upward from the basin for a short distance before it detached. Apparently the pressure difference responsible for the inverted flow of the stream is strong enough to make the stream actually flow uphill before gravity slows it enough to pull it free of the cup.



A spiral on a rod



The cup phenomenon

From extreme environments on earth to remote areas of space — wherever the greatest challenges of modern communications are found, you'll find CSD. Our extensive resources, unmatched experience and total involvement in every area of communications technology makes CSD the pacesetter of the industry. If you can solve difficult communications problems by incorporating divergent technologies and innovations into a complete system design, join CSD on the threshold of new discoveries. Because of our depth, leadership and emphasis on individual achievement, we can create a position specially suited to your interests and abilities. Wherever your career goals lie, you'll find the room to accomplish them at CSD — because we have the future for you.

NETWORK SYSTEMS ENGINEERS

Assist in private line communications systems design for modern defense systems, including applications of commercial practices such as Common Channel Signaling and ISDN Network Management. Qualifications should include BSEE/BSCS and 5+ years experience in modern switched network design. Hardware and software design experience and some defense exposure a strong plus.

Send your resume to: Brian T. Harris, GTE's Communication Systems Division, Suite 441, 1777 North Kent St., Arlington VA 22209.

SYSTEMS ENGINEERS, SOFTWARE ENGINEERS

We are seeking qualified engineers with a BS, MS or PhD in Electrical Engineering or Computer Science, 5+ years experience with military communications systems. Project needs will utilize technical expertise in the areas of:

- Digital Signal Processing
- Computer Controlled Data Communications Systems
- Computer Communications Networks
- Secure Systems
- Troposcatter Modems

Send your resume to: Harriet T. Solomon, GTE's Communication Systems Division, 77 "A" Street, Needham MA 02194. Or call us at 1-800-225-3956.

SYSTEM DESIGN/ ACQUISITION ENGINEER

Packet Communications

System design engineer to lead front-end studies of packet networks. Responsibilities include system trade studies and transforming system requirements into hardware and software design specifications. The qualified candidate should have experience with current technology and systems in either packet radio or packet switching, including analytical determination of correct packet size, network access schemes, coding techniques and other system parameters. BSEE or CS required, MS preferred. 5+ years experience. Familiarity with DoD requirements, specifications and systems is a plus.

Send your resume to: Sherry T. Huffer, GTE's Communication Systems Division, 400 Park Plaza, P.O. Box 13279, Research Triangle Park NC 27709. Or call us at 919-549-1111.

Along with a network of project opportunities offering new growth comes a strong system of support and rewards. Success here is recognized within GTE and across the communications industry. Internal advancement can take the direction of management or development of technical knowledge. And because we provide year after year of challenge, CSD has kept top quality professionals and maintained low turnover in a fiercely competitive market for engineers. For access to the important opportunities in communications, we can be the last place you'll ever have to look. Our benefits include complete tuition reimbursement, life, health and dental insurance coverage for dependents. U.S. Citizenship is required.

An Equal Opportunity Employer M/F

IF IT'S HAPPENING IN COMMUNICATIONS, IT'S HAPPENING AT CSD.

Research Triangle Park NC, Arlington VA, Colorado Springs CO, Rockville MD, Natick & Needham MA

GTE Government Systems

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

THE SCIENCE OF CHECKERS AND DRAUGHTS. Tom Wiswell. A. S. Barnes and Company, 1973.

MATHEMATICAL GAMES. Martin Gardner in *Scientific American,* Vol. 236, No. 2, pages 121–126; February, 1977.

METAMAGICAL THEMAS. Douglas R. Hofstadter in *Scientific American,* Vol. 248, No. 1, pages 14–22; January, 1983.

SPACE-BASED BALLISTIC-MISSILE DEFENSE

BALLISTIC MISSILE DEFENSE. Edited by Ashton B. Carter and David N. Schwartz. The Brookings Institution, 1984.

THE REAGAN STRATEGIC DEFENSE INITIATIVE: A TECHNICAL, POLITICAL AND ARMS CONTROL ASSESSMENT. Sidney D. Drell, Philip J. Farley and David Holloway. Center for International Security and Arms Control, Stanford University, 1984.

SPACE-BASED MISSILE DEFENSE: A REPORT BY THE UNION OF CONCERNED SCIENTISTS. Union of Concerned Scientists, Cambridge, Mass.; March, 1984.

DIRECTED ENERGY MISSILE DEFENSE IN SPACE. Ashton B. Carter. Office of Technology Assessment, U.S. Government Printing Office; April, 1984.

THE STRATEGIC DEFENSE INITIATIVE: DEFENSIVE TECHNOLOGIES STUDY. Department of Defense, U.S. Government Printing Office; April, 1984.

PRIONS

UNCONVENTIONAL VIRUSES AND THE ORIGIN AND DISAPPEARANCE OF KURU. D. Carleton Gajdusek in *Science,* Vol. 197, No. 4307, pages 943–960; September 2, 1977.

NOVEL PROTEINACEOUS INFECTIOUS PARTICLES CAUSE SCRAPIE. Stanley B. Prusiner in *Science,* Vol. 216, No. 4542, pages 136–144; April 9, 1982.

SCRAPIE PRIONS AGGREGATE TO FORM AMYLOID-LIKE BIREFRINGENT RODS. Stanley B. Prusiner, Michael P. McKinley, Karen A. Bowman, David C. Bolton, Paul E. Bendheim, Darlene F. Groth and George G. Glenner in *Cell,* Vol. 35, No. 2, pages 349–358; December, 1983.

SOME SPECULATIONS ABOUT PRIONS, AMYLOID AND ALZHEIMER'S DISEASE. Stanley B. Prusiner in *The New England Journal of Medicine,* Vol. 310, No. 10, pages 661–663; March 8, 1984.

SEISMIC TOMOGRAPHY

WORLDWIDE DISTRIBUTION OF GROUP VELOCITY OF MANTLE RAYLEIGH WAVES AS DETERMINED BY SPHERICAL HARMONIC INVERSION. Ichiro Nakanishi and Don L. Anderson in *Bulletin of the Seismological Society of America,* Vol. 72, No. 4, pages 1185–1194; August, 1982.

ANISOTROPY AND SHEAR-VELOCITY HETEROGENEITIES IN THE UPPER MANTLE. H.-C. Nataf, I. Nakanishi and Don L. Anderson in *Geophysical Research Letters,* Vol. 11, No. 2, pages 109–112; February, 1984.

MAPPING THE LOWER MANTLE: DETERMINATION OF LATERAL HETEROGENEITY IN *P* VELOCITY UP TO DEGREE AND ORDER 6. Adam M. Dziewonski in *Journal of Geophysical Research,* Vol. 89, No. B7, pages 5929–5952; July 10, 1984.

MAPPING THE UPPER MANTLE: THREE-DIMENSIONAL MODELLING OF EARTH STRUCTURE BY INVERSION OF SEISMIC WAVEFORMS. John H. Woodhouse and Adam M. Dziewonski in *Journal of Geophysical Research,* Vol. 89, No. B7, pages 5953–5986; July 10, 1984.

CARTILAGE

ADULT ARTICULAR CARTILAGE. Edited by M. A. R. Freeman. Pitman Medical Publishing Co., Ltd., 1979.

CELL BIOLOGY OF EXTRACELLULAR MATRIX. Edited by Elizabeth D. Hay. Plenum Press, 1981.

PROTEOGLYCANS: STRUCTURE AND FUNCTION. Vincent Hascall in *Biology of Carbohydrates: Vol. 1,* edited by Victor Ginsburg and Phillips W. Robbins. John Wiley & Sons, Inc., 1981.

CARTILAGE, VOL. 2: DEVELOPMENT, DIFFERENTIATION AND GROWTH. Edited by B. K. Hass. Academic Press, 1983.

EPSILON AURIGAE

THE INTERPRETATION OF ϵ AURIGAE. G. P. Kuiper, O. Struve and B. Strömgren in *The Astrophysical Journal,* Vol. 86, No. 5, pages 570–612; December, 1937.

THE ATMOSPHERE OF THE I COMPONENT OF EPSILON AURIGAE. Robert P. Kraft in *The Astrophysical Journal,* Vol. 120, No. 3, pages 391–400; November, 1954.

AN INTERPRETATION OF ϵ AURIGAE. Su-Shu Huang in *The Astrophysical Journal,* Vol. 141, No. 3, pages 976–984; April 1, 1965.

THE ZETA AURIGAE STARS. K. O. Wright in *Vistas in Astronomy,* Vol. 12, pages 147–182; 1970.

INTERACTING BINARY STARS. Jorge Sahade and Frank Bradshaw Wood. Pergamon Press, 1978.

THE MYSTERY OF EPSILON AURIGAE. Francis J. Reddy in *Sky and Telescope,* Vol. 63, No. 5, pages 460–462; May, 1982.

A LATE ICE-AGE SETTLEMENT IN SOUTHERN CHILE

EARLY MAN IN THE NEW WORLD. Richard S. MacNeish in *American Scientist,* Vol. 64, No. 3, pages 316–327; May–June, 1976.

ANCIENT SOUTH AMERICANS. Edited by Jesse D. Jennings. W. H. Freeman and Company, 1983.

EARLY MAN IN THE NEW WORLD. Edited by Richard Shutler. Sage Publications, 1983.

THE CONTINUOUS PROCESSING OF METALS IN THE U.S.S.R.

TEORIA RASCHETA USILII V PROKATNYKH STANAKH (THEORY OF FORCE CALCULATION IN ROLLING MILLS). A. I. Tselikov. Izdatelstvo "Metallurgiya" (Publishing House for Metallurgy), 1962. In Russian.

METOD PODOBIIA V TEORII PROKATKI (SIMILARITY METHOD IN THE THEORY OF ROLLING). V. S. Smirnov, A. K. Grigoryev, A. D. Karachunsky and O. Ya. Mel'nichuk. Izdatelstvo "Nauka" (Science Publishing House), 1971. In Russian.

THE CRYSTAL PALACE

THE WORKS OF SIR JOSEPH PAXTON, 1803–1865. George F. Chadwick. The Architectural Press, London, 1961.

THE GLASS HOUSE. John Hix. The MIT Press, 1974.

DEVELOPMENTS IN STRUCTURAL FORM. Rowland Mainstone. The MIT Press, 1975.

PAXTON'S PALACE. Anthony Bird. Cassell and Company, Ltd., 1976.

IRON BRIDGE TO CRYSTAL PALACE: IMPACT AND IMAGES OF THE INDUSTRIAL REVOLUTION. Asa Briggs. Thames and Hudson in collaboration with the Ironbridge Gorge Museum Trust, 1979.

THE AMATEUR SCIENTIST

THE TEAPOT EFFECT...A PROBLEM. Markus Reiner in *Physics Today,* Vol. 9, No. 9, pages 16–20; September, 1956.

TEAPOT EFFECT. Joseph B. Keller in *Journal of Applied Physics,* Vol. 28, No. 8, pages 859–864; August, 1957.

Choose one:

- ☐ Comfort and luxury.
- ☐ The confidence of 4-wheel drive.
- ☐ Both of the above (in the American Eagle).

By choosing "Both of the above" you've obviously made the right choice. A choice that most car makers, however, don't even give you!

Until the American Eagle. An automobile with the comfort and luxury of a fine road car, plus the capability to let you shift-on-the-fly to the confidence of 4-wheel drive. So simple. So incredibly logical.

After all, why should your family have to give up a roomy five-passenger interior, leather upholstering, plush carpeting, power windows, door locks and more for the privilege of having 4-wheel drive??

You shouldn't. And you don't have to. And that's why your choice is so clear: the one and only American Eagle. Available in wagon or 4-door sedan.

THE 2-WHEEL/4-WHEEL DRIVE AMERICAN EAGLE

AVAILABLE FOR PURCHASE OR LEASE AT PARTICIPATING AMC DEALERS

SEAT BELTS SAVE LIVES

Zeiss

Optical glass. Formulated, ground and polished to predictable standards. Possessed of exact refractive and dispersive properties. Satisfactory performance expected and guaranteed. But it was not always so.

For centuries, optical-glass making was an empirical art, whereas today's product, often of exotic composition, is the result of a precise science. The groundwork for that science was laid just a century ago by Otto Schott, an early associate of Carl Zeiss. It was from Schott's imaginative experimentation in raw materials and melting processes that the methodology for manufacturing optical glass to pre-determined standards evolved.

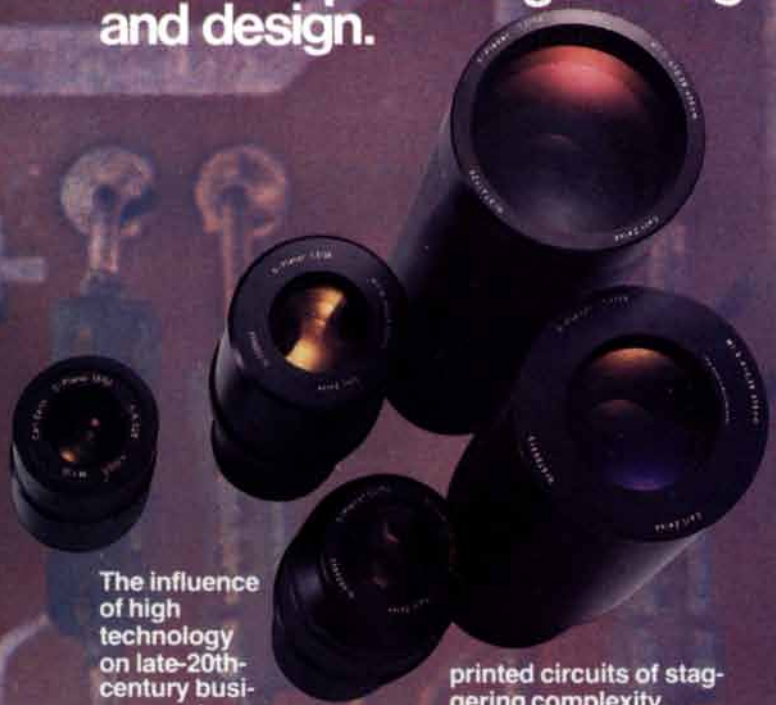
Another associate of Carl Zeiss, Dr. Ernst Abbe, derived the mathematical formulae that remain to this day the scientific basis for all microscope lens design. Moreover, to prove his theories in practice, Dr. Abbe invented and built a dozen or more analytical and measuring instruments, including his famous refractometer and the comparator.

Unswerving pursuit of quality has characterized Zeiss since its inception more than a century ago. It earned for Zeiss the identification as the "great name in optics," a distinction it enjoys to the present day.

To the tradition of superb optics and precision mechanics that made it a leader in the "age of industrialization," Zeiss has added a third skill – advanced electronics – to maintain its pre-eminence in this age of computers.

Today:

The great name in electro-optical engineering and design.



The influence of high technology on late-20th-century business, science, and industry is pervasive. Yet, in truth, "hi-tech" products almost always embody not one but several technologies. When the need dictates the union of optics, mechanics, and electronics, the inimitable source is Zeiss.

It was Zeiss microlithography lenses, for example, that first enabled the semiconductor industry to break the submicron production barrier. Zeiss twenty-year leadership in this supremely complicated technology is acknowledged by the fact that 90% of all such lenses used in the world today come from Zeiss. Now in their third generation, they are used in the design and production of printed circuits of staggering complexity.

Other Zeiss lenses are integral components in such familiar electro-optical equipment as laser scanners, copying machines, telecopiers, TV cameras, and infrared night vision optical systems.

Zeiss also has enviable credentials in several other technologies of significance to the business machines industries. These include: optical character recognition, facsimile, scanning systems, video disc read/write heads, and lasers.

The name "Zeiss" is synonymous with progress in many scientific and technical disciplines. Zeiss surgical microscopes launched—and continue to direct the evolution of—microsurgical procedures. Zeiss electron microscopes lead the field in versatility and ease of operation. In astronomy, Zeiss has just introduced a telescope with a mirror that is 3.5 meters in diameter and weighs 13 tons. The mirror took 20 months to construct and is a tribute to Zeiss skills in optics, mechanics, electronics, and laser technology.

Zeiss technology, uniquely expressed in optics, engineering and electronics, finds increasing welcome across a broad spectrum of business, science, and industry. Currently representative are: computer numerical controlled measuring; computerized mapping and surveying; microscopy and image analysis; optics and electronics for microsurgery; and products for sight.

And if that's Zeiss today...

Who knows what tomorrow will bring?

Carl Zeiss, Inc.
One Zeiss Drive
Thornwood, NY 10594
(914) 747-1800

Who knows what tomorrow will bring?

ZEISS

Carl Zeiss, Inc.

DEWAR'S PROFILE:

JEAN-CLAUDE NÉDÉLEC

HOME: New York City.

AGE: 35

PROFESSION: Chef; co-owner Glorious Food, caterers.

HOBBY: Eating meals prepared by someone else.

LAST BOOK READ: *Tom Jones*, Henry Fielding.

LATEST ACCOMPLISHMENT: Catered the Museum of Modern Art's reopening, with more than 10,000 people in six days.

WHY I DO WHAT I DO: "I've been a food fanatic since I was a little boy in France. And, what may be more important, I'd rather work for me than anyone else I know."

QUOTE: "Professional secret? Just pack well and don't forget anything."

PROFILE: Energetic. Thrives on what the rest of us might call chaos. Sees a sit-down dinner for two thousand as an intimate little gathering.

HIS SCOTCH: "Dewar's 'White Label,' on the rocks. Its taste blends perfectly with the sense of accomplishment I feel when five parties have gone well. On the same night."