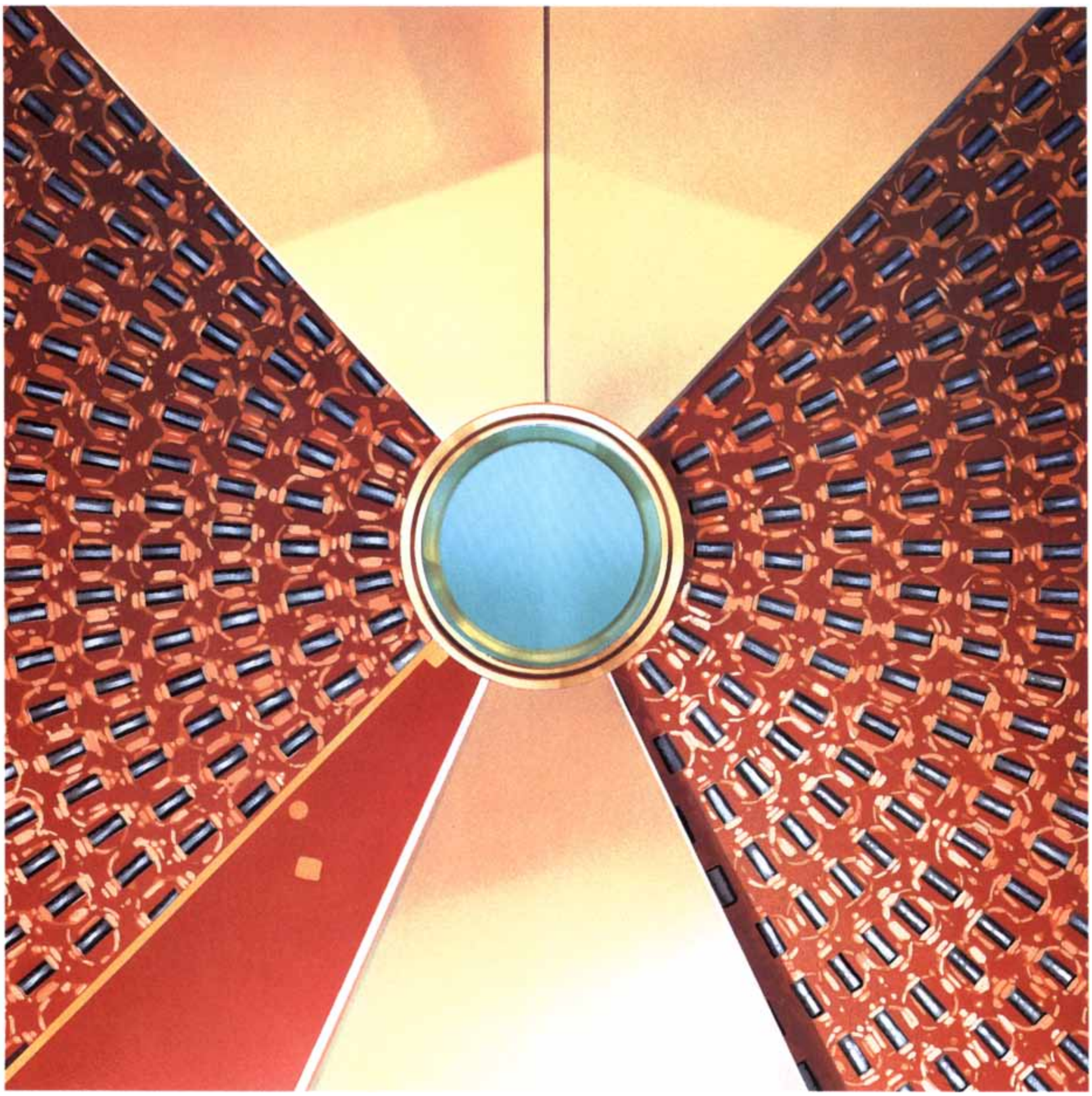
SCIENTIFIC AMERICAN

OCTOBER 1989
$2.95

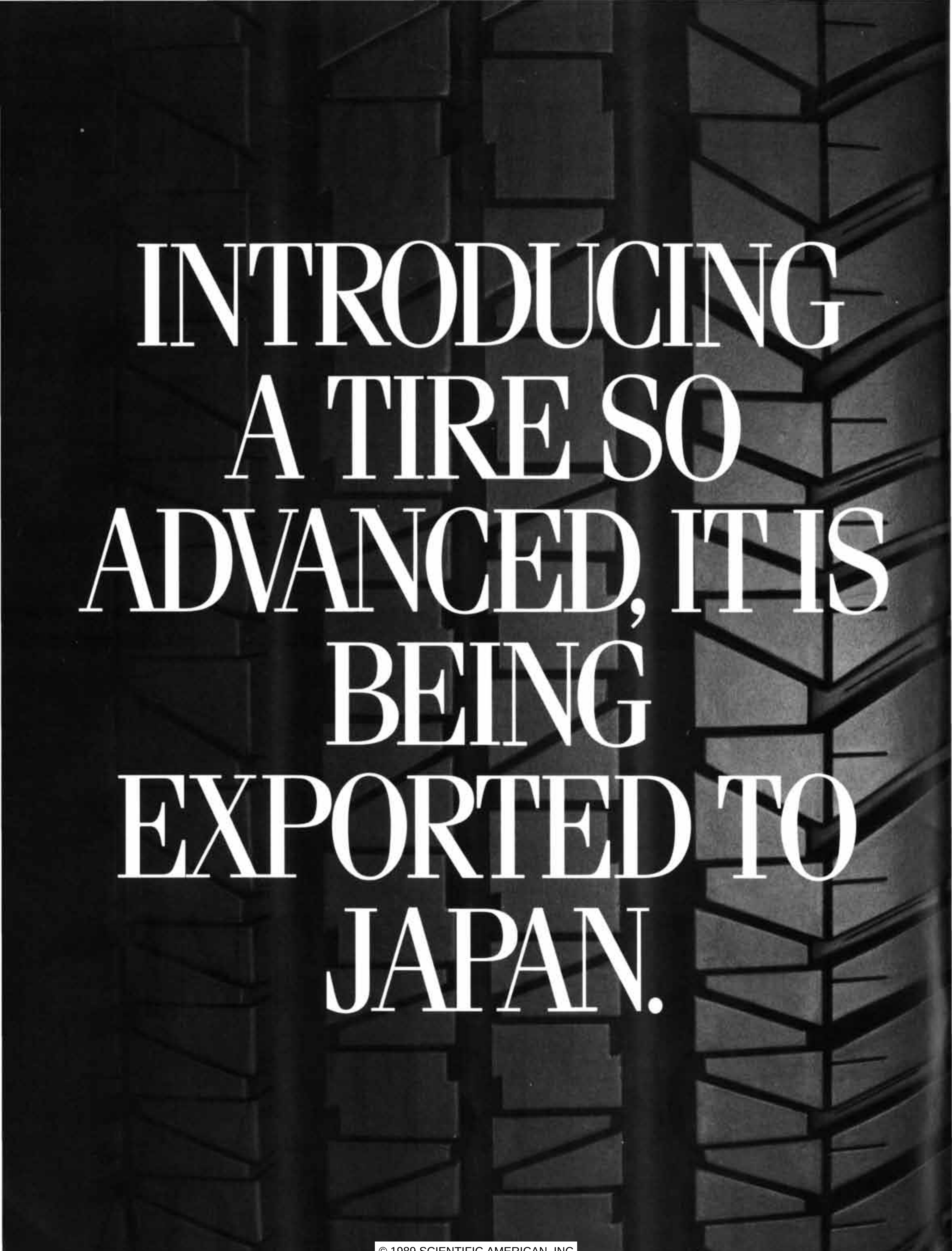
How to turn technology into industrial competitiveness.

Microscopes that probe the ups and downs of atoms.

Where did the Indo-European languages come from?



Seeing into the heart of matter: *the Mark II detector finds Z^0 particles generated at the Stanford Linear Collider.*


INTRODUCING
A TIRE SO
ADVANCED, IT IS
BEING
EXPORTED TO
JAPAN.

The new Eagle GA Touring Radial.

Where luxury meets performance. And both win.

Quite possibly, you have never heard of Goodyear's newest tire, the Eagle GA Touring Radial. But the engineers at Lexus certainly have.

For more than 18 months, Goodyear and Lexus worked together on three continents of the globe.

Their objective: To make a tire with the handling capabilities of an Eagle high-performance radial, and the superior ride suitable for a luxury car.

Developmental tires were tested in America. In Canada. In Germany, Luxembourg, and in Japan.

And now, the Goodyear Eagle GA Touring Radial is being sent to Japan for fitment as the factory-specified, original equipment tire on the new $35,000 Lexus LS 400.

The Goodyear Eagle GA Touring Radial bridges the gap between the aggressive handling, grip and stability of an outright performance radial and the smooth, undisturbed, quiet ride of a quality luxury radial.

What it can offer your car is a quiet, smooth and undisturbed ride over a variety of road surfaces. Plus the ability to handle your car's full performance capabilities.

You can get details of Eagle GA availability for your car from your local Goodyear retailer. (Call 1-800-CAR-1999 for the Goodyear retailer nearest you.)

Every Eagle GA Touring Radial is speed-rated. And it is available in all-season mud and snow versions.

For Lexus, there was no doubt as to the best tire for their $35,000 flagship: the Goodyear Eagle GA Touring Radial.

It is the tire where luxury meets performance. And both win.

GOODYEAR

SCIENTIFIC AMERICAN

October 1989 Volume 261 Number 4

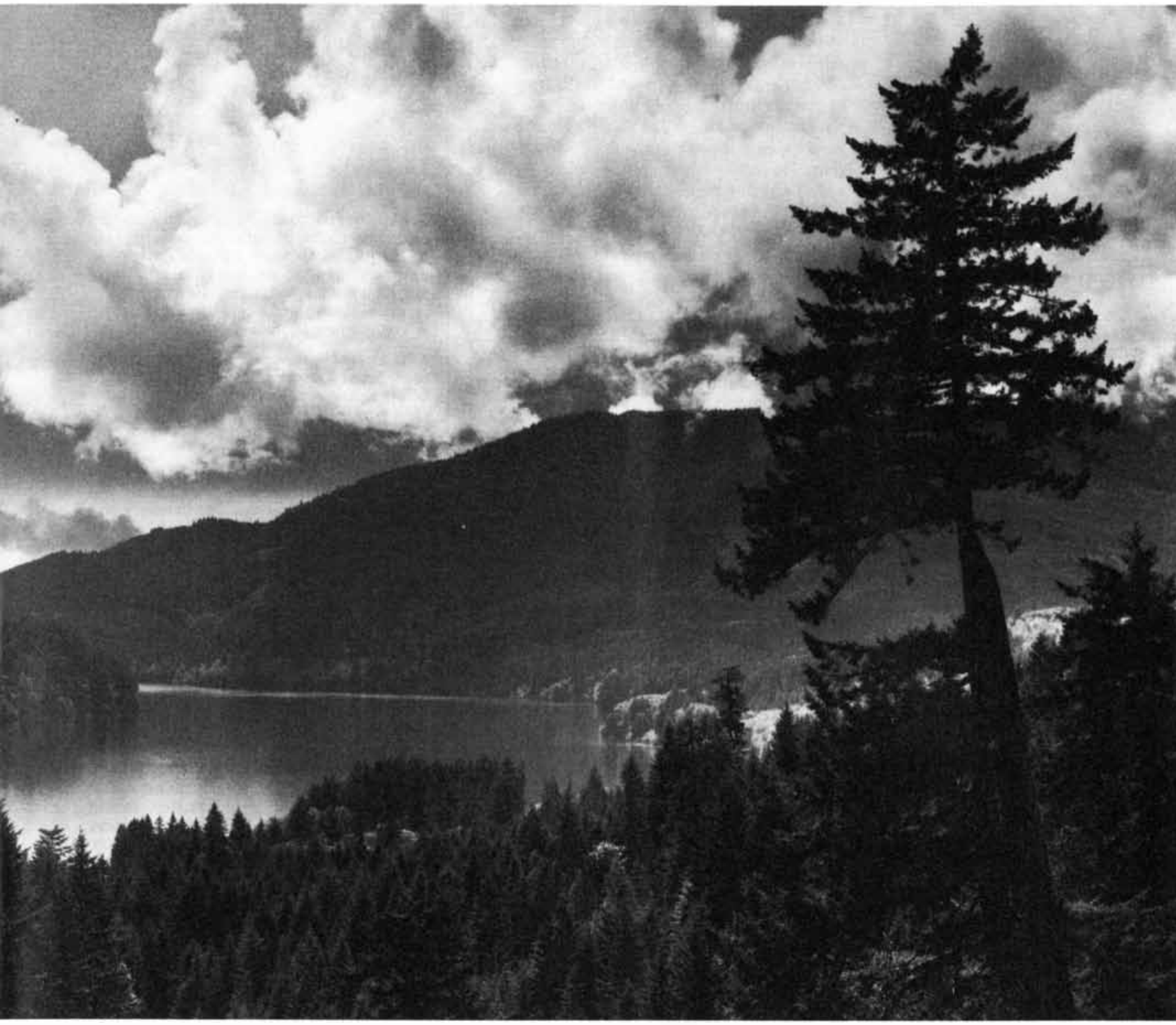
DEPARTMENTS

"BY 1945, AMERICA'S FORESTS WILL BE HISTORY."

In 1920, authorities predicted we'd run out of forest land in 25 years. But today we have nearly *730 million acres* of lush forest land in the U.S.—and more trees than we had 70 years ago.

Thanks, in part, to private landowners and America's forest products companies, who

BOISE CASCADE   BOWATER, INC.   CHAMPION INTERNATIONAL   CONSOLIDATED PAPER, INC.   GAYLORD CONTAINER CORP.   GEORGIA-PACIFIC   INTERNATIONAL PAPER



THE COVER painting is a cross-sectional view of the Mark II detector, which records events generated by the collision of high-energy electrons and positrons at the Stanford Linear Collider (see "The Stanford Linear Collider," by John R. Rees, page 58). The pipe in which the particles collide is seen at the center. A radial array of cables carries data from drift chambers to computers. The detector has found several hundred Z^0 particles, making it possible to refine estimates of the Z^0 mass.

THE ILLUSTRATIONS

Cover painting by Ian Worpole, based on a photograph by Peter Menzel

Page	Source	Page	Source
42-44	Andrew Christie	96	Jeff Foott
46	AT&T Bell Laboratories	97	Spencer C. H. Barrett
49	Mark S. Ladinsky, University of Colorado at Boulder	99	Olaf Wolter, IBM German Manufacturing Technology Center, Sindelfingen
50-52	George V. Kelvin	100	Hank Iken
53	George V. Kelvin (*top*), Matthew J. Schibler (*middle*), Kent L. McDonald (*bottom*)	101	Randall M. Feenstra, IBM Thomas J. Watson Research Center
54	George V. Kelvin (*top*), Hirohisa Masuda (*middle*), George V. Kelvin (*bottom*)	102	Hank Iken (*top*), Paul K. Hansma, University of California, Santa Barbara (*bottom*)
55	Kent L. McDonald (*top*), George V. Kelvin (*bottom*)	103	Hank Iken
56	Mark S. Ladinsky	104	Pierre Levy (*top*), Yves Martin, IBM Watson Center (*bottom*)
58-61	Ian Worpole	105	Hank Iken
62	Stanford Linear Accelerator Center (*left*), Ian Worpole (*right*)	106-107	James Mellaart, Pictures of Record, Inc.
63	Ian Worpole (*top*), John Seeman, SLAC (*bottom*)	108	Andrew Christie
64	Peter Menzel	109-110	Johnny Johnson
65	Ian Worpole	111	Andrew Christie
67-71	Dana Burns	112	Johnny Johnson
82-88	Thomas C. Moore	113	Laurie Grace
91	Steve Buchanan	114	Johnny Johnson
92	Spencer C. H. Barrett	116	Michael Goodman
93	Patricia J. Wynne	117	Jearl Walker
94	Johnny Johnson	118	Jearl Walker (*left*), Michael Goodman (*right*)
95	Noel D. Vietmeyer, National Research Council	119	Jearl Walker
		120-122	Hank Iken
		123	Mid-America Museum, Hot Springs, Ark.

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017. Second-class postage paid at New York, N.Y., and at additional mailing offices. Authorized as second-class mail by the Post Office Department, Ottawa, Canada, and for payment of postage in cash. Subscription rates: one year $27, two years $48, three years $66 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. only 800-333-1199; other 515-247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51593.

SCIENTIFIC AMERICAN

Established 1845

EDITOR: Jonathan Piel

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

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

COPY: Maria-Christina Keller, *Copy Chief;* Nancy L. Freireich; Michele S. Moise

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

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

ADVERTISING OFFICES: NEW YORK: Scientific American, 415 Madison Avenue, New York, NY 10017; Robert F. Gregory, *Advertising Manager;* Lisa Carden; Peter Fisch; John Grant; Meryle Lowenthal. CHICAGO: 333 N. Michigan Avenue, Chicago, IL 60601; Patrick Bachler, *Advertising Manager;* Litt Clark, *Midwest Manager.* DETROIT: 3000 Town Center, Suite 1435, Southfield, MI 48075; William F. Moore, *Advertising Manager;* Edward A. Bartley, *Detroit Manager.* WEST COAST: 1650 Veteran Avenue, Suite 101, Los Angeles, CA 90024; Kate Dobson, *Advertising Manager;* Joan Berend, San Francisco. ATLANTA: Quenzer/Stites. CANADA: Fenn Company, Inc. DALLAS: Griffith Group. PRINCETON: William Lieberman, Inc.

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

INTERNATIONAL: EUROPE: Roy Edwards, *International Advertising Manager,* London; GWP, Düsseldorf. HONG KONG/SOUTHEAST ASIA: C. Cheney & Associates. SEOUL: Biscom, Inc. SINGAPORE: Cheney Tan Associates. TOKYO: Nikkei International, Ltd.

PUBLISHER:
John J. Moeling, Jr.

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

PRESIDENT AND CHIEF EXECUTIVE OFFICER:
Claus-Gerhard Firchow

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

CHAIRMAN OF THE BOARD:
Georg-Dieter von Holtzbrinck

CHAIRMAN EMERITUS: Gerard Piel


© 1988 British Airways

LETTERS

To the Editors:

Elizabeth Corcoran's "A Truffling Matter" ["Science and Business," SCIENTIFIC AMERICAN, May] uncritically presents the unverified, entrepreneurial claims about "La Truffe" products as fact: "He hit on a growth medium that was conducive to growing truffles in a petri dish, [yielding] harvests of roughly 60 pounds per week.... Truffle oil, paste, powder and juice [are produced]." The producers have neither revealed the process nor, to my knowledge, released any live mycelium for examination by others.

A variety of "La Truffe" products, claimed on the labels to contain *Tuber melanosporum,* have been microscopically examined by me, Ruth Taber of Texas A&M University, Joseph Ammirati of the University of Washington and Gerard Chevalier of France's Institut National de la Recherche Agronomique station at Clermont-Ferrand, a foremost authority on the biology and culture of the Périgord truffle. To us, "La Truffe" products smell like generic dried fungi. Rehydrated or cooked, they lack the unique fragrance that typifies fresh Périgord truffles. In international commerce, a truffle is defined as the truffle fungus's sexual fruiting body, containing asci and ascospores. "La Truffe" products we have examined lack these. Extraction and preliminary separation of DNA from "La Truffe" products and its comparison with DNA from authentic Périgord truffles indicate significant differences. We do not know what fungal material is in "La Truffe" products; it is not fruiting bodies of Périgord truffles, as claimed on the label.

Truffles evolved as sexual fruiting bodies of obligate mycorrhizal fungi specifically adapted for spore dispersal by animals. As they ripen, these hypogeous sporocarps produce aromatics to attract animals. The animals eat them, digesting all the tissue but the spores. The spores are dispersed in fecal "packages" of spore inoculum that is washed by rain into the soil. When the spores germinate near receptive host roots, the new mycelium can establish a symbiosis with the host tree. In nature, the fungus cannot fruit in the absence of the host and probably cannot survive.

The complex physiological interactions among truffle fungus, tree host, animal disperser and environment result from millions of years of evolution and have thwarted attempts to produce truffles under artificial conditions or in the absence of host roots. Nothing in "La Truffe" products suggests that the barrier to artificial cultivation has indeed been broken. Breakthroughs are welcome from any source, but those who claim breakthroughs must come forward with their evidence. The producers of "La Truffe" have yet to do so.

Meanwhile, buyers who believe the "La Truffe" label will assume they are getting true *T. melanosporum* and will form a false, inferior impression of the qualities of Périgord truffles. This could seriously damage the market for European and American producers of the real thing, producers who have invested much in careful research, openly published and tested, on truffle cultivation. Many fungi produce toxins and carcinogens; an unknown fungal product ought not to be marketed until its content and safety are reasonably confirmed. *Scientific American* ought not to promote unsubstantiated commercialism.

JAMES M. TRAPPE

Departments of Forest Science
and Botany-Plant Pathology
Oregon State University, Corvallis

To the Editors:

California Truffle Company (which formerly used the trade name "La Truffe") has never claimed to produce sexually mature fruiting bodies of *Tuber melanosporum,* Dr. Trappe to the contrary. The California Department of Public Health has determined, after a thorough examination of the company's truffles and growing conditions, that our products are indeed "cultured truffles, *Tuber melanosporum*" and that they have progressed beyond the mycelial stage, although they cannot be characterized as sexually mature. The term cultured is used to distinguish our truffle from its European counterpart, dug from the ground. The Department of Public Health has also determined that our products are entirely safe for human consumption. Dr. Trappe's preliminary separation of DNA from our cultured truffles is not meaningful without data, especially given that the samples analyzed were taken from processed, dehydrated materials. We are a commercial enterprise engaged in producing gourmet food and not an academic institution. We therefore see no reason to reveal our process and have no intention of doing so.

As for Dr. Trappe's suggestion that buyers will form a false, inferior impression of truffles, our products have received enthusiastic kudos from a number of noted chefs, including Jack Czarnecki, author of *Joe's Book of Mushroom Cookery* and considered the premier mushroom chef in America; ex-White House chef René Verdon; Jean Michel Jeudy, vice-president of the California Culinary Academy; Gilbert Drouelle, executive chef of Sofitel; and Alex Errecarte of Brasserie Chambord. Contrary to Dr. Trappe's assertion that the existing truffle market will be damaged, we believe that making our cultured truffle products available throughout the year to a wider market will improve awareness and appreciation of truffles.

MOSHE SHIFRINE

President, California Truffle Co.
Woodland, Calif.

To the Editors:

I very much enjoyed "The Trireme Sails Again," by John F. Coates [SCIENTIFIC AMERICAN, April]. It calls to mind some sequences of Aristophanes' play *The Frogs.* There Aeschylus says (verse 1073): "Then, when I was living, the oarsmen knew nothing but to talk of their allowance of bread and to sing out their yo-ho!" In verse 207, while crossing a lake in a boat with two oars, Dionysos orders: "Then, beat time!" and Charon follows with "O opop, o opop!" These references indicate that crews must have synchronized their oars by shouting.

In verse 1074, Dionysos says: "By Apollon, so it is, and to break wind into the faces of those oarsmen at the lowest bank...!" From the illustrations on pages 102 and 103 it is clear why Aristophanes specifies the lowest oarsmen: *thalamakes.* Such a practical joke can work only between the second and third banks of oarsmen.

EMANUEL PFEIL

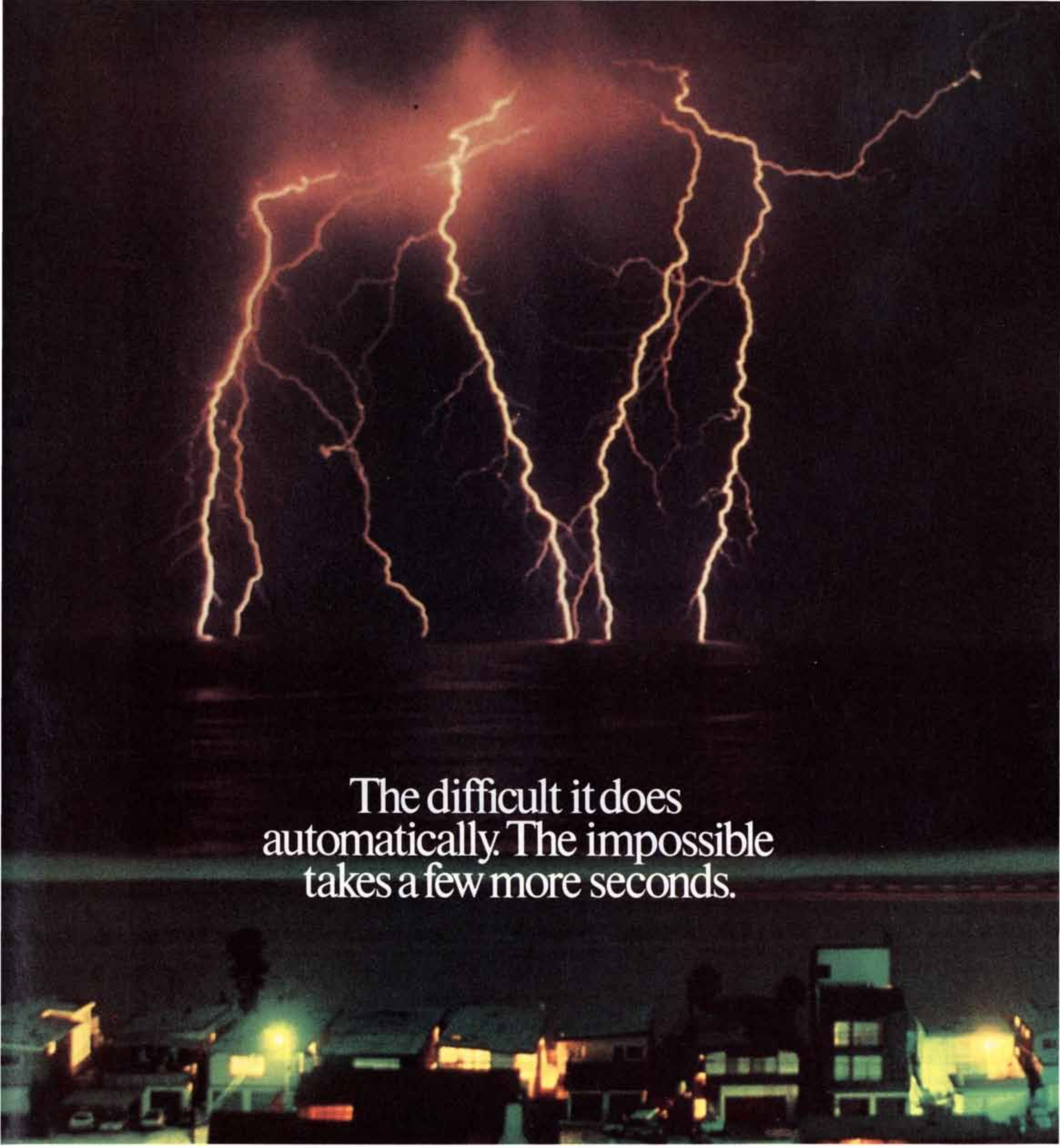
Marburg, West Germany

To the Editors:

In reference to "Aging Comes of Age" ["Science and the Citizen," SCIENTIFIC AMERICAN, May], wasn't it Ponce de León who was said to be looking for the Fountain of Youth and not Hernando de Soto (who was seeking wealth)?

ALICE R. BENSON

Ann Arbor, Mich.


The difficult it does automatically. The impossible takes a few more seconds.

In a world of the flighty and fly by night,

It's our signature, of course. Our sign-off on a product. It is an invitation to your confidence, and if you like the product, it's an R.S.V.P. If for some reason, the product disappoints, you'll move us from your shopping list to your black list. We can't have that, so we see to it that our label promises more.

Take one of our buttondown shirts.

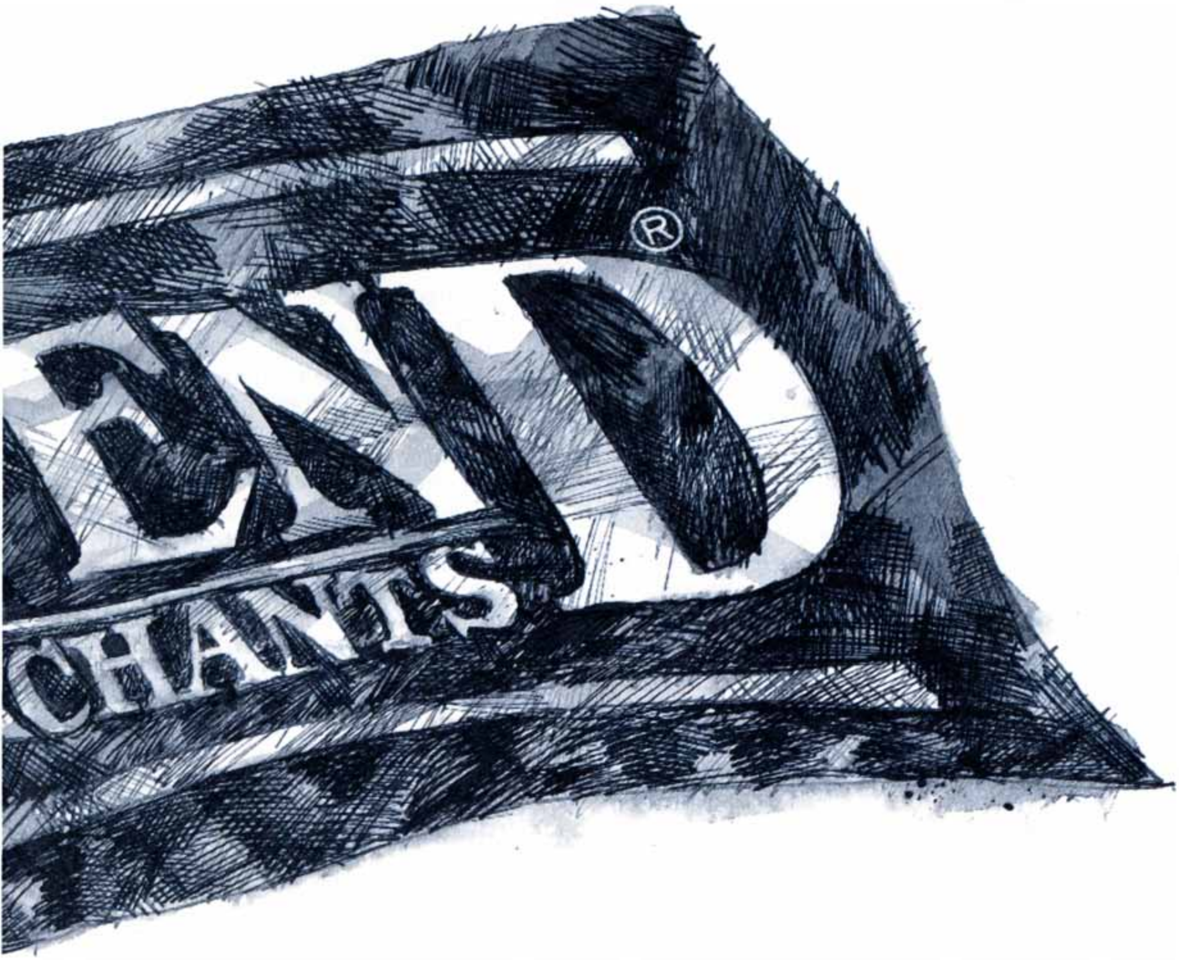
Inside the collar of any of our buttondowns, our label assures you a degree of value unparalleled in the market place. For several reasons.

First, because we're Direct Merchants, spared the mark-up middlemen make necessary. But also because we're not all things to all shirt wearers. We're buttondown people, because we're buttondown wearers. And we share with you what we insist on for ourselves: 100% cotton integrity, extra-long plackets, 7-button fronts, gauntlet buttons on the sleeves, a generous cut, and tails that go all the way to there, so they don't inch up.

Or, consider our label in a tie or a shoe.

Look at a Lands' End label in any of our ties and it says to you, instantly, "handsewn". We handle no other kind of ties. Beyond that, we put the label to exceptional use as a "keeper" that does what a keeper should—keeps the lower half of the tie in place instead of parting company with the tie after the first hurried tying.

In a shoe, the Lands' End label tells you that if your foot feels comfortable in it, it will always feel comfortable in any Lands' End shoe of that exact size, whatever the style. Why?



a label like this can be a wonderful thing.

Because we size our shoes to a common last. You can order our shoes with confidence. (All this without enduring the shoe salesman's hand-squeeze of your metatarsal arch, or the pinch of his thumb and forefinger on your sensitive big toe!)

Not to forget our guarantee.

As vigilant as we are in guarding the integrity of our label, we are not perfect because our business is more art than science, and hence (thank goodness!) a people-intensive business. So behind our label stands our blunt guarantee: GUARANTEED. PERIOD.® There is no conceivable condition under which we do not honor it.

So, give us a call at 1-800-356-4444, and start enjoying the Lands' End experience as millions of people like you (and us!) already do.

Call toll-free:

1-800-356-4444

LANDS' END DIRECT MERCHANTS®

©1989, Lands' End, Inc.

Or write:

Lands' End, Inc. Dept. Q-30
Dodgeville, WI 53595

Name ____________

Address ____________

City ____________

State ________ Zip ________

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

OCTOBER, 1939: "If atom smashing could be made more efficient, power production by means of nuclear fission would not be beyond the realm of possibility. But under present conditions, the process is as inefficient as removing the sand from a beach a grain at a time. The secondary neutrons emitted at the moment of fission and in later reactions, however, bring up an interesting and rather disturbing aspect of the case. These secondary neutrons constitute a fresh supply of 'bullets' to produce new fissions. Thus we are faced with a vicious circle, with one explosion setting off another, and energy being continuously and cumulatively released. It is probable that a sufficiently large mass of uranium would be explosive if its atoms once got well started dividing. As a matter of fact, the scientists are pretty nervous over the dangerous forces they are unleashing and are hurriedly devising means to control them. It may or may not be significant that, since early spring, no accounts of research on nuclear fission have been heard from Germany—not even from discoverer Otto Hahn. It is not unlikely that the German government, spotting a potentially powerful weapon of war, has imposed military secrecy on all recent German investigations."

"It is 25 minutes' truck running time from the Camden Airport, New Jersey, to the Philadelphia Post Office at 30th and Market Streets. A Kellett KD-113 Autogiro makes the trip by air in five or six minutes and shuttles back and forth from the Airport to the roof of the Post Office five times daily, carrying a full load of mail each time."

"A new plastic, developed by the H. S. Polin Laboratory of Research in Physics, is unique, particularly from the standpoint of use in South America, in that it is made wholly from the green coffee bean and requires no additional raw materials. Coffee-plastic making will be a highly self-contained industry, because the coffee provides its own chemical plasticizers and catalysts and its own filler material."

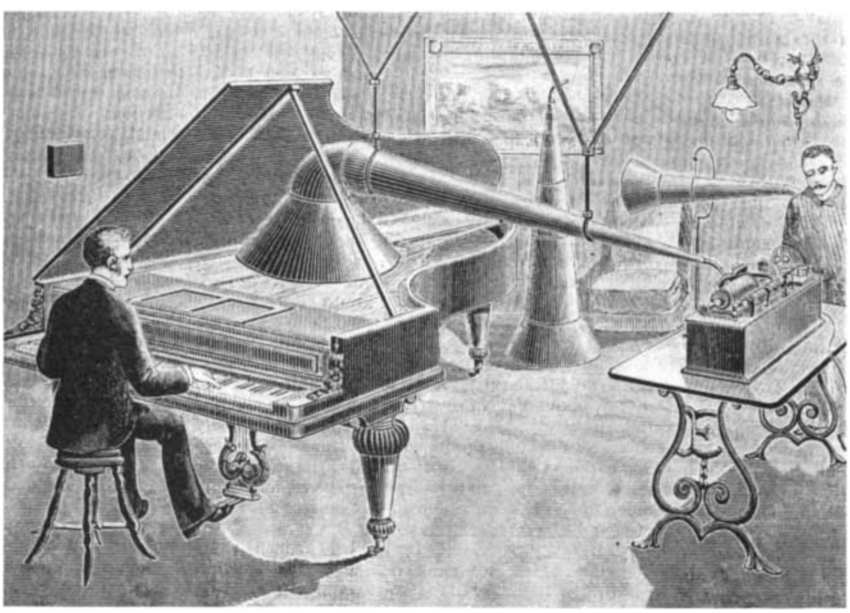
SCIENTIFIC AMERICAN

OCTOBER, 1889: "Still the acquirement of large American plants by foreign concerns goes merrily on, says *The Iron Trade Review.* Not to speak of the breweries and land companies and Western mortgages recently absorbed, the important news of the day is the proposition of an English syndicate to buy the extensive works of the Thomas Iron Company, at Hokendauqua, Pa. These movements again press to the front the question: Is there any danger to our protective system in the absorption of American concerns by foreign capitalists? The general consensus seems to be that we can stand any amount of influx of foreign capital, but it should be distinctly understood that, once invested in America, it becomes American capital, bound to respect American interests, and pay the American scale of wages."

"Atlantic racing has risen to the dignity of an art developed and aided by the resources of science. The hulls of such ships as the Teutonic, the Etruria, the City of New York, or the Columbia have been designed not only to slip through the water with the least possible effort, but to withstand the worst assaults of sea and wind. Within they are palaces, without they are castles. Their engines and boilers are the most perfect as well as the most gigantic examples of steam machinery. They are sailed by men to whom the Atlantic is as well known as Fleet Street to a policeman. In their engine rooms are men who have absolutely nothing left to learn in the art and mystery of getting the last foot-pound of useful work out of their machinery. Keen rivalry prompts the driving of these great ships across the ocean as fast as they can possibly go."

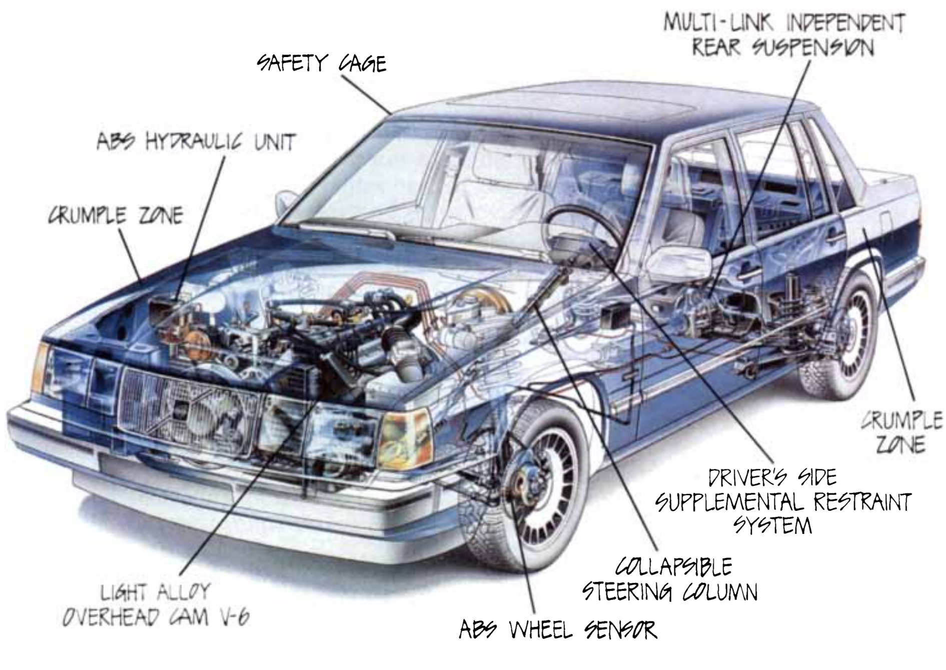
"The centenary of the discovery of uranium has been marked by the finding of a continuous lode at the Union Mine, Grampound Road, Cornwall, which is believed to be the only known lode in the world. It is anticipated that the present discovery will enable two important applications of the metal to be followed up. The first is as a substitute for gold in electroplated ware, inasmuch as with platinum and copper it forms two beautiful alloys, each having the appearance of gold, and the former also resisting the action of acids. The second application is in connection with electrical installations, where its usefulness consists in its high electrical resistance."

"*La Nature* reports from the Paris Exhibition: 'When one has heard the new, improved phonograph speak, he is astonished at the distinctness exhibited in the reproduction of piano and wind instrument music. It has seemed to us of interest to depict the means employed for registering the airs obtained by the aid of these musical instruments. The figure shows the immense ear trumpet which leads the sounds of a grand piano to the wax cylinder of the phonograph.'"



Apparatus for registering piano music by the phonograph

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



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

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

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

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

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

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

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

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

VOLVO

A car you can believe in.

STA FELLOWSHIPS

Postdoctoral Fellowships in Japanese Government Laboratories

The Japanese Science and Technology Agency (STA) is offering postdoctoral fellowships to scientists and engineers of the countries listed below for periods of 6 months to 2 years to be held in any Japanese national laboratory (excluding university and university-affiliated laboratories). Over one hundred Japanese research laboratories covering almost all areas of science, engineering and medicine are participating in the scheme.

The fellowships are open to young PhD holders of under 35 (although older researchers will be considered) from universities, research councils, government research laboratories and industry. Any science or engineering discipline will be considered except military R & D. Applicants will be required to supply a letter of invitation from their Japanese host institution (the organizations listed in Table 1 provide help in contacting suitable host institutions).
Those in the final stages of a PhD may also apply.

There are no closing dates but candidates are encouraged to submit their applications as soon as possible. Fellowships for fiscal 1989 must be taken up by March 1990.

FELLOWSHIP AWARDS

Fellowships include round-trip air tickets (economy class) and the following tax-free allowances:

1) Living allowance: ¥270,000 (about US$2,000) a month
2) Family allowance: ¥ 50,000 a month
3) Housing allowance: up to ¥100,000 a month
   Apartments will normally be provided to awardees. The apartment floor area is 40 m^2 for awardees unaccompanied by their family and 60 m^2 for awardees to stay with their family in other than metropolitan areas such as Tokyo. If an awardee prefers to use a larger apartment because of the family size or otherwise ¥100,000 maximum per month may be paid as housing allowance; any shortfall is to be borne by the awardee.
4) International relocation allowance: ¥200,000
5) Travel allowance: (within Japan) ¥100,000 a year
6) Japanese language
   and their family members in Tsukuba area. Those who live in places other than the Tsukuba area will be entitled to reimbursement of Japanese language school tuition up to a specified amount.
7) Excursions or the like will be held to help make the Fellowship awardees and their family members in Tokyo/Tsukuba areas familiarized with Japan's culture, tradition and history.

In addition, ¥1,480,000 per year will be paid to the host institute to cover research costs and insurance for researchers will be paid by JISTEC during their stay in Japan to cover medical care.

APPLICATION PROCEDURE FOR STA FELLOWSHIP

Management of the STA Fellowship, including recruitment of candidates, is entrusted to the Japan International Science and Technology Exchange Center (JISTEC).

Responsible organizations overseas which represent the governments of their respective countries are given in Table 1.

A researcher wishing to be awarded the STA Fellowship should apply to the responsible organization in his/her country. Candidates are required to contact the desired host institute and obtain a letter of acceptance before filling an application with their government. Further information regarding the STA Fellowship and host Institutes is available from the responsible organizations.

A researcher whose country is not listed in Table 1 could contact a Japanese host institute directly, which in turn may recommend the reseacher to JISTEC as a candidate for STA Fellowship.

Fig. 1 shows the process from consult/contact to receipt of award. (W6352)E

Fig. 1 **PROCESS FLOW FROM CONSULT/CONTACT TO RECEIPT OF AWARD**

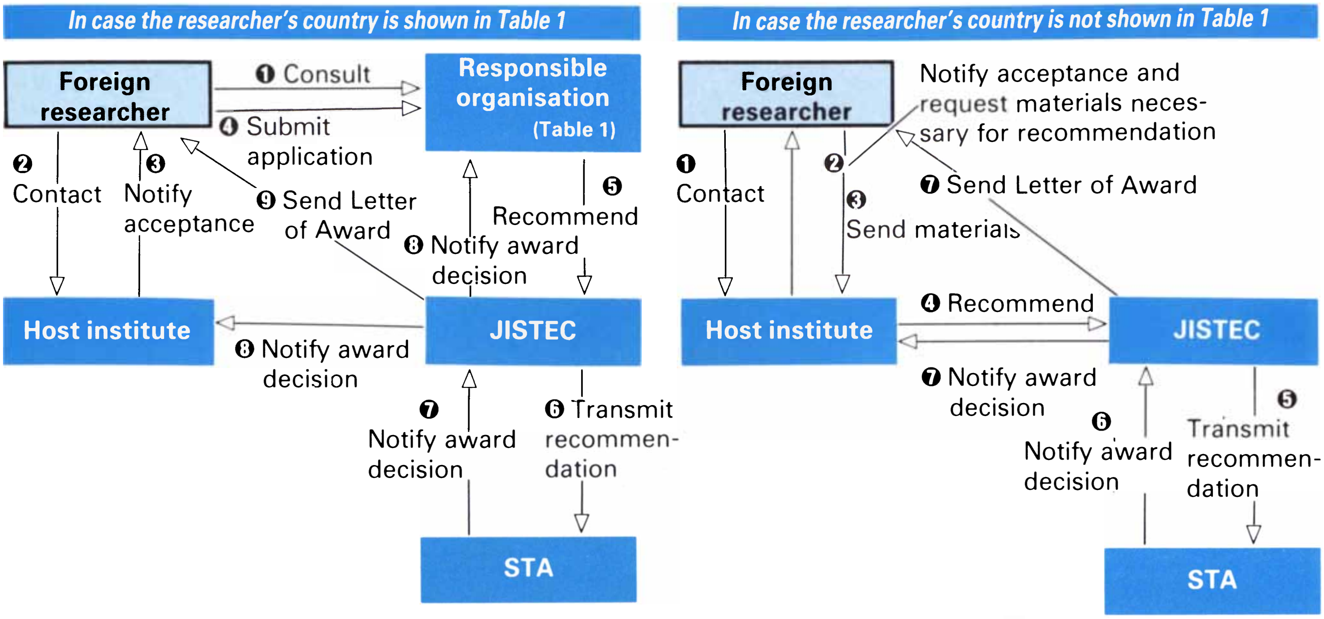


Table 1 **RESPONSIBLE ORGANIZATIONS**

Country	Contact
Australia	DEPARTMENT OF INDUSTRY, TECHNOLOGY AND COMMERCE The Secretary (Attention: Assistant Secretary, Japan Branch) GPO Box 9839, Canberra ACT 2601 Tel: 062-76-1000 Fax: 062-76-1122
Canada	NATURAL SCIENCE AND ENGINEERING RESEARCH COUNCIL Dr. R.J. Kavanagh Director-General (Scholarships & International Programs) Centennial Towers 200 Kent St., Ottawa, Ontario K1A 1H5
Federal Republic of Germany	ALEXANDER VON HUMBOLDT-STIFTUNG Dr. Rolf Hoffmann Selection Department Jean-Paul-Strasse 12 5300 Bonn 2, FRG Tel: (0228) 833-0 Fax: (0228 833-199
Finland	MINISTRY OF TRADE AND INDUSTRY Division for International Affairs Mr. Pertti Valtonen Head of Division Aleksanterinkatu 10, SF-00170 Helsinki
France	CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE Mr. Stuyck Taillandier Direction des Relations et de la Cooperation Internationales 15 Quai Anatole France, 75700 Paris Tel: 1-47 53 1515
Italy	MINISTERO DELLA RICERCA SCIENTIFICA E TECNOLOGICA Ufficio Relazioni Internazionali Dr. Mario Bove Director Lungotevere Thaon di Revel 76, 00100 Rome Tel: 6-369-941 Fax: 6-392-209 Tlx: 612548 RISCIE 1
Netherlands	MINISTRY OF EDUCATION AND SCIENCE Dr. ir. B. Okkerse Director-General for Higher Education and Scientific Research Dr. P. van't Klooster Deputy Director Division Research Organisations Directorate-General for Higher Education and Scientific Research P.O. Box 25000, 2700 LZ Zoetemeer
New Zealand	INTERNATIONAL SCIENCE UNIT Dept. of Scientific & Industrial Research Mr. M.A. Collins Assistant Director-General P.O. Box 1578, Wellington
Sweden	STYRELSEN FOR TEKNISK UTVECKLING Dr. Erik von Bahr Box 43200, 100 72 Stockholm Tel: 08-775 40 00 Fax: 19 68 26 Tlx: 10840 swedstu s
Switzerland	SWISS NATIONAL SCIENCE FOUNDATION Mr. Benno Frey Wildhainweg 20, CH-3001 Bern Tel: 031-24-54-24 Fax: 23-30-09 Tlx: 912-423
United Kingdom	THE ROYAL SOCIETY Ms. Karen Kimpton or Dr Stephen Cox 6 Carlton House Terrace, London SW1Y 5AG Tel: 01-839-5561 Tlx: 917876
United States	NATIONAL SCIENCE FOUNDATION Dr. Charles W. Wallace Senior Program Manager US-Japan, Australia and New Zealand Programs Division of International Programs Washington, D.C. 20550 Tel: 202-357-9558
European Communities	THE COMMISSION OF THE EUROPEAN COMMUNITIES Mr. Giorgio Boggio, DG XII Head of Division, Rue de la Loi 200, 1049 Brussels Tel: 235-5635

QUESTIONS ABOUT STA FELLOWSHIP

Please direct questions about the fellowship scheme to the responsible organization in your country given in Table 1. If your country is not listed, inquiries will be received by JISTEC.

Japan International Science and Technology Exchange Center (JISTEC)

Address: Port One Building 6F, 1-7-6, Minato-machi
Tsuchiura City, Ibaraki Pref. 300
Japan
Telephone: 0298-24-3355
Facsimile: 0298-24-3214

SCIENCE AND THE CITIZEN

Land-Locked

Increasingly, experts question the need for land-based missiles

Fred C. Iklé had to curb his tongue in public for the seven years he served as Ronald Reagan's under secretary of defense for policy. But in a speech shortly after he left the Pentagon last year, he finally uttered the question that had long gnawed at him: "Why on earth, heaven and hell do we still want the land-based ICBM?"

Good question, one that arms analysts left and right are asking these days. It arises in discussions of how the U.S. should cut its forces in the event of a Strategic Arms Reduction Talks (START) treaty. It is also central to the long-grinding debate over how to make the land-based leg of the U.S. nuclear triad less vulnerable to an attack by its ever more accurate Soviet counterpart. The old "window of vulnerability"—remember?

Campaigning for the presidency, George Bush vowed to slam the putative window shut by deploying intercontinental ballistic missiles in "survivable" modes. (The U.S. now has 1,000 ICBM's based in silos: 950 Minutemen, each carrying up to three warheads, and 50 MX's, each with 10 warheads.) Bush was given two options: putting 50 MX's on railcars or hundreds of small, single-warhead Midgetman missiles on all-terrain trucks. Each alternative has powerful supporters: the Air Force brass and Dick Cheney, Bush's secretary of defense, favor the MX; Les Aspin, chairman of the House Armed Services Committee, and Brent Scowcroft, Bush's national security adviser, back the Midgetman.

Bush's choice? Both of the above. His decision may be politically sound, but it has been widely criticized as lacking strategic sense. Barry Blechman of Johns Hopkins University's Foreign Policy Institute points out that the MX and Midgetman each has a distinct drawback: MX missiles, which would roll out from a central garrison only during a crisis, would take hours to disperse over rail lines and so could be destroyed by a surprise attack; Midgetman missiles, although more elusive, would be much more expensive—up to $40 billion, or about three times as much as the MX plan. None of the 60 or so other schemes proposed to modernize the U.S. land-based missile force are any better, according to Blechman. "There is no good way to deploy survivable ICBM's on land," he maintains.

This inherent vulnerability, coupled with blazing speed and pinpoint accuracy, makes land-based missiles much more destabilizing than either bombers or submarine-launched missiles,

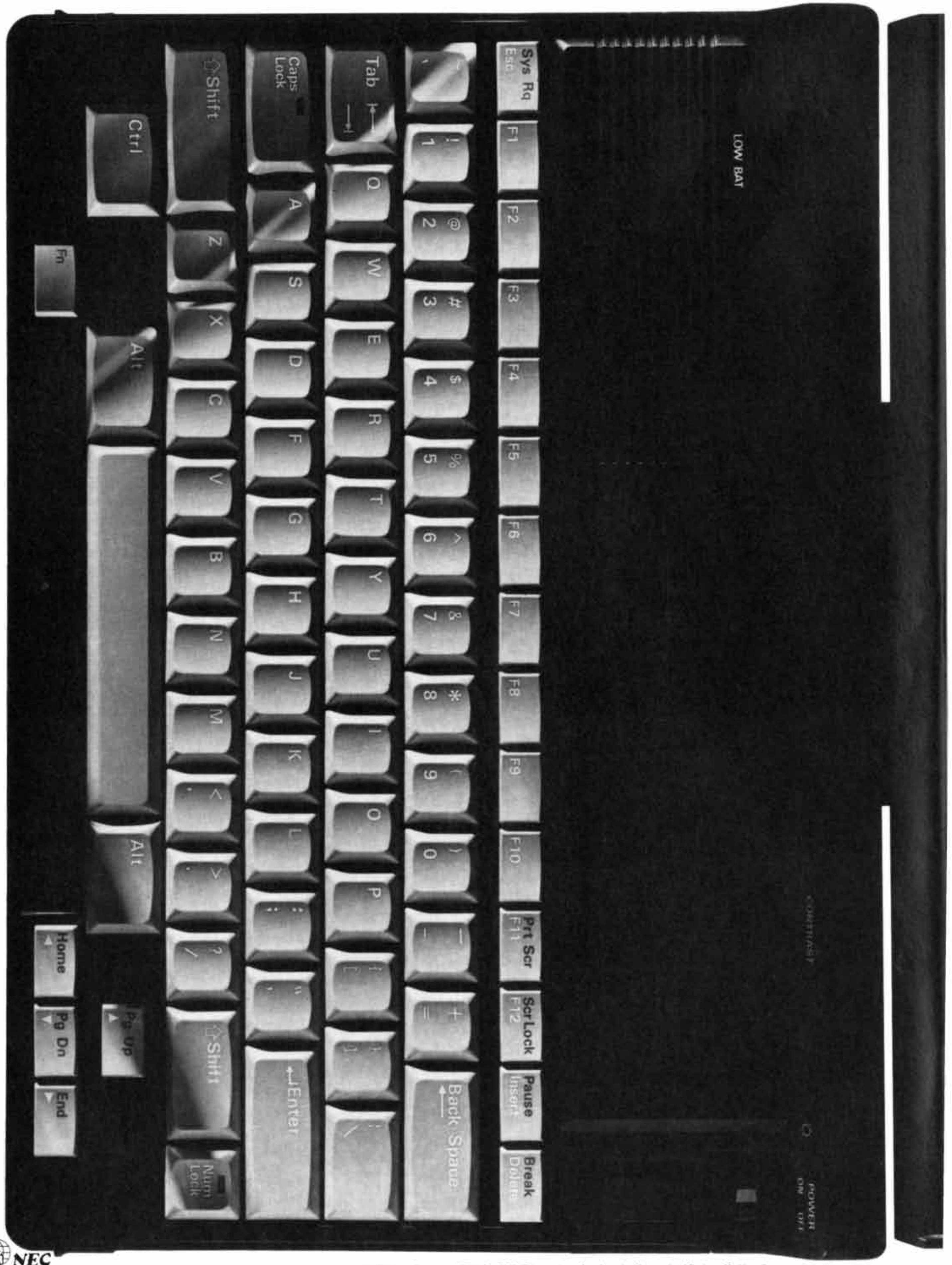
MIDGETMAN-MISSILE prototype was launched from Vandenberg Air Force Base in Calif. in May.

notes Marshall Brement of the Naval War College. The ICBM's represent both the best weapon for a first strike and the most tempting target for a preemptive attack, Brement explains. Moreover, he says, if U.S. satellites give warning of a Soviet attack, the President has about 10 minutes to "use or lose" his land-based missiles; once fired, they cannot be recalled. In contrast, bombers—once they are aloft—and submarines are relatively safe; they provide the President with more flexibility and time.

Iklé—like Brement and Blechman—thinks the U.S. should consider restructuring its strategic triad (or "tripod," as former House Speaker Tip O'Neill once put it) as a diad that would consist primarily of bombers and submarines. Submarines are now invulnerable to a surprise attack. The question is whether a technological advance might someday make the oceans "transparent" to Soviet satellites or other sensing devices. Such a breakthrough is not even on the horizon, according to Iklé. If it did occur, he notes, the Soviets would still need the ability to destroy all the submarines at one blow. If only one Trident submarine escaped, perhaps by hiding under the polar ice cap, it could launch up to 240 warheads in retaliation.

As a backup, there are still the bombers. The Air Force keeps about a third of the 300 or so B52's, F111's and B1B's that make up its strategic bomber force revved up and ready to fly within 15 minutes—enough time to detect and elude most surprise attacks. The Soviets could beat the clock with ballistic missiles that reach targets faster by flying at a shallower angle; launched from submarines near U.S. shores, these so-called depressed-trajectory missiles could reach air fields anywhere in the U.S. in under 15 minutes. But this possibility may soon be eliminated through arms control. A bipartisan group in Congress recently proposed that the administration negotiate a mutual ban on tests of depressed-trajectory missiles.

Brement advocates doing away with land-based missiles entirely. He says Bush could seize the public-relations initiative from Mikhail Gorbachev by calling for a mutual ban on all land-based missiles. Although the Soviet Union is much more dependent on land-based missiles than the U.S. is, Brement contends that "given recent



LOW BAT
Sys Rq
Esc
F1
F2
F3
F4
F5
F6
F7
F8
F9
F10
Prt Scr
F11
Scr Lock
F12
Pause
Insert
Break
Delete
Back Space
Tab
Caps Lock
Shift
Ctrl
Fn
Alt
Alt
Enter
Shift
Num Lock
Home
Pg Dn
Pg Up
End
POWER
ON OFF

history there is a chance Gorbachev will go for it." But for most other U.S. strategists, old habits apparently die hard. Iklé, despite his distaste for land-based ICBM's, still thinks the U.S. should hold on to a few. So does Blechman—just in case the U.S. someday needs their reliable, precise killing power, he explains.

Blechman's position dismays Robert S. McNamara, who served as secretary of defense from 1961 to 1968. He asserts that strategic nuclear arms should never be considered as war-fighting instruments but only as a deterrent. In the early 1960's McNamara himself set forth what is still the standard definition for a deterrent: it should be able to survive an all-out Soviet attack and still inflict "unacceptable damage"—McNamara felt the obliteration of 50 percent of the Soviets' industrial capacity and 25 percent of its population would suffice—in return. Some military analysts call the nuclear firepower needed to do this job an "annihilation unit." Others, in particular Soviet experts, prefer "McNamara unit," or just "McNamara." According to McNamara (the man), the U.S. does not need the MX or Midgetman—or the "Stealth" bomber, for that matter—to retain that capability for the foreseeable future. *—John Horgan*

Batteries Not Included

The space station faces yet another budget crunch

The call by President Bush for a manned expedition to Mars, a long-term goal that would use a space station and a permanent moon base as stepping stones, was what the National Aeronautics and Space Administration wanted to hear. "All roads begin with Space Station *Freedom*," proclaims a space-agency handout. Yet the call to the new frontier did not stir hearts and minds on Capitol Hill. The same day President Bush affirmed the Mars goal, the House of Representatives voted a 20 percent cut in the space-station budget this year, from $2.05 billion to $1.65 billion, and further cuts are possible.

NASA officials hold out some hope that the Senate may restore part of the cut, but they are taking no chances. Previous plans for the baseline space station, known as level 1, envisaged a crew of eight. Among the options now being considered, which have been informally dubbed level zero, is a scaled-down version that would have a crew of only four when permanently occupied in 1996. Station power would be halved to 38 kilowatts. Thrusters for keeping the station in its proper orbit might be fueled with hydrazine, a known technology, rather than with hydrogen and oxygen, a newer technology that offers better long-term efficiency.

Just as the space station finds a clear purpose, then, NASA must entertain cost-cutting options that would drastically reduce the amount of research that the orbiting laboratory could support. The scaled-back plans do not bode well for the future of manned exploration in space. Both the space station and a permanent base on the moon—the latter NASA officials maintain could be in place by the year 2001—are seen as valuable test-beds for long-duration life-support systems and for studies on the effects of extended space missions. In addition, the space station could serve as a base for assembling spacecraft that would shuttle cargoes to and from the moon and ultimately Mars.

NASA's defensive maneuvers have unsettled the foreign partners in the space station: Canada, the European Space Agency (ESA) and Japan. Their representatives were treated to a bureaucratic Freudian slip during a briefing in August on the level zero studies: a viewgraph that depicted only two modules in the initial configuration rather than four. ESA and Japan are each developing their own modules and had expected them to be part of the level 1 station. NASA officials say the viewgraph was a mistake and that no thought is actually being given to reducing the number of modules. The modules may be reduced in size, however, to cut the number of shuttle flights needed to assemble the station. Providing fewer facilities for the first occupants is seen as preferable to delaying the date of initial occupancy, one NASA analyst explains, because the station can grow later. "It's the same program," he says, "but it'll take longer to get there."

Quite apart from the budget retrenchment—should that prove necessary—the space-station design is being reevaluated in the light of the President's Mars goal. A concern that has been raised repeatedly is that a station designed to lead toward a lunar base and Mars expedition—which would have to include treadmills for life-sciences research, bays for assembling spacecraft, fuel-storage facilities and so on—is not the same type of station that would be designed for materials-science research. Life-sciences research tends to produce vibrations (from treadmills, for example), which are not at all desirable for materials science.

Franklin D. Martin, head of NASA's office of exploration, dismisses such technical criticisms as red herrings. He argues that careful scheduling should be able to avoid the worst clashes. The only obstacle he foresees is money. Martin points out that NASA's budget represented 4 to 5 percent of the federal budget at the peak of the Apollo program; the figure today is 1 percent. Sending astronauts to Mars, he estimates, would require 2 to 3 percent. "There's never a good time, but we're not a poor nation," he says. "It's very doable." *—Tim Beardsley*

Who's Minding the Store?

Central gene-research facility is nearly lost in NIH shuffle

As the U.S. biomedical establishment embarks on a plan to spend roughly $3 billion over the next 15 years to sequence the genomes of humans, animals and plants, it is scrambling to find a few million to store the information gathered and make it accessible to researchers. Bionet, a computerized data-base, bulletin-board and electronic-mail system that currently serves about a quarter of the U.S. laboratories supported by the National Institutes of Health, is to be shut down September 30 because its managers have not precisely fulfilled the agreement under which it was funded. They worked too hard at simply providing services and spent too little money on their own research.

According to Charles L. Coulter, program manager for Bionet at the NIH, a study team found that Bionet had not done enough software development—which he classed as research—to merit renewal of its grant. David Kristofferson of IntelliGenetics, Inc., in Mountain View, Calif., the company that runs Bionet, says his group was blindsided: they were instructed to get other funding for most of their research and ended up presenting only their more speculative projects to the NIH reviewers.

Richard H. Roberts of the Cold Spring Harbor Laboratory, a member of Bionet's advisory board, contends that the entire episode has been "a disservice to the molecular biology community." A central facility should never have been funded with such strings attached, he says. Unfortunately, the NIH twice rejected a proposal to establish Bionet through standard

WHAT'S BETTER THAN SPEED READING?

SPEED LEARNING

**Speed Learning has replaced speed reading. It's a whole new way to read and learn.
It's easy to learn . . . lasts a lifetime . . . applies to everything you read.
It may be the most productive course you've ever taken.**

Do you have too much to read and too little time to read it? Do you mentally pronounce each word as you read? Do you frequently have to go back and reread words or whole paragraphs you just finished reading? Do you have trouble concentrating? Do you quickly forget most of what you read?

If you answer "yes" to any of these questions—then here at last is the practical help you've been waiting for. Whether you read for business or pleasure, school or college, you will build exceptional skills from this major breakthrough in effective reading, created by Dr. Russell Stauffer at the University of Delaware.

Not just "speed reading"—but speed reading-thinking-understanding-remembering-and-learning

The new *Speed Learning Program* shows you step-by-proven -step how to increase your reading skill and speed, so you understand more, remember more and use more of everything you read. The typical remark from over one million people taking the Speed Learning program is, "Why didn't someone teach me this a long time ago." They were no longer held back by the lack of skills and poor reading habits. They could read almost as fast as they could think.

What makes Speed Learning so successful?

The new *Speed Learning Program* does not offer you a rehash of the usual eye-exercises, timing devices, costly gadgets you've probably heard about in connection with speed reading courses or even tried and found ineffective.

In just a few spare minutes a day of easy reading and exciting listening, you discover an entirely new way to read and think—a radical departure from anything you have ever seen or heard about. Speed Learning is the largest selling self-study reading program in the world. Successful with Fortune 500 corporations, colleges, government agencies and accredited by 18 professional societies. Research shows that reading is 95% *thinking* and only 5% eye movement. Yet most of today's speed reading programs spend their time teaching you rapid eye movement (5% of the problem) and ignore the most important part (95%) *thinking*. In brief, *Speed Learning gives you what speed reading can't.*

Imagine the new freedom you'll have when you learn how to dash through all types of reading material *at least* twice as fast as you do now, and with greater comprehension. Think of being able to get on top of the avalanche of newspapers, magazines and correspondence you have to read . . . finishing a stimulating book and retaining facts and details more clearly and with greater accuracy than ever before.

Listen-and-learn at your own pace

This is a practical, easy-to-learn program that will work for you—no matter how slow a reader you think you are now. The *Speed Learning Program* is scientifically planned to get you started quickly . . . to help you in spare minutes a day. It brings you a "teacher-on-cassettes" who guides you, instructs, encourages you, explaining material as you read. Interesting items taken from *Time* Magazine, *Business Week, Wall Street Journal, Family Circle, N.Y. Times* and many others, make the program stimulating, easy and fun . . . and so much more effective.

Executives, students, professional people, men and women in all walks of life from 15 to 70 have benefited from this program. Speed Learning is a fully accredited course . . . costing only 1/5 the price of less effective speed reading classroom courses. Now you can examine the same, easy, practical and proven methods at home . . . in spare time . . . without risking a penny.

Examine Speed Learning FREE for 15 days

You will be thrilled at how quickly this program will begin to develop new thinking and reading skills. After listening to just one cassette and reading the preface you will quickly see how you can achieve increases in both the speed at which you read and in the amount you understand and remember.

You must be delighted with what you see or you pay nothing. Examine this remarkable program for 15 days. If, at the end of that time you are not convinced that you would like to master Speed Learning, simply return the program and owe nothing. See the coupon for low price and convenient credit terms.

No-Risk Guarantee

If, after reviewing *Speed Learning* for 15 days, you are not delighted in every way, you may return the materials in the condition you received them *and owe nothing.*

"I was extremely impressed with the course, and with the techniques it taught me. By utilizing these methods, I have more time to spend on important work. I recently ordered another set for my son, a college student."

Mike Gravel
Former U.S. Senator

"*Speed Learning* is the most outstanding course I have taken. It has helped me so much in my business and personal reading that I plan to give it to all my employees."

Vincent Clementi
President
The Clementi Organization

FREE TRIAL ORDER FORM

☐ **YES!** Please send me the complete Speed Learning program for a 15-day Free Trial. After the Free Trial period, if I decide to take the course, I will pay for it in four monthly payments of $30.00* (usually tax deductible). If I decide not to improve my reading skills, I will return the program and owe nothing. **My complete satisfaction is guaranteed.**

☐ **SAVE!** I prefer to pay the first payment of $30.00 now, and save $7.00 shipping and handling charge. If I decide not to improve my reading skills, I will return the program for a full refund. **My complete satisfaction is guaranteed.**

☐ Check enclosed (payable to Learn Incorporated)

☐ Charge my ☐ Visa ☐ Master Card ☐ American Express ☐ Discover

Card # ________________ Exp. Date ________

Signature ________________ Phone (____)

Name ________________

Address ________________

City ________________ State ________ Zip ________

* Plus $7.00 shipping and handling. New Jersey residents, please add sales tax.

Clip this order form and mail to: learn INCORPORATED Dept DZ-27
113 Gaither Drive, Mt. Laurel, NJ 08054–9987

funding channels, and the "cooperative research agreement" that is now running out was the only way to get money at all.

Most of Bionet's services are to be taken over by GenBank, a separate NIH facility also operated by IntelliGenetics. (The shift, whose official authorization was still pending at press time, is being billed as a "natural evolution" of GenBank's capabilities.) Molecular biologists will be able to communicate with each other and retrieve stored gene and protein sequences, but they will be on their own when it comes to analyzing the data in detail. James C. Cassatt, the GenBank program manager at the NIH, says that researchers can buy analysis software that runs on their personal computers; some programs are even available gratis. Kristofferson, on the other hand, estimates that the cost of equipping all Bionet subscribers with equivalent computer hardware and software could go as high as $13.5 million. The price each user will be charged for GenBank's new services is still unsettled; Bionet cost a flat $400 a year, whereas equivalent commercial services charge $40 for an hour of computing.

Also unsettled is how GenBank or any other central repository will cope with the flood of data—at least 100 times the current amount—to be generated by the Human Genome Project and other massive gene-sequencing efforts. Many molecular biologists say that much of the money spent on sequencing genes will be wasted unless new methods and facilities are developed for retrieving and analyzing the data. GenBank's current contract runs for another three years, remarks Mark S. Guyer of the NIH Office of Human Genome Research: "We're only beginning to talk about what happens then." —*Paul Wallich*

PHYSICAL SCIENCES

Cosmic Quarrel

Fearless philosopher finds fly in the ointment of time

Theoretical physicists, equipped with counterintuitive perceptions and a formidable mathematical armory, are considered by philosophers to be armed and dangerous. Their turf—however interesting—is usually avoided. Huw Price, a philosopher from the University of Sidney, belongs to a different breed. He has taken on Stephen W. Hawking of the University of Cambridge, one of the world's leading cosmologists. The rift developed over whether Hawking has, as he claims, found a possible explanation for the arrow of time.

Time occupies a strange place in the cosmological scheme of things. Most physical laws would allow the universe to run equally well forward or backward. The major exception is a relentless tendency for the extent of disorder in the universe, or entropy, to increase. In his best-selling book *A Brief History of Time,* Hawking argues that the tendency toward entropy underlies the psychological experience that we know as time. He makes the connection by observing that living things can exist and record memories, thus gaining a sense of time, only by overcoming the rising tide of entropy within a local region. To do so, they have to use energy supplied by the sun. So, according to Hawking, the deeper question underlying our perception of time is: Why does the universe at this stage of its evolution contain ordered structures such as the sun rather than just total disorder—random radiation?

Like most cosmologists, Hawking believes the known forces of nature can account for galaxies, stars and other ordered systems only if there was a big bang that started the universe expanding rapidly from a very hot, dense and relatively organized state. Hawking and his supporters also think disorder will spread until the universe becomes a bland void that may eventually contract in a big crunch. Hence, for Hawking, the really deep mystery about time is: Why is the universe ordered at one pole of time (the one we call the big bang) but disordered at the big crunch? Why won't nature's epic film run backward?

Hawking finds his answer in a quantum-gravity model of the universe, a theoretical hybrid that combines relativity and quantum physics. He has proposed that cosmologists should focus their efforts on a particular type of quantum-gravity model, one in which the history of the universe is finite in extent and in time but has no boundary. Such a universe lacks an edge or wall because space-time is curved; a straight line eventually meets itself. Hawking calculates that this "no boundary" proposal, together with some other assumptions, leads to the grand conclusion that the real universe is by far the most probable type: ordered at one end of time, expanding to a maximum extent and then, aeons hence, contracting once more, yet becoming more disordered all the while.

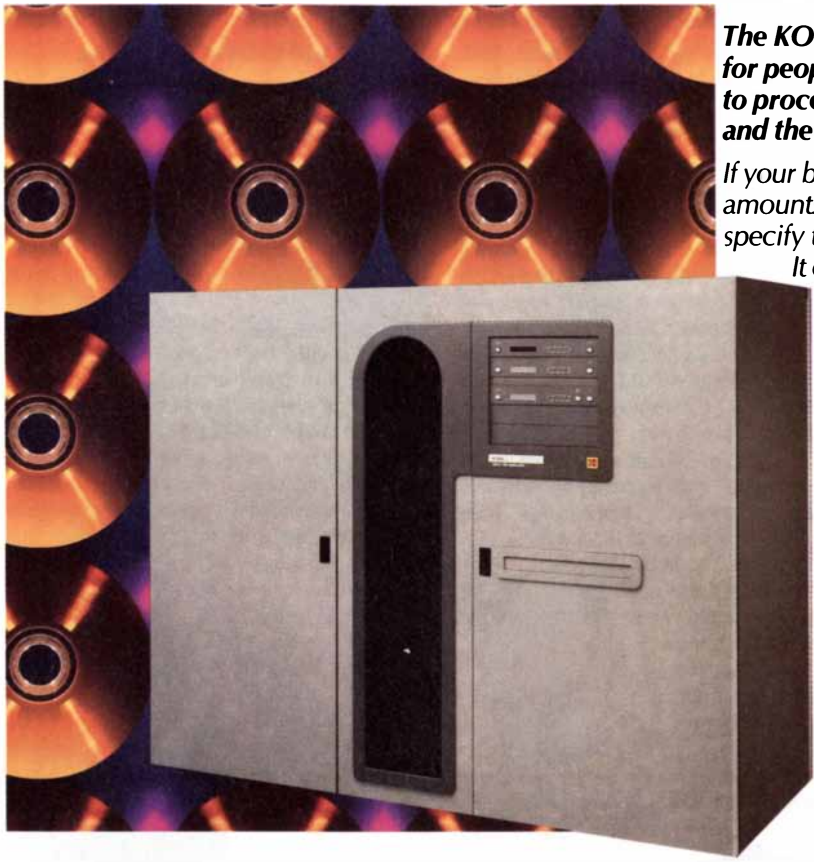
At this point Price charges into Hawking's cosmological briar patch. Writing affably in the pages of *Nature,* he suggests that *A Brief History of Time* fails to explain how Hawking finds the arrow of time embedded in his no-boundary proposal. The book, according to Price, teases the reader with a mystery that lacks a denouement. "It is as if we are assured that the butler did it, without being told how he overcame the evident obstacles (that he was incarcerated in Wormwood Scrubs at the time, for example)," he writes.

Price argues that Hawking's explanation for the arrow of time assumes what it sets out to show: that one end of time is different from the other. Price maintains that in a discussion about time, external final conditions are just as valid as initial conditions: any explanation for the high state of order at the moment of the big bang that is built into Hawking's model should, he suggests, apply equally to the big crunch. And if Hawking's model assumes some kind of temporal asymmetry, Price says, then it does not explain the arrow of time.

Hawking sticks to his guns. He writes to *Scientific American* that the no-boundary proposal does indeed explain the arrow of time because it predicts, essentially, that a universe of any given size has two distinct, highly probable degrees of disorder. In the low-disorder state, entropy increases as the universe expands; in the high-disorder state, disorder increases as the universe contracts. Hawking interprets the low-disorder state as belonging to the early history of the universe and the high-disorder state as coming after the universe has started to contract. "He does not have to assume time," one Hawking sympathizer comments. "If he is right, this answers the question."

Other cosmologists are not quite so sure. Don N. Page of Pennsylvania State University, who collaborates with Hawking, observes that "it's not clear what fraction of simple models of universes have an arrow of time." If the property of being simple itself predicted an arrow of time, then Hawking's no-boundary proposal might not explain very much.

Nevertheless, Page is confident that further work will clarify the situation. "It may be that Hawking's argument is wrong for other reasons, but I don't think it's because he's smuggling the arrow of time in there," he says. In fact, Page thinks, Hawking may have



THE LIFESTYLE RESOURCE™

The Lifestyle Resource gives you the facts necessary to make an informed purchase. Your satisfaction is our primary concern. If your purchase doesn't meet your expectations, return it in original condition within 30 days for prompt refund.

THE BEST CORDLESS PHONE JUST GOT BETTER

Southwestern Bell FF-1725 Offers–
- *1000-foot range*
- *Hearing aid compatible*
- *Hold button*
- *18-number memory*
- *Intercom*
- *Tone/pulse switchable*
- *Memory dialing*
- *Free 5-year service plan*

When a leading consumer magazine rated the FF-1700 Cordless Phone tops for range and quality amongst 21 brands and models in 1988, Southwestern Bell didn't rest on its cordless laurels. Instead, it went back to the drawing board and came up with something even better! The FF-1725 Freedom Phone breaks the "mediocre sound" barrier of cordless phones, combining the outstanding speech quality and convenience features of its predecessor with amazingly clean, crisp sound within a range of 1000 feet. The base unit serves as a freestanding speakerphone with its own dialpad, you get two phones in one. Take or make calls around your home — even next door! Plus an 18-number memory, intercom, paging, automatic redial, hold button and 10-channel selection. Personalized security code protects line from outside access. Free 5-yr. service plan. **$179.95** #2131.

Want even more freedom? The latest answer from Southwestern Bell is the FA-1525 Telephone Answering Machine. Features dual-cassette operation for quality recording. Pick up any household extension and auto-disconnect feature interrupts outgoing message and resets tape. Beeper or beeperless remote operation, so you'll never miss an important call. Plus two-way record for conversations, call screening, household memo function and voice-activated record. Soft-touch keys are designed for easy operation. **$129.95** #2141.

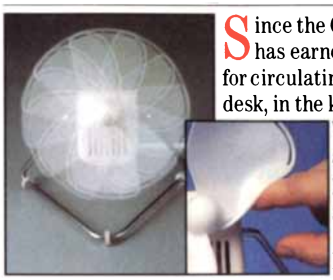
CAB DRIVERS' SECRET

A New York City cab driver tipped us off about the Wooden Bead Seat's "*mystical*" massaging action. It enables you to sit for long periods of time without discomfort or fatigue — whether you're at a desk or in traffic. It works like modern-day acupressure, gently *massaging* your back and legs, stimulating tired muscles and improving circulation. It creates a "zone of comfort" between you and the seat, allowing air to circulate so you stay cool in summer and warm in winter. The Wooden Bead Seat has health benefits, too. It's *recommended* by chiropractors because it supports the lower back and helps improve posture. The smooth lacquered wood beads are handstrung with flexible, heavy-duty nylon cord. Our Wooden Bead Seat will keep you comfortable all year long. **$29.95** #2690; Share the comfort — two for **$49.95** #2700.

NEVER A DULL MOMENT

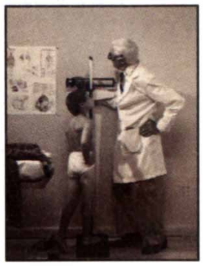
The new Chef's Choice® Professional Model 110 sharpener is on the cutting edge of knife-sharpening technology, honing kitchen, pocket, fillet and sporting knives to better-than-razor sharpness in minutes...over the entire length of the blade! Patented bi-level magnetic guides hold knives at the precise sharpening angle, eliminating guesswork. Diamond abrasives safely sharpen steel or alloy knives in three stages to form a perfectly smooth, triple bevel edge that will stay sharp longer than conventional edges. World class chef Pierre Franey in his *N.Y. Times* column, hailed the Chef's Choice® as "unparalleled for putting an exceedingly sharp and long-lasting edge on knives of all kinds." Manufacturer's 2-year limited warranty. UL listed, measures 9″ x 3⅝″ x 4½″. **$79.95** white #2451; black #2455.

THE CLASSIC COMPACT FAN



Since the Caframo Desk Fan was introduced only a few short years ago, it has earned its reputation as the best little compact fan. It can't be beat for circulating and freshening the air around you. Goes anywhere—on your desk, in the kitchen, out in a weekend cabin, or near a radiating woodstove. Whisper quiet, almost unnoticeable, highly energy efficient, and safe. Its patented blade-edge design absorbs impact and prevents injury if accidentally touched. And it's well-made, to give you years of steady service. Two-speeds; tilts to any angle; rubber feet prevent "crawling". Only 6½″ high, 33 oz. light. UL listed. **$39.95** #2600.

WORTH WEIGHTING FOR



For over 60 years, Health o Meter® has been the leading manufacturer of scales for those who demand 100% accuracy — doctors and other medical professionals. Chances are, you've been weighed on one of them! Now Health o Meter has incorporated the advanced technology, superior workmanship and unsurpassed accuracy of its finest physician-quality scales into a personal scale. The Professional Dial Scale features a huge (8″ diameter), easy-to-read speedometer dial with large, bold numbers. A heavy-duty professional rack and pinion weighing mechanism ensures lifelong accuracy and durability on any surface, and massive heavy-gauge steel platform and base support up to 325 pounds. Color-coded weight minders help you work towards your goal. With its tough, non-skid textured vinyl mat and sturdy feet, the Professional Dial Scale is equally at home in the bathroom, bedroom or exercise room. Full 5-year manufacturer's warranty. **$99.95** #2620.

FRESH AIR MACHINES

Bionaire 1000

The Bionaire 500 is like moving a mountain breeze indoors. Removes up to 99% of all particulate air pollutants: soot, dust, animal dander, pollen, cigarette smoke, smog, molds and fungi. Everybody breathes more healthfully including allergy sufferers. The patented electret filter removes particles 1/10,000 the thickness of a human hair. Helps clean an average (100 sq. ft.) room three times an hour. Includes a switchable negative ion generator — negative ions are well-known for their presumed psychological good effects. Its attractive case with woodgrain finish measures 11″ x 7″ x 5″; uses only 45 watts. The Bionaire 500 **$139.95** #2710. A larger model Bionaire 1000, has three speeds and cleans 118 CFM of air. Measures 14″ x 8″ x 8½″ with a lifetime steel housing. Bionaire 1000 **$199.95** #2720. Both are UL listed. Manufacturer's 1-yr. limited warranty. Set of two replacement filters for the 500 **$19.95** #2715; for the 1000 **$24.95** #2725.

Bionaire 500

THE LIFESTYLE RESOURCE™

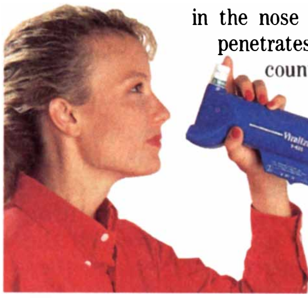
A WATCH FOR ALL REASONS

This handsome digital-analog chronometer will swim with you (to a depth of 150 feet), fly with you, calculate your speed, wake you in the morning, and tell you the day and date. A push of a button will convert the digital display to 24-hour timekeeping. All functions—digital and analog timekeeping, calendar, alarm, chronometer and tachymeter—are powered by a highly accurate electronic quartz movement. The analog hands and hour markers are luminous. The stopwatch times to 1/100th of a second, and the rotating bezel can be used for such practical matters as telling you when the parking meter will need another quarter. Adjustable fit. **$39.95.** Teflon-coated durable black-matte anodized stainless steel #1051; Stainless steel and gold-plated #1061.

TAKE YOUR CONTACTS FOR A SPIN

Contact lenses — so convenient to wear...so *inconvenient* to clean! The FDA found the risk of eye damage to be significantly reduced by proper lens care. The Clensatron™, developed specifically to *minimize* problems of lens care, is clinically proven to offer contact lens wearers an alternative to the traditional, "finger-rubbing" method which can scratch or tear lenses. Revolving at a rate of 150 cycles per minute, it thoroughly scrubs your lenses in *two minutes*, removing protein deposits and other contaminants that may damage your eyes. Works with hard, soft and gas-permeable lenses. Compact for travel, uses two AA batteries. AC adaptor included; UL listed. Manufacturer's 1-yr. warranty. **$59.95** #2680.

ALLERGY AND COLD BUSTER



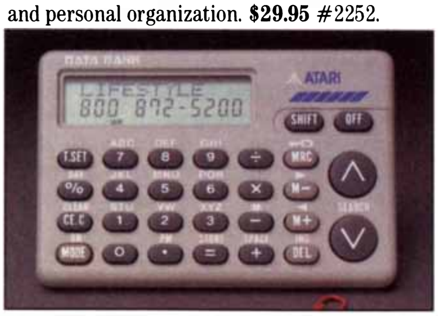
A major scientific breakthrough — The Viralizer System — gives you relief from cold, sinus and allergy symptoms. It's the newest development of a concept pioneered at the Pasteur Institute in Paris. The cause of the common cold is the heat sensitive Rhinovirus family which lives and multiplies in the nose and throat. Viralizer delivers a gentle, controlled heat which penetrates the nose and throat and dispenses either of two mild, over-the-counter, medicated sprays. Vira-Spray I is an analgesic, anti-bacterial spray. Vira-Spray II is a decongestant that relieves nasal congestion due to colds, hay fever, sinusitis or allergies. The Viralizer can produce effective relief by using it for only a few minutes, several times a day. Clinical tests prove the Viralizer System 90% effective in treating symptoms of upper respiratory infection. Doctor-recommended, the Viralizer works in 24 hours or less without pills. Safe for children and adults. Comes with a 3-pack refill of Vira-Sprays. **$39.95** #1691.

GOODBYE BACK PAIN

More than 6.5 *million* Americans suffer from back pain daily...and more than 8 out of 10 will eventually suffer from it. Research has shown that 83% of back problems are caused by muscles that have become weak due to stress and lack of exercise. But that's actually good news, because it means that most back pain is preventable! The "Say Goodbye to Back Pain" video is derived from the YMCA's "Y's Way to a Healthy Back" program which is based on research by noted back care expert, Dr. Hans Kraus. This doctor-recommended program is used at YMCA's nationwide, and has helped hundreds of thousands of people where other methods have failed. *Scientifically proven* to reduce or *eliminate* back pain in more than 80% of cases. Consult your doctor before beginning this six-week program of easy, relaxing and amazingly effective exercises which will help restore the strength and flexibility of your back...no matter how long you've neglected it. 96 minutes. **$39.95** #2370.

POCKET-SIZED MEMORY

The Atari Data Bank is the most innovative, easy-to-use, multi-function Data Bank in its class. No more fumbling through notes or phone books. Touch a key and information is at your fingertips. Holds up to 50 names, numbers and memos in its 2K memory. Raised rubber keys for easy data entry, built-in clock with date display, 8-function calculator and a secret function to safeguard confidential data, all on a 2-line 24-character display! Offers a scroll feature for access to information, all in alphabetical order, or direct search for instant retrieval. Turns itself off if you forget. Atari's multi-function Data Bank is the perfect solution for a complete business and personal organization. **$29.95** #2252.



TO YOUR HEALTH

Ancient Mandarins dating back 800 years believed that these Chinese Exercise Balls induced well-being of the body and serenity of spirit. These treasured gifts were given to President Reagan and his wife while visiting the People's Republic of China. The Chinese say that rotating the balls in the palm of each hand stimulates the fingers and acupuncture points, and improves circulation of vital energy throughout the body. Sports enthusiasts, musicians, computer users and health-conscious people everywhere consider them great muscle conditioners. Arthritis sufferers feel a decided benefit from this gentle but challenging exercise. Very effective for relaxation and meditation, Chinese Exercise Balls emit a distantly mysterious chime as you turn them. Beautifully handcrafted, 45mm. hollow polished chrome balls are perfectly weighted and fit comfortably into the average man's or woman's hand. In exquisite silk brocade box. **$29.95** #1702.

THE LIFESTYLE RESOURCE™

FOR FASTEST SERVICE — CREDIT CARD ORDERS CALL TOLL-FREE 24 HRS A DAY **800-872-5200**

DEPT. SFAK09; 921 EASTWIND DR. SUITE 114: WESTERVILLE, OH 43081

SEND TO (PLEASE PRINT)

ADDRESS

CITY STATE ZIP

☐ CHECK OR MONEY ORDER ☐ MASTERCARD ☐ VISA ☐ AMEX SIGNATURE

ACCT. # EXP. DATE

ITEM NO	QTY	DESCRIPTION	ITEM PRICE	TOTAL PRICE
			SUB TOTAL	
			SHIPPING (See table at left)	
			TOTAL	

Shipping Charge covers UPS, handling and insurance for guaranteed delivery. UPS Second Day available for an additional $7.50 per order.

UPS 2ND DAY AIR

Up to $20 .... $ 3.95	$50.01 to $60 .... $ 7.95
$20.01 to $30 .... $ 4.95	$60.01 to $70 .... $ 8.95
$30.01 to $40 .... $ 5.95	$70.01 to $100 .... $10.95
$40.01 to $50 .... $ 6.95	Over $100 .... $12.95

Canadian residents we cannot accept mail orders, please call (614) 794-2662.

ORDER WITH CONFIDENCE

- *Most orders ship within 48 hours of receipt.*
- *Credit Card orders billed only upon shipment.*
- *No risk 30 day return privilege.*

THE MEDIA DEVELOPMENT GROUP

©1989 C92 ADVERTISEMENT

come on a truly profound insight. "If the time asymmetry is borne out by more complicated models, this would seem to be an amazing fact about the universe that this model might explain.... I know of no previous explanation for the arrow of time." —*T.M.B.*

Out of Its Field

How the neutron responds to a field that exerts no force

Neutrons, the nuclear particles that have no charge, and electrons, the primary carriers of negative charge, interact in a way that would have boggled even the mind of Michael Faraday, the father of electrodynamics. To all appearances, a recent experiment by workers at the University of Missouri at Columbia and the University of Melbourne in Australia suggests that a neutron can sense an electron from a distance without experiencing a force generated by the particles' electric or magnetic field.

Rather than a quantum leap of faith, belief in this subtle effect requires an appreciation of the elegant theoretical work of Yakir Aharonov and A. Casher of Tel Aviv University. The theorists hold that such an interaction is possible because the electron creates an entity called a potential field, from which electric and magnetic fields can arise [see "Quantum Interference and the Aharonov-Bohm Effect," by Yoseph Imry and Richard A. Webb; SCIENTIFIC AMERICAN, April]. The potential field originally appeared in the theories of quantum mechanics and relativity as a matter of mathematical convenience, but the field had no known physical significance. Physicists no longer invoke the potential just for the purposes of theoretical bookkeeping, however. A pure potential field—one that is not manifested as a magnetic or electric field—can influence the behavior of a neutron.

How does the neutron sense the potential field? The neutron possesses a magnetic moment, that is, it has a certain magnetic strength that acts in a particular direction. The neutron's magnetic moment is readily revealed when the neutron crosses a magnetic field. For instance, in the earth's magnetic field, the neutron's magnetic pole precesses about 1,000 times per second. Aharonov and Casher posit a far more elusive effect: the interaction between a neutron and a pure potential field.

The task of detecting the Aharonov-Casher effect challenged the experimental team of Alberto Cimmino, Geoffrey I. Opat and Anthony G. Klein from Melbourne and Helmut Kaiser and Samuel A. Werner from Missouri. Writing in *Physical Review Letters,* the workers report that they directed a beam of neutrons toward an electrode so that an individual neutron could pass either above or below it. In the realm of particle physics, one cannot know whether the neutron will take the top or bottom route without disturbing the experiment. Instead the neutron is described as a wave that encodes information about the probability of the neutron being on either side of the electrode at a particular time. The neutron wave can be regarded as having been split so that it could travel above and below the electrode simultaneously.

Both the upper and lower neutron waves are then directed toward a silicon crystal, where they recombine and interfere. If the waves reinforced one another, the chance that the neutron would be observed was great. If the waves canceled each other out, then the chance that the neutron would be detected was small.

By means of two detectors, the physicists counted the number of neutrons observed while the charge on the electrode was negative and then positive. Because in each state the electrode generated a different potential field, the upper neutron wave shifted with respect to the lower neutron wave. These shifts caused small changes in the way the neutron waves interfered and hence affected the number of neutrons detected.

During several months the workers observed about 50 million neutrons; they found a change of about one count per 1,000 when the polarity of the electrode was switched. The investigators say that although the result could not be explained in terms of ordinary electric and magnetic fields, it agreed with Aharonov and Casher's predictions. —*Russell Ruthen*

Low-Zone

The infamous hole has influence beyond Antarctica

The continent-size hole in the stratospheric ozone layer that forms every October over Antarctica has the saving grace that virtually no one lives under it. But a recent study, reported in *Nature,* confirms what some researchers have suspected: parcels of ozone-depleted air drift away from the hole when it breaks up in December and dilute the ozone over distant, inhabited regions. The researchers worry that the dilution may cause increases in ultraviolet radiation levels just when people are out sunbathing, during the Southern Hemisphere's high summer.

Roger J. Atkinson of the Australian Bureau of Meteorology, R. Alan Plumb of the Massachusetts Institute of Technology and their colleagues studied air movements in the Southern Hemisphere during December, 1987, a few months after the appearance of the biggest ozone hole ever detected. By combining satellite and ground-level data with analyses of wind patterns, the workers could track large bands of ozone-depleted air that pulled away from the hole. One band extended across southern New Zealand and southern Australia; another skirted South Africa and swept over South America. The authors write that their tracking exercise produced "a strong prima facie case" that ozone-poor air from the hole contributed to record-low ozone levels of about 7 percent below normal hovering over five major cities in Australia, New Zealand and Tasmania.

A 7 percent decrease in total ozone should raise ultraviolet levels at the ground by some 14 percent, according to Paul A. Newman, a member of the study team from NASA's Goddard Space Flight Center. Even for one month, that is "highly significant," observes Janice Longstreth of Clement Associates, an environmental consulting firm in Fairfax, Va. Longstreth suggests that such a decline would increase the number of cases of sunburn and therefore enhance the risk of melanoma, an aggressive skin cancer. Sunsor, Inc., in Pittsburgh is hoping to sell hand-held ultraviolet monitors to school authorities in Australia so that teachers can caution their charges to stay indoors on high-ultraviolet-level days.

In another study published in *Nature,* David J. Hoffman of the University of Wyoming and his colleagues report ozone depletion in association with polar stratospheric clouds (the same pattern seen in Antarctica) near the Arctic Circle. Such clouds are essential for rapid ozone depletion. Normally the Arctic is not cold enough for them to form in significant amounts, but this year the Northern Hemisphere stratosphere was the coldest it has been for at least 25 years. The ozone depletion that Hoffman and his colleagues detected in January (with instruments on balloons launched from Kiruna in northern Sweden) was slight:


Lufthansa

"Find us computers that will let us do many jobs at once, Sinkfield, and it'll be another feather in your cap."

How're you going to do it?

No one ever made their mark in business by doing just one thing at a time. So to stay competitive, you need your computers to do a lot of business tasks at once.

The Genius Of Micro Channel. That's why IBM developed the Personal System/2® computers with Micro Channel.™ Micro Channel is like a multilane highway, so information can travel in lots of directions at once. With optional bus master cards, it's like having many computers in one, some managing printers and fax machines, another busy calculating a spreadsheet and others working on graphics or transferring data. And Micro Channel

PS/2 it!

does all of this with remarkable speed and reliability. Naturally, a PS/2® with Micro Channel runs DOS applications, and add OS/2 to this equation, and the benefits of multitasking can make you and your people even more productive.

The Solution Is IBM. The PS/2 with Micro Channel will provide you with solutions not only today but far into the future. See your IBM Authorized Dealer or IBM marketing representative. For a dealer near you, call 1 800 IBM-2468, ext. 139. And see how much more you can do when you PS/2 it!

IBM®

3 percent, compared with Antarctic depletions of 50 percent. —*T.M.B.*

TECHNOLOGY

Heavy-Ion Fusion

A dark horse emerges in the race toward fusion energy

As cold fusion pulls up lame, a more serious contender seems poised to break out of the pack in the fusion-energy competition. Called heavy-ion fusion, the technology would generate miniature thermonuclear explosions by accelerating charged particles of lead or other massive elements into capsules of hydrogen isotopes. Many experts maintain that this technique represents the only plausible means of harnessing fusion; it elicits praise even from those pursuing rival techniques, a rarity in this hypercompetitive field. Stephen E. Bodner, who directs laser-fusion research at the Naval Research Laboratory, calls heavy-ion accelerators "the only possible approach"—in addition to his own, of course—to a commercial fusion generator.

Heavy-ion fusion is a latecomer to the field of inertial-confinement fusion, which involves making hydrogen targets implode by blasting them with various types of radiation. Since the late 1960's the U.S. Department of Energy has invested about $2 billion in the field; its annual budget is now about $155 million. Most of the money has been spent building huge drivers, or radiation-generating machines, at the national nuclear-weapons laboratories. Lawrence Livermore and Los Alamos laboratories are focusing on lasers and Sandia Laboratories on machines that accelerate light ions such as lithium.

Heavy-ion fusion, by contrast, has received only about $4 million a year since the Energy Department first began funding it in the late 1970's. The Lawrence Berkeley Laboratory in Calif., the prime U.S. facility for heavy-ion research, has built a small, experimental heavy-ion accelerator. It generates beams whose energy totals only about a joule; the machines at Livermore and Sandia produce beams many thousands of times more powerful.

Nevertheless, in the past few years the reputation of heavy ions has been quietly growing. In 1986 a review of the entire inertial-confinement-fusion field by the National Academy of Sciences concluded that "heavy-ion beams may well be the best eventual driver for energy applications." A study by several Energy Department laboratories came to a similarly optimistic conclusion a year later. "It could be a case of the less we know about it the better it looks," says Donald J. Dudziak of Los Alamos, who headed the study. "But no one has come up with a major problem yet."

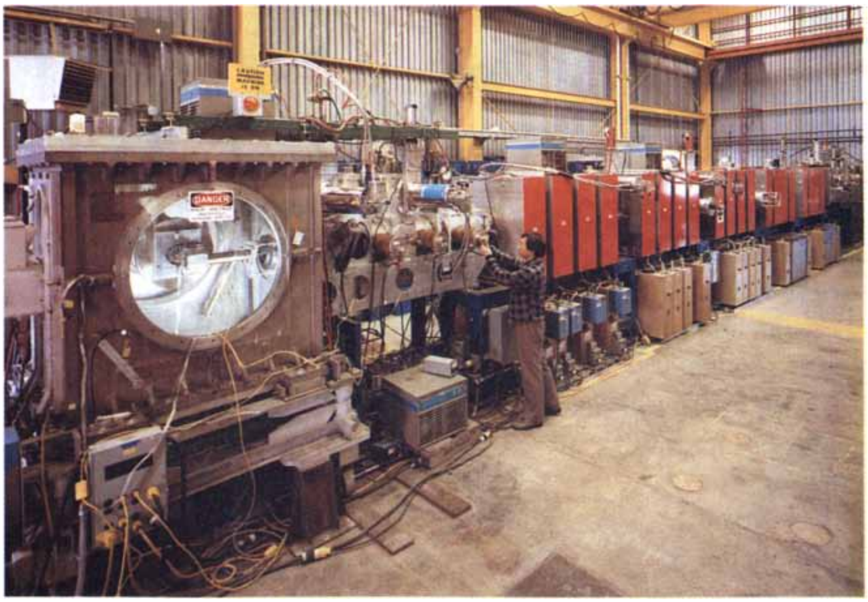
Proponents of heavy ions say this approach offers several advantages. The accelerators being considered for heavy-ion fusion are variants of those used for high-energy physics research; they can convert electricity into beams of energetic particles with great efficiency. The accelerators can also fire rapidly—hundreds of times per second in many cases—without breaking down. Ten shots a second would be more than enough for a commercial power generator.

High-energy lasers, on the other hand, are notoriously inefficient and prone to breakdowns; the Nova laser at Livermore, for example, typically fires only a few times a day. The chief advantage of heavy ions over the light-ion approach taken by Sandia is that fewer heavy ions are needed to produce the same impact; the heavy-ion beam is therefore much easier to focus. Perhaps the biggest drawback of a heavy-ion generator would be its size. Denis Keefe, who heads the Lawrence Berkeley program, estimates that the accelerator for a commercial generator would be some five kilometers long. "That scares some people," he acknowledges. "But then any fusion driver would be a big machine."

The technology is being pursued in Japan, the Soviet Union and—most aggressively—Europe. West Germany's $400-million Institute for Heavy-Ion Research in Darmstadt represents the most advanced heavy-ion facility in the world. With the encouragement of Nobel laureate Carlo Rubbia, the director of the European laboratory for particle physics (CERN), Italy recently initiated an ambitious heavy-ion program at the School of Plasma Physics in Varenna; plans call for an eventual annual budget of $100 million.

Several recent events could presage greater support for heavy ions in the U.S. Recently Robert O. Hunter, Jr., the Energy Department's director of energy research, proposed taking $50 million from the program for magnetic-confinement fusion (in which a hydrogen plasma is heated while being held in a powerful magnetic field) and giving it to the inertial-confinement-fusion program. Hunter also called for placing greater emphasis on the energy applications of inertial fusion; the official goal of the program is now nu-

Heavy-ion accelerators can fire rapidly, reliably and efficiently—valuable traits for a generator



MULTIPLE-BEAM EXPERIMENT (MBE-4) is a heavy-ion accelerator at the Lawrence Berkeley Laboratory. Whereas most accelerators boost particles by means of powerful radio waves, the MBE-4 employs induction. Ions are generated in the box with the window at the left and accelerated in the red boxes at the right.

clear-weapons research. Finally, Congress has asked the National Academy of Sciences to review the potential of all inertial-fusion approaches, including heavy ions, for producing energy; a preliminary report is due in January. Keefe has his hopes up. "Heavy-ion fusion is too good an idea not to pursue vigorously," he says. *—J.H.*

Spacecraft on a String

Satellite "tethers" could provide power and thrust

In May, 1991, the future of spaceflight could literally be hanging by a thread: a spaghetti-thin 20-kilometer cable stretching between the space shuttle *Atlantis* and a satellite in orbit. The Tethered Satellite System (TSS), a project sponsored jointly by the National Aeronautics and Space Administration and the Italian Space Agency (ASI), will draw electrical energy to the shuttle from the earth's magnetic field. If the *Atlantis* mission is successful, tethers may find applications in adjusting the orbits of spacecraft, recharging the batteries of failing satellites and transporting wastes away from space stations.

Although it sounds exotic, TSS will work like any ordinary electric generator in which coiled wires rotate around a magnet. The motion of the conducting tether through the planet's magnetosphere will induce a flow of current as electrons collected by the satellite from surrounding ionized gases move through the tether to the shuttle. The strength of the current is proportional to tether length; the 20-kilometer tether on the first TSS mission should produce from 4,000 to 5,000 electron volts of energy or about four kilowatts of power.

"Of course, you don't get something for nothing," points out Thomas D. Stuart, the NASA program manager. "The energy has to come from somewhere." Consequently, when current flows toward the shuttle, electromagnetic drag between the tether and the earth causes the shuttle to lose speed and altitude. Yet this loss points the way to a different application: if the direction of current flow is reversed, with the shuttle pumping electrons through the tether toward the satellite, electrostatic repulsion between the charged satellite and the ionized gases around it will cause the shuttle to gain altitude. Thus, TSS could also serve as a source of thrust.

The tether—a multilayered cable—that NASA and its prime contractor, the Martin Marietta Corporation, have created for the mission is only one tenth of an inch thick. At its core is a copper-wound, plastic filament that is surrounded for strength with braided Kevlar, the lightweight plastic used in bulletproof vests. The outermost layer is made of Nomex, a synthetic fiber that will protect against oxidation by the ionized gases around the orbiting spacecraft. The tether can hold up to 420 pounds without breaking, but it should experience only about 20 pounds of tension. Ordinary fishing line can stand that much pull.

NASA has completed work on the tether and on its deployment mechanism; tests of the deployer system will begin this November. ASI will start environmental tests on the satellite in January, 1990. The first TSS mission is scheduled for the May, 1991, shuttle launch. NASA and ASI are planning a second mission that could take place in late 1994. On that flight, a tether 100 kilometers long would be deployed to demonstrate the feasibility of tethered systems for upper atmospheric research.

The TSS missions have an interesting provenance. In 1974 Mario D. Grossi and Giuseppe Columbo of the Harvard-Smithsonian Center for Astrophysics proposed in their "Skyhook" report that tethered satellites would have useful electrodynamic properties. Credit also belongs to earlier visionaries, including Yuri N. Artsutanov, a Soviet engineer. In the 1950's he proposed an elevator, or "funicular," traveling into space on cables 38,000 kilometers in length that would extend from the ground to a satellite in geostationary orbit.

Modern materials are still too weak by an order of magnitude to support such a structure. Yet if the *Atlantis* mission and its possible follow-ups are successful, Artsutanov's dream may assume a slightly stronger tinge of reality. *—John Rennie*

BIOLOGICAL SCIENCES

Diluvian Tremens

Policy initiatives could turn the tide on wetland loss

"From a geologist's point of view, we're sitting on a gold mine," says Shea Penland. To the layman it looks like a bayou: clapboard shanties, lush cow pastures, long fields stippled with young sugar cane and the muddy Mississippi, trapped in a levee 12 feet high. As we drive upriver, Penland, a coastal geologist at the Louisiana Geological Survey, explains that by containing sediment as well as water in the Mississippi, the levees are starving the southern Louisiana wetlands of the silt that would normally rebuild them—one of the many reasons the five-million-acre Mississippi delta is fast becoming part of the Gulf of Mexico.

And for geologists such as Penland who have made coastal erosion and wetland loss their specialty, times are suddenly flush. This past summer the state legislature created a mechanism for funding wetland protection and restoration that should provide $25 million in the next year; it also created a cabinet-level position to administer the money and coordinate state and federal efforts. The move was in part an answer to growing constituent concern and in part an effort to convince the federal government of Louisiana's commitment to its own turf: Congress is currently considering bills that would send roughly $150 million a year to the state for wetland protection and restoration.

In Louisiana, home to 40 percent of the country's coastal wetlands and 80 percent of its coastal-wetland loss, a battle is beginning that could have repercussions far from Cajun country. Projections of global sea-level rises resulting from greenhouse warming as well as from greater awareness of the value of wetlands—as wildlife habitats, sources of groundwater and sinks for floodwater—have put the wetlands' plight on the national agenda. In January the Environmental Protection Agency formally adopted the goal of preventing the net loss of wetlands in the continental U.S.; the current EPA administrator, William K. Reilly, was president of the Washington conservation group that recommended the objective. President Bush referred to the "no net loss" pledge in his budget address in February and named a task force under the Domestic Policy Council to determine how to implement the pledge.

The goal can be achieved by a number of means. One unlikely extreme would be a complete halt to development on or near wetlands; the other, equally unlikely, would involve the complete destruction of natural wetland environments balanced by the construction of new wetlands. The responsibility for deciding how much of each happens lies largely with the U.S. Army Corps of Engineers, the agency that issues permits for the development of coastal and riverine areas. The

Corps receives 15,000 individual permit requests a year; the EPA provides the environmental criteria by which the Corps judges those requests and has the power to veto permits the Corps grants. The EPA says it puts more emphasis on avoiding the destruction of existing wetlands than the Corps does, a point of contention between the agencies.

"When you destroy an existing wetland, you destroy the best guide you have to creating a new wetland," says Jon A. Kusler, executive director of the Association of State Wetland Managers and co-editor of a recent EPA report on the science of wetland creation and restoration. Such guides are precious: restoration projects are largely empirical, and Kusler says roughly half of them have failed so far in one respect or another. "The problem is that there are many types of wetlands, and they can be managed for many different functions," he says. Even though some of his colleagues have supervised more than 200 restoration projects, he says "there are very few experts in wetland restoration."

Because each wetland site has a distinct personality, a manager's experience with one site does not necessarily contribute to his or her understanding of another. Forested wetlands are harder to manage than marshes, and inland areas tend to be more complex than coastal ones. Different strategies promote the proliferation of certain kinds of vegetation or species of fish or songbirds. The flora and fauna of a wetland can be extremely and unpredictably sensitive to oxygen exchange, nutrient cycling, soil gradients and wave energy; Kusler says there may be other critical parameters that have not been recognized yet.

Those parameters that have been recognized are conspiring against the wetlands in southern Louisiana. Until it was leveed some 60 years ago, the Mississippi River routinely replenished the delta wetlands with almost 200 million tons of sediment a year, compensating for the compaction of the loose delta soil. Now the river dumps its sediment beyond the continental shelf, and in some places the land subsides at a rate of more than two centimeters a year. Offshore oil and gas extraction may also contribute to the subsidence of the delta by undermining its substrate.

Furthermore, the barrier islands a few miles off the coast are migrating west, leaving the delta exposed to waves straight from the gulf. Navigation channels cut into the delta by oil and gas companies carry saltwater inland, killing trees and grass whose roots hold the bayou together. Some coastal experts even blame nutrias, ratlike rodents the size of terriers that were imported from South America to control the spread of water hyacinth [see "Waterweed Invasions," by Spencer C. H. Barrett, page 90] and ended up eating everything else, too.

"Nowhere in the U.S. is the subsidence and erosion rate anywhere near the one in coastal Louisiana," says John W. Meagher of the EPA's Office of Wetland Protection. The natives have a number of ways of quantifying it: a Rhode Island-size chunk gone in 20 years, 50 square miles awash per year, a football field sunk every 15 minutes. In the wake of the legislature's recent move, however, the apocalyptic mood has become more upbeat. "We'll be building a coastal defense and restoration industry that will be needed throughout the world as sea levels rise," says G. Paul Kemp, executive director of the Coalition to Restore Coastal Louisiana, whose 80 members include churches, chambers of commerce and Indian tribes. "Louisiana will be a mecca for coastal engineers. They'll be seeing in 50 years what we're seeing today." *—Karen Wright*

First Impressions

Genes from mother don't equal genes from father

Experiments and clinical observations are leading to a revision of one of the basic tenets of Mendelian genetics: the principle that it makes no difference whether a gene was inherited from an animal's mother or its father.

As long ago as 1959 Janice B. Spofford of the University of Chicago found a gene in *Drosophila* whose effect differed depending on whether it came from the male or female parent. Like Mendel's own discoveries, the observation was largely ignored. Now studies are picking up the theme, demonstrating that in mice, human beings and other species some genes are inactivated, in a process known as imprinting, when they are inherited from the parent of one sex but not when inherited from the other parent. The inactivation may last throughout an individual's lifetime, but genes are "wiped clean" and reimprinted when they are passed on to subsequent generations. Indeed, puzzling aspects of some human genetic diseases and cancers might find an explanation in imprinting.

One of the key observations concerned mouse embryos that contained either two sets of maternal chromosomes or two sets of paternal chromosomes. Such embryos, which can be made by transplanting nuclei of fertilized eggs, always fail to develop. The failures of development are, however, very different in the two cases, indicating different roles for maternal and paternal DNA.

Then two years ago Carmen Sapienza of the Ludwig Institute for Cancer Research in Montreal and M. Azim Surani of the Institute of Animal Physiology and Genetics Research in Cambridge, England, showed independently that in mice the sex of the parent of origin determines to what degree the DNA of an inserted foreign gene is chemically modified by methylation—the addition of a methyl (CH_3) group to certain components. Genes from one parent, in other words, are more readily methylated than are genes from the other. Methylation is a major mechanism of gene regulation, and so such selective chemical modification may possibly explain how imprinting is achieved, although little is known about the process.

Newly available techniques seem to be making clinical workers as well as mouse geneticists aware of the different roles of maternal and paternal genes. Sapienza and others have established that in Wilms' tumor, a cancer of the kidney in which malignant cells characteristically lose all or part of chromosome 11, the loss usually turns out to involve the maternally inherited chromosome, not the one inherited from the father.

Sapienza and others have evidence that imprinting is involved in other tumors, including rhabdomyosarcoma, a muscle tumor, and some cases of retinoblastoma, a heritable eye tumor. Sapienza says he has indications that the primary defect in some cases of retinoblastoma may lie not in the known retinoblastoma gene site but in a different regulatory gene that is responsible for imprinting other genes. "What you're inheriting is an aberrant imprinting mechanism," he suggests.

In a literature search for parental effects, Judith G. Hall of the University of British Columbia found numerous examples among genetic diseases, both in humans and in mice. She believes many of them could be explained by imprinting; for example, the mechanism might explain why patients who inherit the gene for Huntington's disease from their father sometimes acquire a more severe form of the disease that begins earlier in life than do those who inherit it from their mother. The reverse is true of myotonic dystrophy. Nancy S. Wexler of Columbia University, president of the Hereditary Disease Foundation, is now examining Huntington's pedigrees for evidence of imprinting.

"Imprinting is a genuine phenomenon, and it has not been well delineated hitherto," comments Victor A. McKusick of Johns Hopkins University, perhaps the preeminent authority on human genetic diseases. "However, I would not want to give the impression that it turns Mendel upside down or that it's anything revolutionary; it's a phenomenon that explains some of the irregularities of transmission of genes in families." —*T.M.B.*

Trans-Kingdom Sex

E. coli *mate with yeast!*
(Will Jesse Helms cut off funds?)

Bacteria are a promiscuous lot. These primitive prokaryotes lack not only nuclei but also the inhibitions that bar eukaryotic species—from yeast to humans—from breeding with one another. Indeed, different species of bacteria commonly get into each other's genes through a process called conjugation. It begins when one bacterium, the donor, presses against another, the recipient. A pore opens between the cell walls, and the donor slips a ring of DNA known as a plasmid through the pore and into the recipient. The recipient takes up the new DNA and passes it on to all its descendants, which are generated through the more traditional form of reproduction—cellular division.

Might the sexual propensities of bacteria extend even to eukaryotes? Biologists have long known that at least one species of bacterium is able to manipulate the genes of eukaryotic cells. Called *Agrobacterium tumefaciens*, this much studied pathogen causes plant cells to form tumors, which in turn produce nutrients that nourish the bacterium. Until recently many investigators thought this case of natural genetic engineering involved a unique, highly specialized mechanism. Then three years ago Patricia C. Zambryski of the University of California at Berkeley and other workers discovered that *A. tumefaciens* alters the genes of plant cells in essentially the same way that bacteria alter each others' genes—through conjugation. The finding suggested that bacteria might be even more promiscuous than anyone had supposed.

Inspired by the work of Zambryski and her colleagues, Jack A. Heinemann and George F. Sprague, Jr., of the University of Oregon began investigating whether *Escherichia coli*, a bacterium common to human digestive tracts and a favorite subject for biologists, could conjugate with yeast, a eukaryotic denizen of the plant kingdom. The workers first altered the genome of the yeast cells so that they could no longer synthesize leucine, an amino acid needed for growth and reproduction. They then introduced a leucine-producing gene into the *E. coli*, along with other DNA sequences known to promote conjugation among bacteria. When placed in a petri dish with the *E. coli*, the yeast cells quickly regained their ability to grow and reproduce. Evidently, Heinemann and Sprague report in *Nature*, conjugation had taken place.

This finding of yet another case of trans-kingdom sex has important implications for biotechnology. It suggests that, perhaps with some prodding, many kinds of bacteria might serve to insert genes into many kinds of eukaryotic organisms. *A. tumefaciens* has already become such a tool—in the hands of researchers trying to make crops more resistant to disease or drought, for example—but it has certain limitations; most notably, it cannot mate with so-called monocots, which include wheat, corn and other important cereals. "Conjugation involving other bacteria might provide a way to introduce genes into monocots," Sprague says, "and possibly into animal cells."

To change an animal's genome, Sprague notes, a bacterium would have to conjugate with one of the animal's germ cells—an egg or sperm. How likely is it that such a union would happen in nature, and what would be its outcome? The answers to these questions could have a profound bearing on evolutionary theory, which now holds that eukaryotic species evolve as a result of random genetic mutations. Zambryski and Scott E. Stachel of Genentech, Inc., remark in a comment accompanying Heinemann and Sprague's report in *Nature* that "trans-kingdom conjugation, unlike interspecies mythological matings, will give rise to neither minotaurs nor satyrs. Nevertheless, it might mediate a low level of horizontal gene transfer in eukaryotes. It remains for future experiments to . . . deduce what roles such disparate sex might play in nature." It's enough to make a eukaryote blush. —*J.H.*

MEDICINE

Sudden Impact

Why a little heart disease can be worse than a lot

Here is something else to worry about. People in whom heart disease has just begun to develop and so has not produced symptoms run a higher risk of dying if they suffer a heart attack than do people in whom the disease is more advanced.

A report in the *New England Journal of Medicine* links this "fascinating but disturbing fact" to the body's complex response to heart disease. Heart disease commonly begins when plaque accumulates on the inner walls of the coronary arteries, which causes them to narrow. Eventually, this narrowing results in the reduction of blood flow through the arteries, a condition known as ischemia. In many people, however, as ischemia becomes more severe, the collateral arteries feeding the heart begin to enlarge, compensating to an extent for the diminished capacity of the primary arteries.

People whose disease has progressed to this point have two advantages. First, their disease can be readily detected—by a cardiogram, for example, or by their own sensation of chest pains, known as angina—and treated. Moreover, if a main coronary artery is suddenly blocked, the enlargement of the collateral arteries softens the blow of the resulting attack, or myocardial infarction.

Conversely, those in whom ischemia and the resulting enlargement of the collateral arteries have not yet occurred have a double disadvantage. They show no signs of trouble and so have no reason to take remedial action. If blockage of a main coronary artery does occur, its effects are likely to be devastating, because the collateral arteries provide no backup.

That is why the "total number of deaths occurring over time in an asymptomatic population screened by stress testing is far greater among subjects with negative tests than among those with positive tests," write Stephen E. Epstein, Arshed A. Quyyumi and Robert O. Bonow of the National Heart, Lung and Blood Institute in Bethesda, Md. There is no practical way to identify those at risk of out-of-the-blue attacks so that they can be treated, according to the authors. Such fatalities can only be reduced, they conclude, if more people

SCIENTIFIC AMERICAN
In Other Languages

LE SCIENZE
L. 3,500/copy L. 35,000/year L. 45,000/(abroad)
Editorial, subscription correspondence:
Le Scienze S.p.A., Via G. De Alessandri, 11
20144 Milano, Italy
Advertising correspondence:
Publietas, S.p.A., Via Cino de Duca, 5,
20122 Milano, Italy

サイエンス
Y950/copy Y10,440/year Y14,000/(abroad)
Editorial, subscription, advertising correspondence:
Nikkei Science, Inc.
No. 9-5, 1-Chome, Otemachi
Chiyoda-ku, Tokyo, Japan

INVESTIGACION Y CIENCIA
500 Ptas/copy 5500 Ptas/year 6200 Ptas (abroad)
Editorial, subscription, advertising correspondence:
Prensa Cientifica S.A.,
Calabria, 235-239
08029 Barcelona, Spain

POUR LA SCIENCE
27FF/copy 265FF/year 315FF/year (abroad)
Editorial, subscription, advertising correspondence:
Pour la Science S.A.R.L.,
8, rue Férou,
75006 Paris, France

Spektrum DER WISSENSCHAFT
9.80 DM/copy 99 DM/year 112.20 DM/(abroad)
Editorial, subscription correspondence:
Spektrum der Wissenschaft GmbH & Co.
Moenchhofstrasse, 15
D-6900 Heidelberg,
Federal Republic of Germany
Advertising correspondence:
Gesellschaft Für Wirtschaftspublizistik
Kasernenstrasse 67
D-4000 Duesseldorf,
Federal Republic of Germany

科学
1.40RMB/copy 16RMB/year $24/(abroad)
Editorial, subscription correspondence:
ISTIC-Chongqing Branch, P.O. Box 2104,
Chongqing, People's Republic of China

В МИРЕ НАУКИ
2R/copy 24R/year $70/(abroad)
Editorial correspondence:
MIR Publishers
2, Pervy Rizhsky Pereulok
129820 Moscow U.S.S.R.
Subscription correspondence:
Victor Kamkin, Inc.
12224 Parklawn Drive,
Rockville, MD 20852, USA

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

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

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

take steps, dietary or otherwise, to avoid heart disease entirely. —*J.H.*

Status Symbol
Affluent women have cesareans more often than poor women do

In a perfect world, everyone would receive the same medical care regardless of his or her economic status. A study described in the *New England Journal of Medicine* reveals that when it comes to childbirth, reality diverges widely from this ideal.

The study focused on 245,854 women who delivered a child in Los Angeles County during 1982 and 1983. Of this group, 17.8 percent underwent a so-called primary, or first-time, cesarean section. (The study excluded women who had previously undergone a cesarean, because abdominal delivery is usually—and sometimes needlessly, some obstetricians now say—prescribed for them.)

Using birth certificates, hospital records and U.S. census data, the investigators found that the rate at which cesareans were performed rose and fell with the patients' income. Only 13.2 percent of women with a median family income of less than $11,000 received cesareans; of those whose income exceeded $30,000, 22.9 percent underwent the procedure. The correlation held up independently of other potentially significant factors, such as maternal age, ethnic background, fetal weight and previous childbearing experience.

The investigators had expected low-income women to exhibit more of the complications—signs of fetal distress, for example, or prolonged labor—that necessitate a cesarean. After all, poor women are much less able to obtain proper prenatal care than are affluent women. But records indicated that poor women had the fewest reported complications and affluent women the most. This finding, according to the investigators—Jeffrey B. Gould, Becky Davey and Randall S. Stafford of the University of California at Berkeley—may result from "the overreporting of complications in the affluent or the underreporting of complications in the poor."

The implications of the study are all too clear, according to Sidney M. Wolfe of the Public Citizen Health Research Group, a consumer advocacy group. Physicians overdiagnose complications in affluent childbearing women and then overprescribe cesareans, Wolfe asserts, because the surgical procedure provides a bigger payoff than vaginal delivery does. Indeed, he charges that avarice on the part of hospitals and physicians is largely to blame for the fourfold increase in the rate of cesareans in the U.S. since 1970. Physicians have less incentive to perform cesareans on poor women, Wolfe adds, because the financial compensation is usually less.

Mortimer G. Rosen, who heads the department of obstetrics and gynecology at the Columbia-Presbyterian Medical Center in New York City, agrees with Wolfe that cesareans are overprescribed for affluent women but sees factors other than greed behind the phenomenon. He points out that whereas poor women are usually treated by whoever happens to be on duty when they arrive at the hospital, affluent women often retain a private obstetrician. These obstetricians have more emotional interaction with their patients, Rosen says, and they are much more likely to be sued for malpractice if something goes awry. (Obstetricians in general pay higher malpractice insurance rates than most other types of physicians.) As a result, Rosen says, private obstetricians are often too quick to see trouble during labor and delivery and to recommend the surgical remedy.

So affluent women receive too many cesareans. Do poor women receive too few? Rosen says that rarely are cesareans clearly called for (for example, when vaginal delivery might lead to brain damage in the infant) but not performed. Ironically, he observes, poor women may receive more judicious care—at least when it comes to method of delivery—than affluent women do. —*J.H.*

OVERVIEW

Glass Menageries
Can cryopreservation freeze the assets of endangered species?

Within the next 200 years an estimated 815 species of mammals will slip irretrievably into extinction, unless specimens can be successfully bred in captivity. The ultimate success of such efforts may depend on the use of reproductive biotechnology to breed animals in sufficient numbers so that they can recapture their niches when they are returned to the wild. Techniques such as in vitro fertilization, cryopreservation of sperm and eggs, and surrogate

TAKE ANY 3 FOR $1 EACH

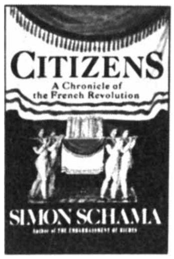
Plus a 4th at the low Member's price.

NO RISK, NO COMMITMENT.

1057. $19.95/**$17.50**

1644. $24.95/**$17.50**

5884. $27.50/**$18.50**



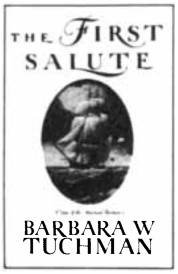
1396. $29.95/**$21.95**

1206. $24.95/**$19.50**

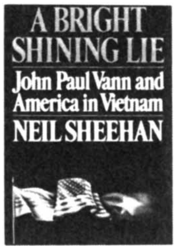
6767. $29.95/**$18.95**

3012. $14.95/**$10.95**

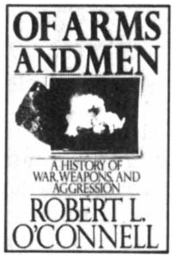
1859. $15/**$10.50**



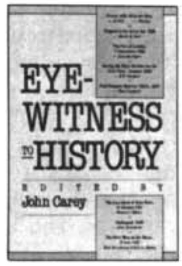
6080. $22.95/**$17.50**



1495. $24.95/**$18.50**

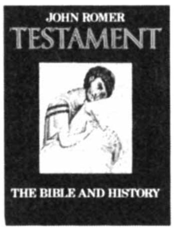


1578. $24.95/**$18.95**

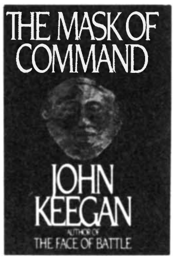


6007. $24.95/**$17.95**

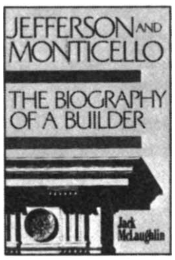
7526. $17.95/**$15.50**



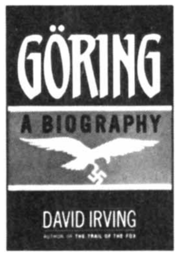
1560. $29.95/**$22.95**



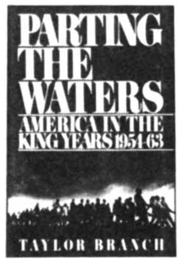
4473. $18.95/**$15.50**



1503. $29.95/**$19.95**



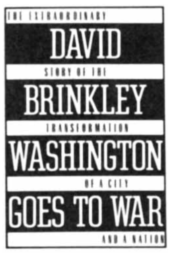
8615. $22.95/**$19.50**



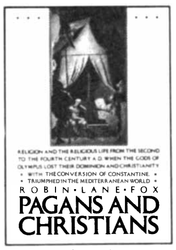
1081. $24.95/**$18.95**

1024. $22.95/**$17.95**

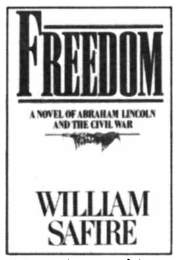
3988. $35/**$22.50**



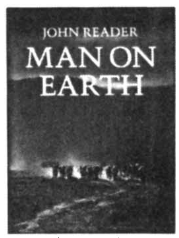
6130. $18.95/**$15.95**



3806. $40/**$23.50**



3327. $24.95/**$17.95**



1677. $29.95/**$22.95**

(First price is Publisher's List. **Boldface** is Member's Price.)

Save on the best recent history titles.

No matter what area of history you enjoy reading about most, you'll find that History Book Club offers some of the finest selections being published today. And no book club we know of offers greater savings —as much as 30% off publishers' list prices.

You can save even more by taking advantage of our Introductory Offer. Select any three books on this page for $1 each when you take a fourth book at the low Member's price—plus shipping and handling. Thereafter, you're not obligated to order any more books. You may cancel membership at any time by notifying History Book Club. We may cancel your membership if you elect not to buy at least one book in any six-month period.

How the Club works. You'll be able to choose from 150 to 200 books featured each month. History Book Club always offers its members well-made, long-lasting editions.

You'll receive our *Review* and a dated Reply Form 14 times a year (about every 4 weeks), plus up to 4 more offers of special selections. If you want the "Editors' Choice," do nothing—the book will come automatically. If you want another book, or no books at all, return the Reply Form by the date specified. A shipping and handling charge is added to each shipment.

Your HBC Guarantee: If you receive an unwanted "Editors' Choice" because you had less than 10 days to decide, simply return it and pay nothing.

History Book Club®
Camp Hill, PA 17012-8805

Please enroll me in History Book Club according to the no-risk, no-commitment terms outlined in the accompanying ad. Send me the four books whose numbers I have listed below. Bill me $1 each for the 3 choices on the left, and the fourth at the low Member's price—plus shipping and handling. **H170-10-0**

_____ _____ _____ _____

9-13

Name________________________________
(Please print clearly.)

Address________________________Apt. #________

City______________State________Zip________

All orders subject to approval. Prices generally higher in Canada.

mothering in hospitable host species are the focus of intensive research, because they offer the hope of overcoming serious difficulties that hamper natural breeding programs.

The National Zoological Park in Washington, D.C., which is celebrating its 100th anniversary this year, has emerged as a major center for research in such work. Workers at the zoo's New Opportunities in Animal Health Sciences Center are not underestimating the obstacles. "There's not an example of any species that is anywhere near being saved from extinction by the use of reproductive biotechnology," says David E. Wildt, the zoo's chief reproductive physiologist. "I'm concerned that people will start believing this is going to be simple, but it's not going to be simple."

In principle, artificial insemination and in vitro fertilization (IVF), aided by the freezing of germ plasm and embryos, can alleviate two problems that plague all efforts to increase the size of very small populations, such as those facing extinction. One is inbreeding and the other is loss of genes from the gene pool. Inbreeding tends to amplify genetic defects, leading to early death and reduced fertility; inbreeding also contributes to the loss of genes, which reduces the long-term fitness of a species. Sperm and eggs frozen in liquid nitrogen can be used to mitigate such effects by breeding animals from different sides of the world or even from widely separated generations. Frozen embryos can also be stored until a suitable recipient mother is available. C. Earle Pope of the Cincinnati Zoo notes IVF could be particularly effective in species, such as big cats, for which mating in captivity is difficult or dangerous.

In addition, when there is a shortage of females of a diminishing species, embryos can in principle be implanted in females of a closely related species. One hope for the future is that a way will be found to broaden the range of foster species. The rule of thumb is that only species that can hybridize can foster each other's offspring, according to the National Zoo's Mitchel C. Schiewe. But published experiments suggest that some of the barriers to interspecies embryo transfer might be circumvented by transplanting an early embryo's inner core of cells into the tissue that gives rise to the placenta in a different species.

When might artificial-breeding techniques bear animals? There have already been some impressive achievements. The San Diego Zoo has hatched rare Chinese Monal pheasants produced by artificial insemination. At the Cincinnati Zoo a bongo calf has been born from a surrogate eland mother, and an Indian desert cat was fertilized in vitro and brought to term by a domestic cat. Both the San Diego and the Cincinnati zoos have banked sperm, eggs and other cells from dozens of rare species in liquid nitrogen for research purposes and in the hope that they will preserve genetic diversity. National Zoo researchers helped to produce a Suni antelope by nonsurgical embryo transfer and are now establishing an embryo bank.

Yet in practice the biology of rare species is so idiosyncratic that only in a few well-studied cases can artificial techniques be relied on, according to Wildt. There are numerous obstacles. Little of the frozen semen in "frozen zoos" has been shown to be viable when thawed, for example. And the conditions needed to freeze and thaw embryos successfully vary even between strains of the same species.

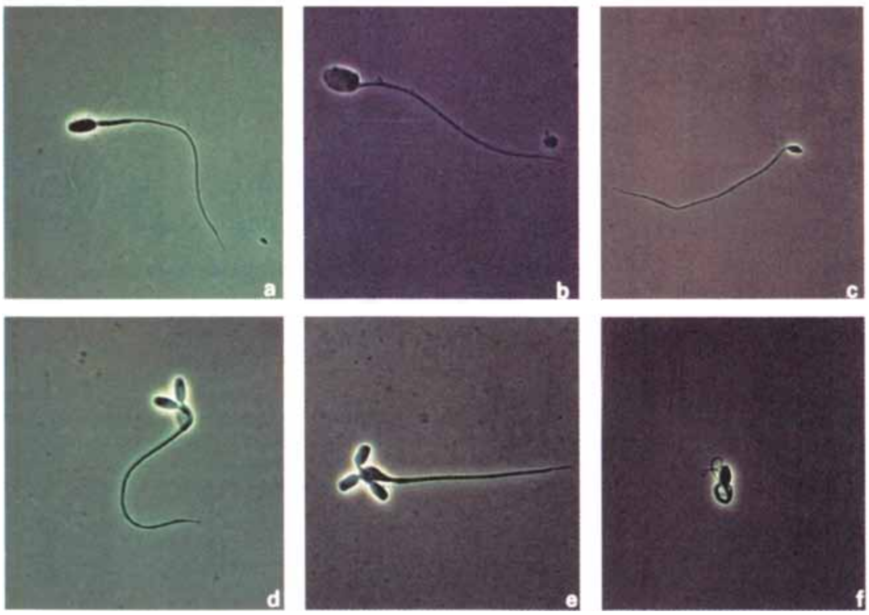
Merely discovering the timing of the estrous cycle in an exotic species—essential for knowing when to implant embryos or to collect eggs—can be a challenge. "Species tend to be more different than similar," says Steven L. Monfort, an endocrinologist at the National Zoo's Conservation and Research Center, who is studying, among other creatures, Eld's deer, the golden lion tamarin and Dall's sheep.

Some high-strung animals cannot even be restrained long enough to take a blood sample without causing trauma, and so Monfort and his colleagues have developed methods of measuring hormone metabolites in urine and feces. Even those efforts are sometimes inadequate. In order to learn about the reproductive cycle of the endangered scimitar-horned oryx, which researchers hope will someday foster the even more threatened Arabian oryx, workers have to pen females with a vasectomized male and watch their behavior.

The National Zoo's research program reflects Wildt's philosophy that the variability between species means far more basic research is needed before artificial-breeding techniques can routinely help to save rare species. "We're coming to a consensus that the major zoos ought to identify a few keystone species that will be used for a data base on the basic reproductive biologies and then see how far we can take it," Wildt says.

The strategy rests on the hope that there will be physiological similarities within taxonomic categories—among all cats, for example. At the National Zoo and at the Cincinnati Zoo, domestic cats are being studied as models

Most members of the cat family are suffering ill effects from a loss of genetic diversity



MALFORMED SPERM CELLS from the leopard cat, a small Asian cat, illustrate a problem that afflicts most felids. The leopard cat is actually less affected than most cat species: 60 percent of its sperm cells are normal, whereas in other species 60 percent are abnormal. Photographs by JoGayle Howard.

for reproduction in the other 36 members of the cat family, all of which are endangered and suffer from high rates of sperm defects caused by loss of genetic diversity. Cat embryos can now be frozen, thawed and successfully reimplanted to produce kittens.

The lessons seem to have been valuable. National Zoo researcher Annie M. Miller recently implanted tiger embryos produced by IVF into four tigresses, which may now be pregnant; she hopes eventually to use Siberian and Bengal tigers as surrogate mothers for the much rarer Sumatran tiger. Likewise, IVF research on pumas has made it possible to produce embryos of the rare Florida panther. Eggs and sperm are now being extracted from panthers found killed on highways.

Other species may yield to the approach. JoGayle Howard of the National Zoo is studying domestic ferrets as a model for the critically endangered black-footed ferret. Even so she must cope with technical problems. Freezing sperm is far from straightforward: the optimal rate of cooling and the type of "cryoprotectant" (antifreeze used to prevent ice crystals from damaging the cells) are critical and vary from one species to another. Howard needed three years to develop a technique for artificial insemination of ferrets with thawed semen, a procedure that must be done surgically.

Reproductive investigators readily agree that no amount of technology can safeguard a species if there is no habitat left to which the animals can be returned. "Habitat preservation is the most important thing anybody can do," Monfort says. Wildt emphasizes that field research to identify threatened species early and study them in the wild is essential if breeding programs are to make a difference.

Then there is the bottom line. Without money, rescues will be as rare as the Florida panther. "People don't understand that you've got to have a lot of money" if artificial-breeding technologies are to realize their potential, Wildt says. He points out that artificial insemination and embryo transfer in cattle became routine only through the efforts of hundreds of researchers and the expenditure of millions of dollars. The worldwide effort for endangered species, in contrast, is carried out by a handful of groups.

"People tend to think these reproductive technologies are going to save the world, and they're not," Monfort adds. "We're only going to be able to target a few species.... If I was responsible for just one species, that would be really something." —*Tim Beardsley*

SCIENTIFIC AMERICAN October 1989 Volume 261 Number 4

The Quiet Path to Technological Preeminence

The U.S. government is relying on ambitious research projects to spur commercial competitiveness. Instead it should speed the commercialization of new technologies wherever they may be developed

by Robert B. Reich

Alarmed by the continuing erosion of America's position in world markets, the U.S. government has embarked on ambitious research and development projects thought to have important commercial applications. Although neither the Reagan nor the Bush administration has announced it as such, the initiatives of the past three years amount to a major change in direction: policymakers otherwise devoted to the free market are pursuing what is in effect a targeted industrial policy for high technologies. Some of the projects may be motivated more by political considerations or the interests of pure science than by commercial requirements. Nevertheless, each has been justified to the public in terms of the nation's competitive needs.

In January, 1987, for example, the Reagan administration approved a $4.4-billion plan for a superconducting particle accelerator; a White House official deemed the project "critical" to the nation's future competitiveness and predicted that American companies would benefit from its spin-offs. Later in the year the president announced a "superconductivity initiative" aimed at developing practical applications for superconducting materials and called the technology "absolutely essential to our future competitiveness."

Early in 1988, warning that Japanese supercomputers were already on the market "with better performance than expected," the White House unveiled a five-year, billion-dollar "high-performance computing strategy." The administration also announced that the Defense Advanced Research Projects Agency (DARPA) would contribute $100 million a year to SEMATECH, a joint research venture by leading U.S. semiconductor makers.

Soon thereafter, initial construction contracts were awarded for the space station, buttressed by claims that the orbiting laboratory would be "vital to enhancing the nation's international competitiveness in the decades ahead." This year has been marked by a sudden focus on high-definition television: DARPA has proposed a two-year, $30-million research program to develop the technology, and the Bush administration is considering plans to grant antitrust immunities and tax breaks to American firms that produce HDTV equipment.

This high-visibility strategy for restoring U.S. technological prominence appears to require not only ever-greater government expenditures on the research and development of new technologies but also special incentives to boost private-sector research and development spending in areas not covered by such projects. Some policymakers are urging increased tax credits for research and development to spur commercial R&D spending. They point out that the portion of the gross national product devoted to research and development is lower today than it was 20 years ago and that commercial R&D makes up a smaller part of America's GNP than it does in either West Germany or Japan.

The alarm expressed by U.S. policymakers over the nation's loss of competitiveness in global high-technology markets is not misplaced. The U.S. share of the world semiconductor market, for example, dropped from 50 percent in 1984 to 37 percent in 1988 while Japan's share rose to more than 45 percent. American companies have virtually stopped selling dynamic random-access memory chips on the open market, and American makers of semiconductor-

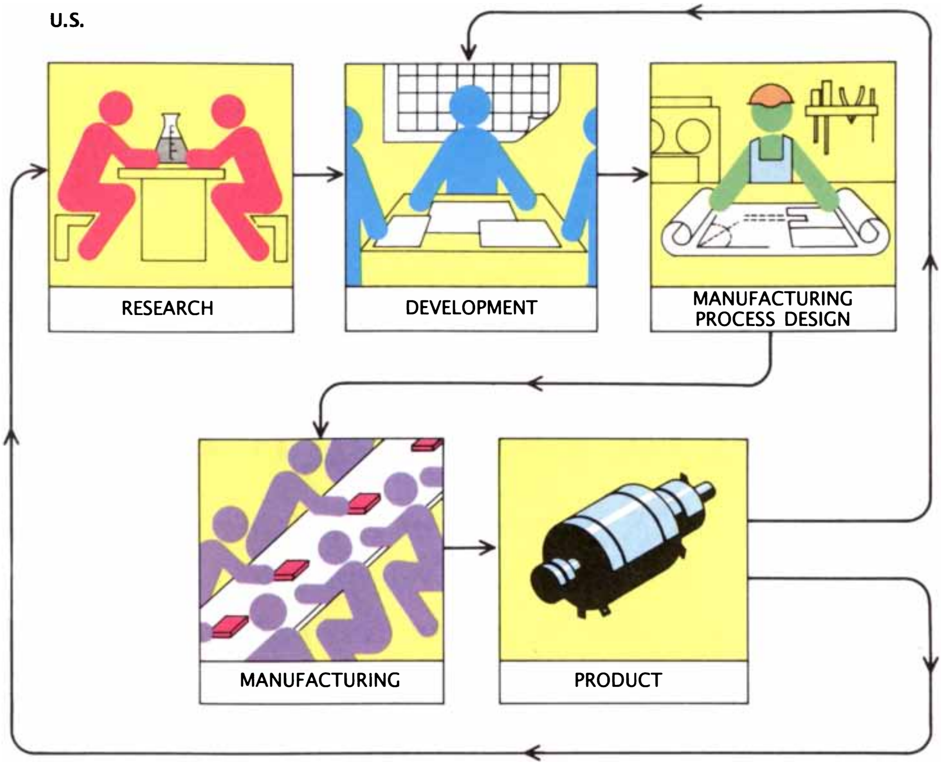
ROBERT B. REICH has taught political economy and management at Harvard University's John F. Kennedy School of Government since 1981. He was director of policy planning at the Federal Trade Commission under President Carter and served as assistant to the solicitor general during the Ford administration. Reich received a bachelor's degree from Dartmouth College, a master's degree from the University of Oxford and a law degree from Yale University. He is the author of several books on public policy and competitiveness, and he currently chairs the biotechnology review committee of Congress' Office of Technology Assessment.

manufacturing equipment are closing shop. Japan is taking over these fields. America's share of the world market for the consumer electronics products that incorporate a large fraction of semiconductors—products such as videocassette recorders, motor-driven 35-mm autofocus cameras and compact disk players—has dropped to about 5 percent since 1975; Japan's share has risen from about 10 percent to more than 25 percent. No U.S. company manufactures fax machines, for which the world market grew to $3 billion in 1988; again Japanese companies are in the forefront. Japan now dominates the numerically controlled machine-tool industry. Japanese companies are far ahead of American ones in the commercialization of high-definition television. Recent government studies warn that Japan is also ahead of the U.S. in applying superconducting materials.

In 1986 America's trade balance in high-technology goods such as semiconductors and communications equipment turned negative for the first time since data have been collected on high-technology trade. In 1987 and 1988, despite a sharp drop in the value of the dollar with respect to foreign currencies, the U.S. posted only a modest surplus in high-technology trade. This poor performance has been caused partly by Americans' desire for imported goods of all kinds but mainly by a loss of competitiveness in world markets.

Will the new path of ambitious research and development projects restore America's technological lead? It seems doubtful. The U.S. already leads the world in the quantity and quality of its research and development, but this lead has not yielded commercially competitive products. America's research universities, taken as a whole, are the best in the world. The research laboratories of America's largest corporations are unrivaled. American researchers write more than a third of all scientific and technical articles published around the globe, and they receive more U.S. patents than the rest of the world combined. Total spending on research and development in the U.S. is significantly higher than that in any other country and still three times the R&D spending in Japan.

There is no reason to suppose that more spending on research and development—even spending targeted to specific technologies—will result in commercial success. The problem lies in the inability of American companies (or, more accurately, the U.S.-based portions of what are fast



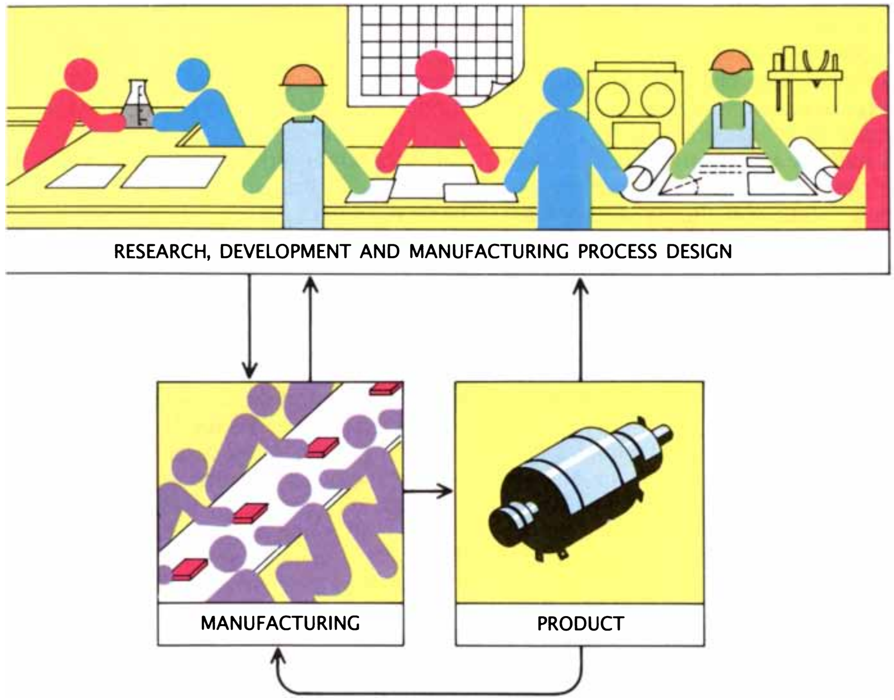
LINKS BETWEEN RESEARCH AND PRODUCTION in most U.S. companies are sequential (*left*); in Japan (*right*), research, product development and the design of manufacturing processes are carried out concurrently so that knowledge from one

becoming global technology firms) to transform discoveries quickly into high-quality products and into processes for designing, manufacturing, marketing and distributing such products. The fruits of research and development—new data, insights, inventions, prototypes—are easily disseminated across national borders. Increasingly the winners in the competitive race are the companies and nations that make use of those fruits most rapidly and comprehensively.

The Japanese have become adept at taking the essential insight of a major discovery—often made elsewhere—as the starting point in a process of application and refinement. American scientists invented the transistor. In 1953 Western Electric licensed the technology to the Sony Corporation; Sony rapidly made improvements on the transistor and launched a host of high-quality consumer electronics products. In 1968 Unimation, an American company, licensed Kawasaki Heavy Industries to make industrial robots; by the early 1970's industrial robots were on the job in Japan; by the end of the decade Japanese robots had come to the U.S. Meanwhile the nascent U.S. robotics industry never quite got on its feet. The video recorder was pioneered by California's Ampex Corporation and perfected by the Japanese. Canon, Inc., improved the basic concept behind Xerox Corporation's plain paper copier and then reworked the inexpensive copying technology it had developed to make inexpensive laser printers as well (another invention never fully exploited by Xerox).

The same pattern has held for composite materials and ceramics, color televisions, computer disk drives, basic oxygen furnaces, microwave ovens, computerized machine tools and other inventions. In the year ending March 31, 1987 (the last date for which such data are available), Japan purchased more than $1 billion of technological information from North America; less than half of that amount was purchased by Americans from Japan.

Japanese firms have been in an ideal bargaining position to buy much of this innovation cheaply. American companies often compete to sell technology to Japan, fearing that otherwise another firm will develop similar know-how and sell it first. (This logic may have motivated the Boeing Company, for example, to share its advanced airframe technology with Japan rather than see Airbus Industrie profit from a similar sale.) Further-



area can readily influence decisions made in other areas. A new concept moves back and forth among the different groups until it is perfected. The Japanese method speeds the transformation of new discoveries into commercial products.

more, small American producers of advanced technologies often lack the capacity to manufacture and market products on a global scale or to defend their patents worldwide, and so they fear that if they refuse to sell their most advanced designs they will eventually fall prey to reverse-engineered versions made at lower cost.

Japan has also coordinated its purchases so that its companies do not have to bid against one another. On several occasions its Ministry of International Trade and Industry (MITI) has acted as a clearinghouse, forcing foreign firms to license their patents and negotiating deals for all Japanese industry. Between 1956 and 1978, largely because of MITI's prohibition of auctions for licenses, Japanese firms paid only $9 billion for access to American technologies that cost between $500 billion and $1 trillion to develop.

It may be tempting to erect barriers to the international flow of technology: increasing patent and copyright protection for American firms, preventing foreign firms from purchasing U.S. technology or high-tech companies, barring foreign firms from gaining access to U.S. government-subsidized research, pressuring MITI to stop coordinating patent purchases. The core problem, however, has little to do with Japanese firms' easy access to U.S. technology. The real Japanese advantage lies in the ability to transform new inventions into high-quality products.

If the U.S. is to regain its technological prominence, it must improve the capacity of Americans to use technology. This quiet path back to competitiveness depends less on ambitious government R&D projects aimed at specific technology areas such as supercomputers or high-temperature superconductors than on improving the process by which technological insights—wherever they may be discovered around the globe—are transformed by American workers into high-quality products.

Close examination of Japan's success in this endeavor and America's comparative weakness suggests six steps along the path to competitiveness: scanning the globe for new technologies, linking government R&D funding to commercial products, integrating corporate research and development with production, managing the establishment of technological standards, investing in the technological learning of workers and providing a good basic education to all citizens. Taking these steps will be necessary, but not sufficient, for the U.S. to regain its competitive edge in the world marketplace.

The first step in rapidly assimilating new technologies is to discover what they are. American firms are often slow to learn of a new technological insight achieved elsewhere—whether it is a breakthrough invention, a more efficient method of fabricating and assembling products or a new way of organizing production and distribution. Some American researchers and engineers, whose formal education and early job experiences occurred when the U.S. was far ahead of other nations in developing and using technology, are simply skeptical of foreigners' abilities; anything "not invented here" is considered of little value.

Most U.S. firms are not organized for global scanning: they do not send their researchers, engineers and technicians to international conferences and trade shows or to visit their global competitors; they do not systematically gather data on the results of government-funded research in other nations (or even, for that matter, in the U.S.); nor do they systematically review technical and scientific journals and newspapers published in other countries. The U.S. government does little to aid them in such efforts.

Japanese firms regard global scanning for technological insights as an integral part of their business strategies. They use insights achieved elsewhere as a means of supplementing their own technological experience. Japanese companies organize study teams to visit American and European companies and university research laboratories, attend all relevant conferences and trade shows and carefully scrutinize foreign publications, including government reports. They even help to finance research and development in American universities and corporate laboratories—and then ensure that Japanese scientists, engineers and technicians monitor the results. (Last year about 5,000 Japanese scientists worked in U.S. laboratories; by most estimates fewer than 150 U.S. scientists worked in Japanese laboratories, most for only a few months.) Japanese government agencies also gather technological information from around the world and make it available to industry (MITI's Agency for Industrial Science and Technology, for example, funds an elaborate system of data gathering).

Japanese firms also enter into joint ventures with American firms for the express purpose of learning how to design and manufacture new products. Japan's long-term goal to produce jet aircraft, for example, is being furthered by the current joint venture with Boeing for the development of a new generation of airframes for mid-sized jetliners and by the recently approved agreement with General Dynamics Corporation to co-produce a successor to the F-16 jet fighter.

A second, closely related step in improving the assimilation of new technology is to link publicly funded research and development to commercial production. New R&D insights must be continuously available to be used in the production process, and new production insights must be continuously fed back into research and development.

One reason that Japan's publicly funded R&D efforts tend to be better integrated into commercial production than those of the U.S. is the dominance of defense-related R&D in this country. Defense accounts for only 3 percent of government-funded research and development in Japan, compared with 70 percent in the U.S. (a figure that makes up more than one third of U.S. R&D spending).

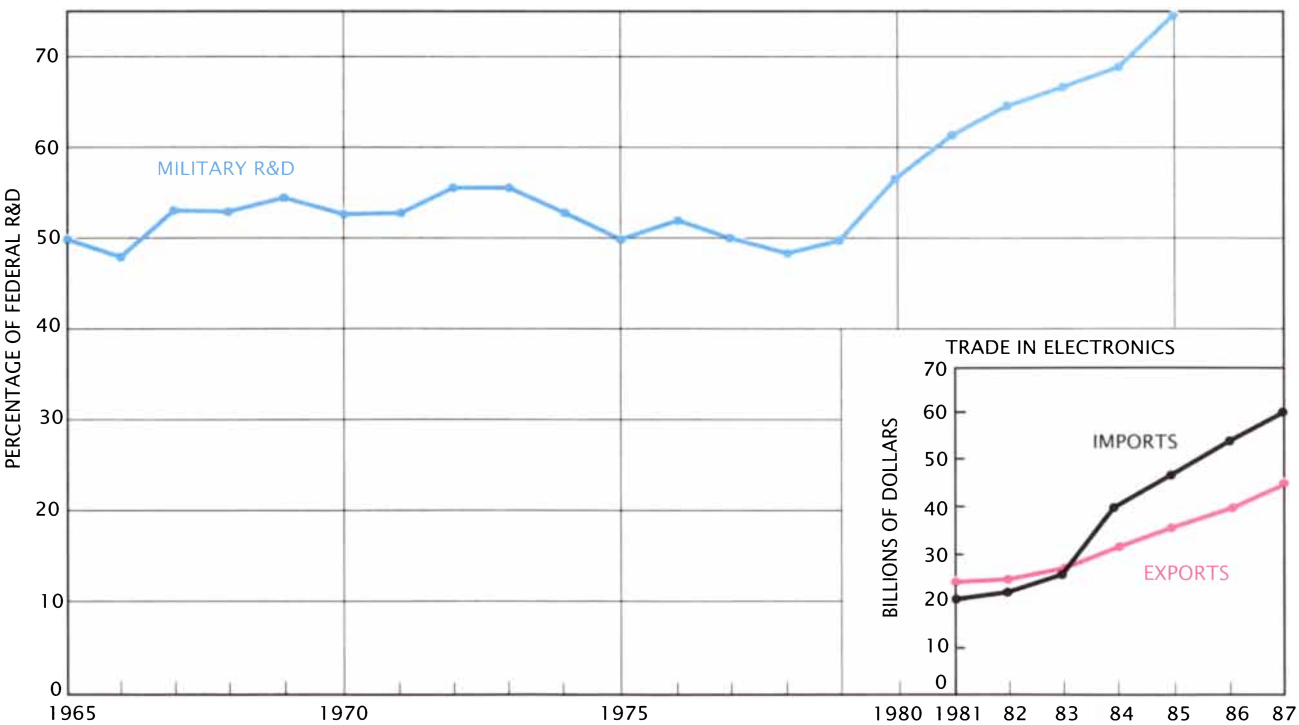
Many defense technologies have civilian uses, and there is a long history of civilian spin-offs from defense inventions, including computers, integrated circuits and high-strength materials. Nonetheless, several factors impede technology transfer from military to commercial applications. Much defense research is not accessible because it is classified. In many cases military specifications call for higher performance—and higher prices—than civilian consumers need or want. In others, commercial technology has outpaced the military state of the art so that defense research and development offers little knowledge of commercial value. The Department of Defense lags well behind the commercial sector in the use of digital electronics of all kinds, for example. Finally, those who develop military technology, be they defense contractors or employees of government laboratories, are often oriented toward the specific missions of their funding agencies rather than toward commercial development of their ideas.

Even though military R&D has become an inefficient means of generating commercial spin-offs, almost half of the federal funds allocated to research on superconductivity, for example, are directed at military applications such as infrared sensors, submarine detectors and electromagnetic rail guns. The Japanese government, on the other hand, has organized and funded a wide variety of projects focused expressly on potential commercial uses for superconductors.

Even nondefense research funding agencies in the U.S.—such as the National Institutes of Health, the Department of Energy and the National Science Foundation—tend to have concerns far removed from commercial applications. Primarily they support basic research.

Here again the contrast with Japan is stark. Although that government is spending more on basic research than in previous years, most of its R&D efforts are aimed at rapid commercialization. One of its explicit goals is to help small and medium-sized businesses adapt new technologies. The government operates a network of 195 regional laboratories specifically charged with providing technical assistance to small and medium-sized firms. The central government absorbs half of the cost; regional authorities and firms pay the rest.

In addition, Japan organizes groups of companies to undertake joint research on emerging technological problems. MITI's Agency for Industrial Science and Technology negotiates the terms of such consortiums and provides them with modest funding. MITI has recently established 28 regional technology centers designed to enable large firms to pool their research efforts.



DEFENSE-DOMINATED FUNDING skews U.S. research priorities and reduces the commercial benefits of research. More than half of U.S. government funding for research and development comes from the military, compared with less than 5 percent in Japan. Meanwhile (*inset*) the U.S. has become a net importer of electronic equipment and other high-technology goods.

Japan's strategy for HDTV illustrates the process. NHK, the national broadcasting company, began research on HDTV in 1970 and shared its results with the nation's 11 television manufacturers. Two government ministries took the lead in coordinating the research, parceling it out among firms so that they did not duplicate one another's work and ensuring that they shared their results. The Japan Development Bank matched corporate funding with public money. Even the Japanese post office began promoting HDTV by establishing savings plans for customers so that they will be able to provide a ready market for HDTV once it is commercially available.

Third, U.S. companies must link their own R&D efforts more closely to commercial production. Compared with Japanese firms, most American firms draw a sharper distinction between R&D on the one side and production and marketing on the other. Most U.S. corporate researchers and design engineers work in laboratories that are separated geographically as well as culturally from the factories, warehouses and distribution facilities where their ideas might eventually be implemented. Research facilities typically occupy modern, campuslike buildings in bucolic surroundings. Researchers and design engineers are often considered to be doing more important or more prestigious work than their compatriots on the factory floor are.

R&D often has relatively little connection to the rest of a company's undertakings. The proposals that emerge from the laboratories are scrutinized by market and financial analysts on a project-by-project basis and only then, if approved, turned over to manufacturing engineers, technicians and production workers who design and execute processes for making and distributing the products.

The implicit assumption behind this series of events is that product designs emerge essentially fully formed from research, after which they are put into production. It is not unusual for American companies to put off any consideration of manufacturing until researchers have come up with a generic solution to a broad problem and design engineers have transformed the broad solution into a specific design. This division prolongs product-development times and causes marketing opportunities to be lost.

In Japan research and development proceed simultaneously with manufacturing engineering and the design of production processes. There is no gap, either geographic or cultural. The process involves continuous trial and error as an innovation shuttles between development and production. Theory does not necessarily precede application, nor does product design necessarily precede process design; all phases occur together.

Only rarely is an entirely new product design formally introduced, reviewed and set into motion. Instead existing products and processes undergo continuous, incremental review as researchers, engineers and technicians gain insights and experience that may reveal possible improvements. One example of such incremental development is the use of ceramics in internal-combustion engines. Rather than developing fundamentally new engines incorporating ceramics, such as the ceramic turbine engines some U.S. companies are attempting to develop with federal help, Japanese firms are developing ceramic piston engines; they will be able to introduce ceramic parts incrementally into an otherwise conventional design.

Fourth, the U.S. must manage the early adoption of industrywide standards that render emerging technologies compatible with each other and speed commercial acceptance. Such standards make it easier for purchasers to experiment with equipment embodying new technology and reduce the risk of committing to a technology that quickly becomes obsolete. Standards adopted too early, however, may freeze the development of as yet immature technologies. Proper timing and widespread acceptance of standards is an important determinant of technological success.

In the U.S., technological standards are set with little regard to such issues. Large companies or government agencies set de facto standards; other governmental bodies or nonprofit organizations set official standards. Unfortunately, none of these sources of standards has explicit responsibility for managing the standards process to best promote a new technology. Nor do they have the staff to undertake such a job effectively.

The Federal Communications Commission recently rejected Japan's proposed HDTV standard on the grounds that it was incompatible with existing television sets. This decision had the effect of giving Zenith Electronics Corporation, the one remaining U.S.-owned television manufacturer (and any other U.S. firms that sense a profitable opportunity in HDTV) additional time to develop technology for the American market. That consequence, however, was inadvertent: the FCC has no mandate to spur American competitiveness and possesses no special insight into how to do so. Now that the FCC has rejected initial Japanese proposals for HDTV, the Japanese government is coordinating the development of a new standard that will be compatible with television sets now used in the U.S.

The Japanese take a more strategic approach to setting standards. Whereas they may allow competitors to vie with one another to establish the standard for a new consumer product, as occurred with the VHS and Beta standards for videocassettes, a department within MITI is specifically responsible for coordinating industrial standards to assure the efficient adoption of new technologies.

The fifth requirement for effective utilization of technology is a sophisticated work force, adept at discovering ways of incorporating new insights into products and production processes. Japanese companies typically spend many years developing and perfecting a technological insight once they have uncovered it or purchased it from anywhere around the globe. Sony's acquisition of video-recording technology from a U.S. company marked the start of a 19-year process of product development; the Betamax, introduced in 1975, was the company's fourth generation of video-recording technology. (Thanks to the automated production processes Sony developed, the Betamax cost only one hundredth as much to manufacture as the first generation of video recorders.) The same sequence holds for high-definition television: research got under way in 1970, and commercial production will not begin until 1991.

Contrary to popular belief in the U.S., however, such doggedness is not due solely to the longer-term view that the Japanese take of the potential gain from a given stream of technology. Japanese firms have shown themselves willing to invest in some technology streams that have no foreseeable results—indeed, even after problems to which the technology might be addressed are solved by a competing system. The investment in such cases is not so much in the technology itself as in the technological education gleaned by workers who grapple with it. Once educated, such technically sophisticated workers can recognize the potential value of a broad range of new technologies and adapt them continuously to new products and processes.

Long-term experience in trying to solve technical problems associated with recording on videotape and in manufacturing videotape recorders gave Sony's workers insights into many other consumer electronics problems—a valuable asset even though Betamax was a market failure. The company was successful in promoting a single standard for the new generation of 8-mm videotape recorders, and it was also able to incorporate some of its Beta technology into the VHS units it now manufactures.

Today many Japanese firms are searching for ways to use high-temperature superconductors. Their success will not be measured solely by whether the particular products or processes they aim for (such as Nippon Steel Corporation's attempt to develop continuous-strand casting techniques using superconducting magnets to levitate and confine molten steel) are ever achieved; it will also be a function of how broadly and deeply their work forces understand the commercial possibilities of this new stream of technology and can adapt it to other purposes in the future. Similarly, the 11 Japanese television makers now investing in HDTV understand that only a few of them will emerge as industry leaders, but they also know that their workers must gain experience in this new technology if they are to develop other applications of it in the future.

American firms are less willing than their Japanese counterparts to invest in the long-term technological learning of their workers, partly because the U.S. financial community demands short-term profits. This past April, for example, the president of Control Data Corporation announced that the company was abandoning its supercomputing business; he noted that CDC could no longer afford to lose $100 million a year trying to develop the technology. The company had been under intense pressure from its investors to increase the price of its shares. In contrast, an executive of Hitachi told the *New York Times* that his firm would continue to spend $100 million yearly on supercomputer R&D. "Financially speaking it is not such a good business," he said. "But supercomputers are the flagship of all computing technology."

American firms also have another, perhaps sounder reason for taking a short-term view of investments in their work force: American engineers change jobs frequently, and so the benefits of training acquired at one firm may well be reaped by another. (Japanese engineers and production workers, by contrast, tend to stay put for life, and so investments in their technological learning are more certain to return to the companies that trained them.) The restlessness of U.S. workers creates centers of technology like Route 128 around Boston and Silicon Valley in California, but it tends to discourage investment in human capital by any single firm.

As a result many American firms opt to purchase components or manufacturing processes that have already been proved cost-effective elsewhere. In U.S. managers' eyes the higher costs and greater risks of in-house development are not outweighed by the opportunity for engineers and production workers to learn a new technology from the inside and to improve on it in the future.

Sixth, the U.S. must improve the basic education of all citizens. Even if companies are willing to invest in the technological sophistication of their workers, the success of those investments depends fundamentally on workers' ability to learn, which depends in turn on the quality of their basic education. Here again the U.S. falls short. Whereas it is as successful as Japan in preparing the most talented and fortunate 20 percent of its population for managerial and professional jobs, the U.S. is far less successful in educating the remaining 80 percent. American 13-year-olds came in last in mathematics and almost last in science among 11 nations on one recent test administered here by the Department of Education and the National Science Foundation. Workers who are ignorant of science and mathematics will be unable to absorb new technological insights on the job and assimilate them into existing products and processes.

In short, the U.S. currently trails nations such as Japan in many areas crucial to commercial success because Americans are not prepared to apply new technologies effectively. The nation has begun to improve its ability to use new technology, but progress has been uneven and is not nearly fast enough to regain world markets.

PASTORAL SETTING of many U.S. research centers (such as this AT&T Bell Laboratories facility in Holmdel, N.J.) fosters creativity, but the separation of research laboratories from the factory floor can make U.S. companies less efficient than their Japanese counterparts at translating new discoveries into profitable products and the processes for producing them.

A few prominent firms such as IBM, Hewlett-Packard Company and 3M have stepped up their efforts to scan the world for technological insights and are seeking new ways to integrate their R&D with commercial production. Meanwhile some government agencies have begun to emphasize the utilization of new technology: Michigan's Modernization Service, for example, offers technical assistance in applying computer-aided design and manufacturing technologies to more than 6,000 small tool-and-die plants, machine shops and metal fabricators in that state. Recent federal legislation has authorized the National Institute for Standards and Technology (formerly the National Bureau of Standards) to help small businesses improve their productivity through the application of new manufacturing tools and methods. The National Science Foundation has funded a number of Engineering Research Centers around the country, each focusing on a different set of manufacturing process problems. Several federal laboratories are establishing offices of research and technological applications to help disseminate their discoveries; they have been authorized to grant exclusive licenses to commercial developers and thereby increase the financial incentives for product development.

Meanwhile primary and secondary school education is being overhauled in Arkansas, Minnesota and a few other states. Many states have tightened science and math requirements for high school graduation, and some states have raised teachers' salaries. (Although standardized tests are not the best of guides, average SAT scores in South Carolina, for example, have gone up by 36 points since the enactment of educational reform there in 1984. In California the proportion of high school graduates with more than three years of math or science has risen by roughly a sixth in the same period.) Many firms and local governments have exhibited new interest in on-the-job training to help workers handle a host of new technologies.

Much more than this will be needed. On the public side the quiet path back to technological leadership does not require significantly larger expenditures on research and development; instead it suggests a different allocation of money. Large-scale military-funded R&D projects should give way to joint development and production projects among U.S. firms, funded in part by a nondefense agency expressly charged with spurring the commercialization of new technologies. The emphasis in such projects should be on giving production engineers, design engineers and technicians experience in applying those new technologies. SEMATECH, the semiconductor joint venture, is a good start in this direction, but its funding agency, the Department of Defense, has no particular expertise in developing commercial technologies, and the project appears to be focused mainly on the development of semiconductor-fabrication equipment rather than on the diffusion of learning throughout the participating firms.

The U.S. must also invest much more than it does now in basic education and the training of its workers. Much has been written about this subject, and it is clear that simply providing more public money is no solution. In addition to conveying basic skills, primary and secondary school curriculums must emphasize critical thinking—a capacity to identify problems, raise questions and find structure in apparent disorder—rather than the mere regurgitation of facts. Teachers need to be given, and to accept, more responsibility for what is taught in the classroom and how it is taught. Parents and other members of the community must contribute more effectively to the educational mission.

On the private side the quiet path to technological preeminence necessitates a more collaborative relationship between researchers, design and production engineers and marketers within each firm and far greater efforts to retain skilled engineers.

Laws and regulations governing financial markets may have to be altered to allow companies greater latitude in funding long-term product development. For example, the federal government in effect subsidizes leveraged buy outs and takeovers; firms that borrow huge sums for such purposes can deduct interest payments from their taxable income. Tax subsidies could more effectively be targeted at activities that promote long-term growth. One step in the right direction would be to remove tax incentives for takeovers and buy outs, increase capital-gains taxes on the sale of assets held for less than six months and reduce such taxes on the sale of assets held for more than six years.

Perhaps the most difficult obstacle to overcome along the quiet path is the relative lack of visibility and drama. Large-scale research projects make headlines and provide tangible evidence that something is being done to improve U.S. competitiveness. Closer and more effective working relationships among government, business and educational institutions at all levels are difficult to measure and quantify. As a result, politicians, educators and business executives who follow the quiet path may have difficulty in claiming credit for progress along the way. Among all impediments this factor alone may prove to be the major stumbling block.

SIX STEPS BACK TO TECHNOLOGICAL PREEMINENCE

Scan the globe for new insights

Integrate government-funded research and development with commercial production

Integrate corporate research and development with commerical production

Establish technological standards

Invest in technological learning

Provide a good basic education to all citizens

RETURN TO PREEMINENCE in technology requires investment in workers' skills at all levels as well as in innovations tied to specific commercial goals. The U.S. has thus far lagged in both of these areas.

FURTHER READING

PROFITING FROM TECHNOLOGICAL INNOVATION: IMPLICATIONS FOR INTEGRATION, COLLABORATION, LICENSING AND PUBLIC POLICY. David J. Teece in *Research Policy*, Vol. 15, No. 6, pages 285-305; December, 1986.

DOES TECHNOLOGY POLICY MATTER? Henry Ergas in *Technology and Global Industry: Companies and Nations in the World Economy.* Edited by Bruce R. Guile and Harvey Brooks. National Academy Press, 1987.

THE DEFENSE INDUSTRIAL AND TECHNOLOGY BASE, VOL. 1. Defense Science Board, Office of the Undersecretary of Defense for Acquisition, October, 1988.

INDUSTRIAL POLICY OF JAPAN. Ryutaro Komiya et al. Academic Press, Inc., 1988.

THE SOURCES OF INNOVATION. Eric von Hippel. Oxford University Press, 1988.

TECHNOLOGY POLICY AND ITS EFFECT ON THE NATIONAL ECONOMY. U.S. House of Representatives, Committee on Science, Space and Technology. Technology Policy Task Force, U.S. Government Printing Office, Serial R, December, 1988.

ORGANIZING FOR MANUFACTURABLE DESIGN. James W. Dean, Jr., and Gerald I. Susman in *Harvard Business Review*, No. 1, pages 28-36; January/February, 1989.

The Mitotic Spindle

Just how this spindle-shaped biological machine parcels the DNA of a dividing cell into two equal clusters is only now becoming clear. The spindle turns out to be as dynamic as it is accurate

by J. Richard McIntosh and Kent L. McDonald

For an organism to grow, repair damaged tissues and reproduce, its cells must proliferate. They do so in two steps. First the parent cell grows; it synthesizes the material for two cells within one cell membrane, including a duplicate of every chromosome, or double-strand thread of DNA, in the nucleus. Then the cell divides.

Division begins with the process of mitosis, in which the previously duplicated chromosomes (and therefore the genes they carry) are separated from each other and parceled into two matching, well-segregated packages. Mitosis ensures that when the rest of the cell divides, in a process called cytokinesis, each daughter cell will contain the genetic information it needs to grow and divide.

Because the accurate segregation of chromosomes is profoundly important to the ability of daughter cells to survive and reproduce, many generations of investigators have attempted to learn how it is accomplished. The work has been marked by both success and frustration.

More than 100 years ago, biologists identified the paths followed by the chromosomes as they move to opposite poles in the parent cell. They found that the movements are determined to a great extent by an extraordinary system of fibers known (because of its shape) as the mitotic spindle. The importance of the spindle is perhaps made clearest by its absence: if spindle development is blocked experimentally, the chromosomes move very little and fail to segregate into two distinct sets.

J. RICHARD MCINTOSH AND KENT L. MCDONALD share a long-standing interest in the mitotic spindle. McIntosh is professor of molecular, cellular and developmental biology at the University of Colorado at Boulder. McDonald is a research associate in the High-Voltage Electron Microscope Laboratory at the same university.

Yet the details of spindle structure and behavior resisted discovery for a long time, in part because the component fibers are quite fragile. They are also too small and tightly clustered to be seen clearly in the light microscope, and they are too long and curvy to be tracked easily with the electron microscope.

Over the past 15 years, however, new labeling techniques and other improvements in microscopic methods have enabled us at the University of Colorado at Boulder and workers elsewhere to see the spindle in what is literally a new light. We now know a great deal about the changes the spindle undergoes in each stage of mitosis. We can also begin to discern an answer to what may be the most basic question of all: What controls the intricately choreographed activities of the spindle—and hence the movement of the chromosomes—during mitosis?

Mitosis is often described as having five consecutive stages, which collectively take about an hour to complete. The first stage, called prophase, begins when the DNA threads that were replicated during interphase (the period between cell divisions) condense to form distinct chromosomes, each one consisting of two genetically identical parts, called chromatids, bound together in a single unit. DNA compaction is essential, because the strands are normally so long and thin that in an uncondensed form they would become impossibly tangled by any attempt to segregate the chromatids.

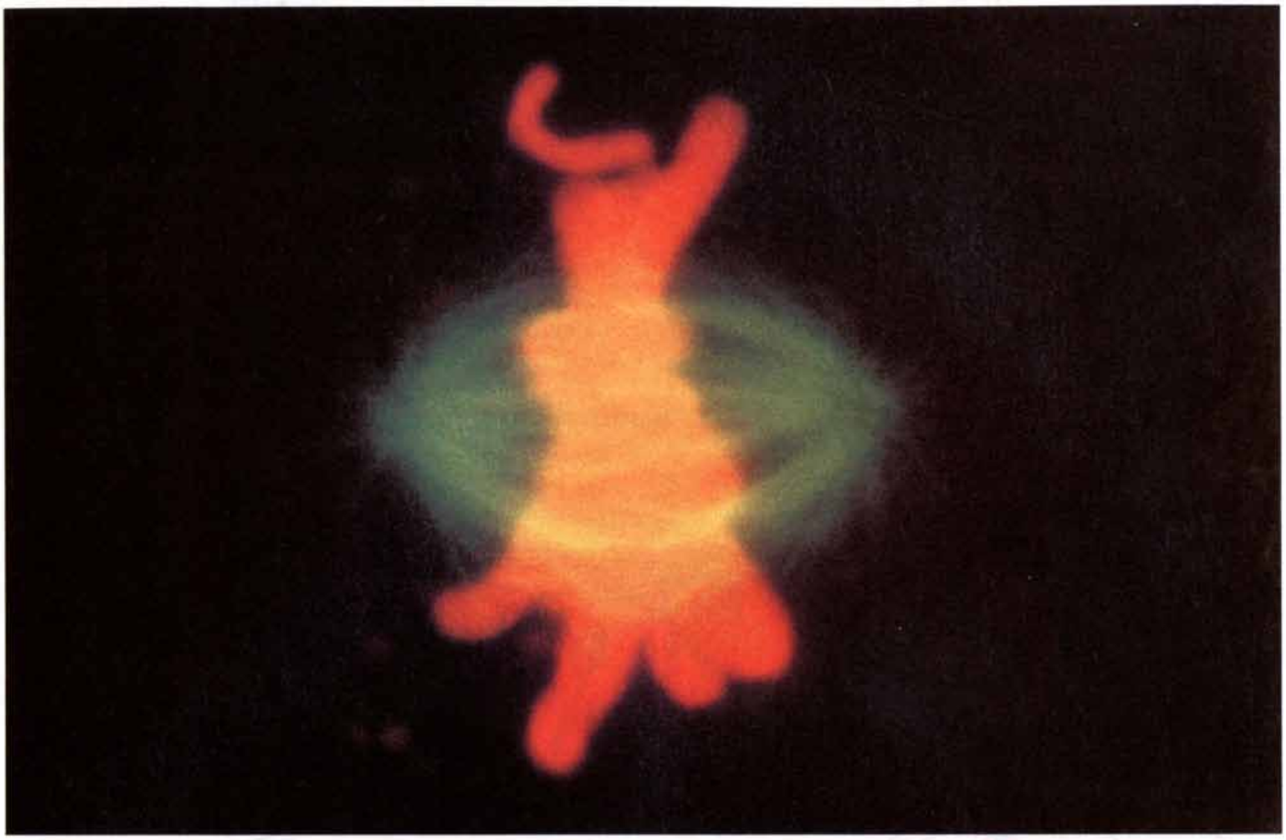
Toward the end of prophase, the spindle begins to take shape. It is formed primarily by fibers that extend from two specialized extranuclear regions called centrosomes. During the next stage—prometaphase—the fibers interact with the chromosomes, usually as a consequence of the fact that the envelope surrounding the cell nucleus breaks up.

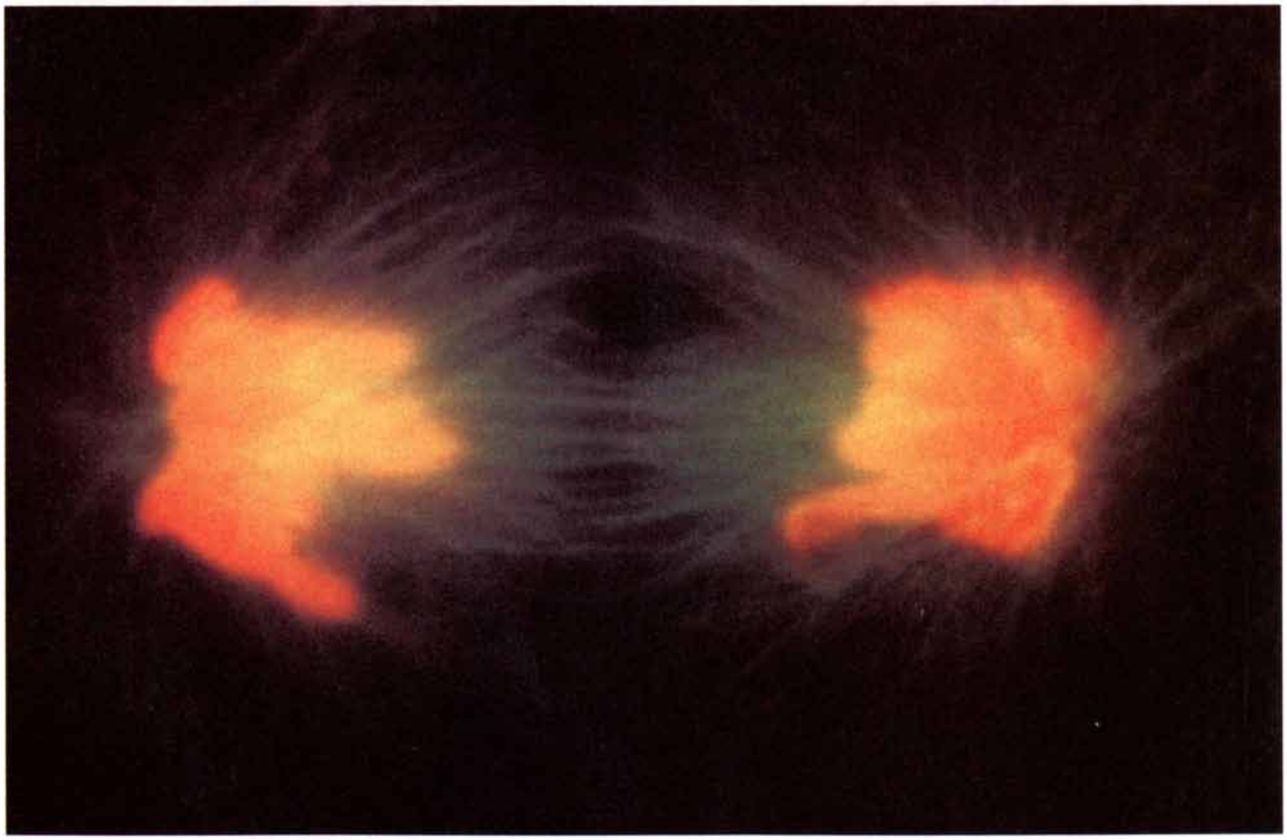
Initially the chromosomes are scattered throughout the nuclear region. However, once they become linked to fibers emanating from both centrosomes, they are drawn toward the equator of the spindle—the plane lying midway between the centrosomes. Within minutes the chromosomes become aligned at the midplane like football players at the line of scrimmage, at which point the cell is said to be in metaphase.

Soon after the chromosomes reach the midplane, the sister chromatids separate, becoming independent chromosomes—an event that initiates the fourth mitotic stage, anaphase. Now the divorced chromosomes migrate toward their respective centrosomes, which lie at opposite poles of the spindle. Also in this stage the spindle elongates, increasing the distance between the centrosomes and thereby increasing the distance between the separated chromosomes.

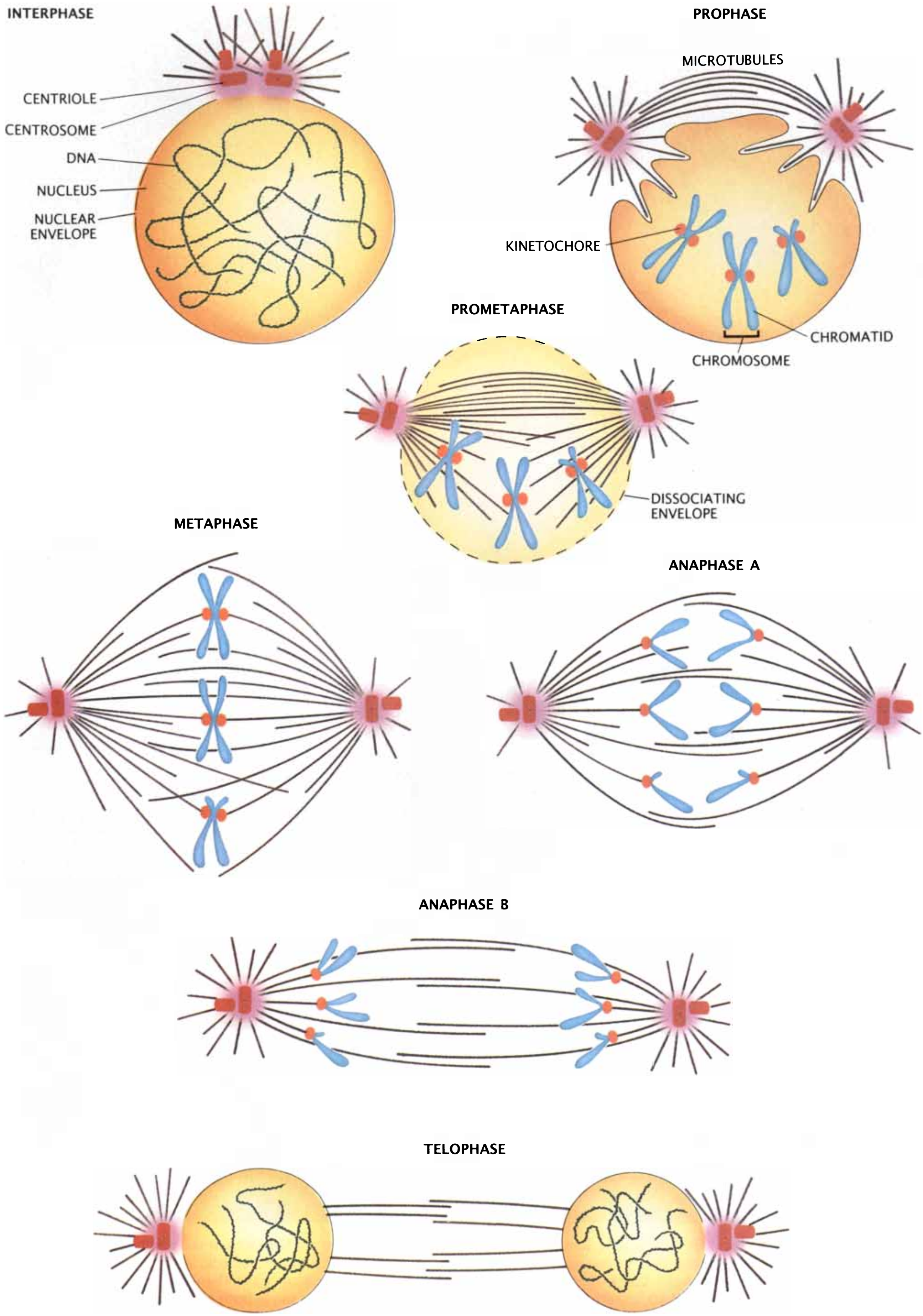
Late in anaphase a new nuclear envelope begins to form around each of the two clusters of chromosomes, initiating the final stage, telophase. During telophase the two sets of chromosomes decondense, giving rise to functional interphase nuclei. Soon after the two chromosome masses become well separated, cytokinesis ensues: the cytoplasm begins to divide and each nucleus, together with the material surrounding it, is partitioned into a discrete cell.

Such a description offers a serviceable overview of mitosis and more or less reflects the state of knowledge as of some 30 years ago. Yet its superficiality is unsatisfying to anyone interested in knowing exactly how a cell segregates its chromosomes. For an answer to that question, one has to delve into the mechanism of action of the spindle itself—a broad scientific problem that is being studied with





FIBERS of the spindle (*green*) in two mitotic kangaroo cells have been made visible by staining, as have the chromosomes (*orange-red*) attached to each spindle. One cell is in the stage of mitosis called metaphase (*top*): its chromosomes, which were initially scattered throughout the nucleus, are aligned at the spindle equator. The other cell, with two distinct sets of chromosomes, has advanced to the next stage, anaphase (*bottom*). Mark S. Ladinsky made both of these photomicrographs.


INTERPHASE
CENTRIOLE
CENTROSOME
DNA
NUCLEUS
NUCLEAR ENVELOPE
PROPHASE
MICROTUBULES
KINETOCHORE
CHROMATID
CHROMOSOME
PROMETAPHASE
DISSOCIATING ENVELOPE
METAPHASE
ANAPHASE A
ANAPHASE B
TELOPHASE

many approaches, including structural, biochemical and genetic analyses.

Most of our own work is focused on structure, because we regard the spindle as a machine for chromosome motion; if you want to understand how a machine works, you must first know something about the nature and behavior of its moving parts. This focus has recently yielded some valuable insights into the architecture of the spindle and how the structure changes as mitosis progresses. Let us, then, review some of the basic features of the spindle's construction before taking another look at the specific stages of mitosis and examining how each stage is accomplished.

The spindle is composed from a variety of elements. The most prominent of these, the fibers, are microtubules: polymers built from protein subunits that consist of two closely related molecules, alpha- and beta-tubulin. Microtubules have many roles in cells. For instance, they help to define a cell's shape, and they form the scaffolding for the cilia that promote cell motility [see "Microtubules," by Pierre Dustin; SCIENTIFIC AMERICAN, August, 1980].

One property of microtubules that is crucial to their functioning is the polarity, or asymmetry, they display along their surface and at both ends. Such polarity derives from the fact that the subunits, which are themselves asymmetric, are arranged head-to-tail down the length of the polymer. As a result, the microtubules can be thought of as long arrows, composed of smaller arrows (the subunits) all pointing in the same direction.

The asymmetry has at least two functional consequences, whose importance will become clear below: the polarity at the ends causes one end—the so-called plus end—to add and lose subunits faster than the other (the minus) end; the polarity along the surface influences the orientation with which proteins will bind to the microtubule surface.

Proteins that bind to microtubules are known to be integral parts of the mitotic spindle, although their specific roles in mitosis have not yet been identified. In cilia and many other microtubule-based biological machines, microtubule-binding proteins include not only structural molecules but also enzymes called motor molecules that convert chemical energy into mechanical work. Presumably the same is true for the spindle.

The centrosomes are another important part of the spindle. They differ considerably in structure from one organism to another, but they all share several properties. During early interphase, when each cell has only one of them, the centrosome is the cell's most important microtubule-organizing center. It serves as a seed to start microtubule polymerization, and it both influences the arrangement of the resulting microtubule clusters and defines their polarity. The centrosome replicates before mitosis, generally when the chromosomes are replicating. At first the sister centrosomes remain close together and function as one. They usually move apart during prophase, when they begin to initiate an increased number of microtubules, many of which will become the fibers of the mitotic spindle.

Although the chromosomes themselves are not parts of the spindle proper, each mitotic chromatid includes a protein-rich region known as a kinetochore that is an important spindle component. All kinetochores serve as couplings that enable the spindle fibers and chromosomes to interact; as soon as microtubules become attached to a kinetochore, the associated chromosome begins to move.

With these components defined, it became possible to determine their organization in the spindle with the electron microscope. For instance, Bill R. Brinkley, then at the M. D. Anderson Hospital in Houston, and other workers showed that there are two major classes of spindle microtubules. So-called kinetochore microtubules have one end anchored in a kinetochore; usually, but not always, their opposite ends are at or near the centrosome. Typically 15 to 35 microtubules growing from a centrosome bind to each kinetochore.

Most microtubules, however, do not interact with chromosomes. Studies we carried out on parts of mammalian spindles and studies we did in collaboration with other laboratories on the small, well-ordered spindles of a variety of microorganisms have shown that the nonkinetochore microtubules typically have one end in a centrosome, with the other end—which extends toward the other centrosome—free.

In collaboration with Jeremy D. Pickett-Heaps and David H. Tippit here at Boulder and with Urs-Peter Roos of the University of Zurich, we cut spindles in thin slices, photographed the slices in sequence with the electron microscope, and by computer recorded the location of each microtubule in each section. This approach enabled us to follow the trajectory of the individual microtubules and to locate their ends. It thereby provided important details about the length and function of the nonkinetochore fibers.

Specifically, we found that some of the nonkinetochore microtubules are short, but others are long enough to interdigitate with nonkinetochore microtubules from the opposite side of the spindle. Particularly during prometaphase and metaphase, many of the interdigitating microtubules extend a large fraction of the distance to the opposite centrosome, forming a framework that serves to keep the two centrosomes apart at the spindle poles. A few microtubules have both ends free; their significance for mitosis is unknown. (The spindles in many animal cells also include what are called aster microtubules that project out from the centrosome toward the cell periphery; they have no obvious role in the segregation of chromosomes but are important in cytokinesis.)

Even after this level of structural knowledge had been attained, we still did not know the polar orientation of the different classes of spindle micro-

ORGANIZATION AND SEGREGATION of chromosomes during mitosis involves many steps. The cell lays the groundwork during interphase, the period between cell divisions. In interphase the DNA in the nucleus replicates, as does a nonnuclear body known as the centrosome, which in animals consists of two rodlike structures called centrioles surrounded by a matrix. The centrosomes initiate the growth of microtubules, which become the spindle fibers during mitosis. In the first mitotic stage—prophase—the replicated DNA condenses to form distinct chromosomes composed of two identical, joined chromatids. Also, the centrosomes move apart, and the newly emerging spindle fibers deform the nuclear envelope. The envelope breaks up in prometaphase, and the chromatids become attached to the spindle fibers at a specialized region known as the kinetochore. Once linked to both centrosomes, the chromosomes move toward the spindle equator, where they eventually form what is called the metaphase array. Soon thereafter the joined chromatids separate, becoming independent chromosomes and initiating anaphase, a two-part stage. First each chromosome approaches the centrosome to which it is connected. Then the spindle elongates, increasing the distance between the separated chromosomes. Finally, a nuclear envelope forms around each of the two DNA clusters, marking the cell's entry into telophase, during which the DNA decondenses.

tubules. Such information was important to obtain because it would provide clues to how the spindle forms; it would also help to limit the universe of possible explanations for various events in mitosis.

To make better sense of the latter reason, recall that the polarity of microtubules influences the orientation of bound proteins, presumably including bound motor molecules. The orientation, in turn, influences the direction in which certain motor molecules travel along a microtubule; it can also be expected to influence the force exerted by a motor molecule on the microtubule itself or on other bodies to which the motor might also be attached, such as a chromosome or a nearby microtubule. Hence, knowing where, say, the plus ends of kinetochore microtubules are located would make it possible to determine such things as whether a motor molecule known to move in the plus-to-minus direction could potentially be involved in moving chromosomes toward the centrosome during prometaphase or anaphase.

Microtubule polarity is not visible by conventional light and electron microscopes, but in the early 1980's Steven R. Heidemann and Ursula Euteneuer at Boulder and Leah T. Haimo and Bruce R. Telzer in Joel L. Rosenbaum's laboratory at Yale University developed methods for visualizing such polarity. Their studies agreed that all spindle microtubules, including those attached to kinetochores, are oriented with their plus (fast-growing) ends pointing away from the centrosomes to which they are connected.

Recently Marc W. Kirschner and Timothy J. Mitchison of the University of California at San Francisco have drawn a mass of structural findings and data about microtubule polymerization into a single powerful model of how the spindle forms. They suggest that during mitosis microtubules are continually being initiated at the centrosomes and growing by subunit addition only at their plus ends (the ends distant from the centrosome). Microtubules are known to grow for a while at a constant rate and then to change to a state in which they rapidly disassemble—a property Mitchison and Kirschner have called dynamic instability. As a consequence, many of the microtubules disappear essentially as soon as they form. Others, however, become stabilized and survive long enough to build the spindle and carry out its functions. These subpopulations become stabilized either by binding to a kinetochore (in the case of the kinetochore fibers) or by interacting with microtubules from the opposite pole (in the case of the interdigitating fibers).

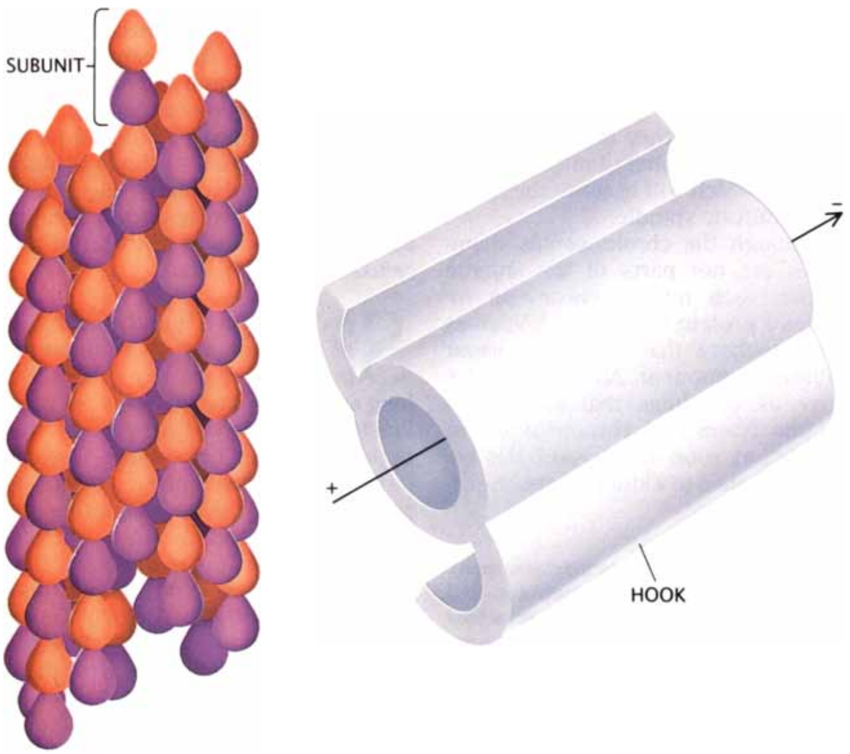
We are now in a position to recast our earlier description of mitosis and to speculate about just how the spindle engineers the segregation of chromosomes during mitosis. Let us pick up the discussion at the point when the microtubules begin to interact with the kinetochores—that is, at the start of prometaphase.

As soon as even one microtubule binds to a kinetochore, the attached chromosome (which still consists of joined chromatids) begins to move toward the centrosome to which it is now linked. This motion might result either from an active kinetochore exerting force on a static microtubule or from the shortening of a microtubule to which an essentially passive chromosome is bound.

The chromosome does not remain near the centrosome for long, however, because as soon as the sister kinetochore is bound by one or more fibers from the opposite chromosome, the chromosome becomes an object of a tug-of-war: it behaves as if it were being pulled to and fro by the two opposing centrosomes. Ultimately all the to-ing and fro-ing moves the chromosomes to the equator of the spindle, where they become a part of the metaphase array. In metaphase the chromosomes are arranged so that each kinetochore faces the centrosome to which it is linked.

Many of the mechanisms of prometaphase are still poorly understood, but the mechanism by which a cell assures that the two kinetochores on each chromosome are connected to opposite centrosomes has been partially clarified. R. Bruce Nicklas and Donna F. Kubai of Duke University have shown that kinetochores can associate initially with microtubules coming from practically any direction, although the association is weak at first. Once a bipolar attachment is achieved, the association of a chromosome with the spindle becomes firm.

Nicklas's work suggests that the stability of the chromosome's attachment to the spindle is established by the

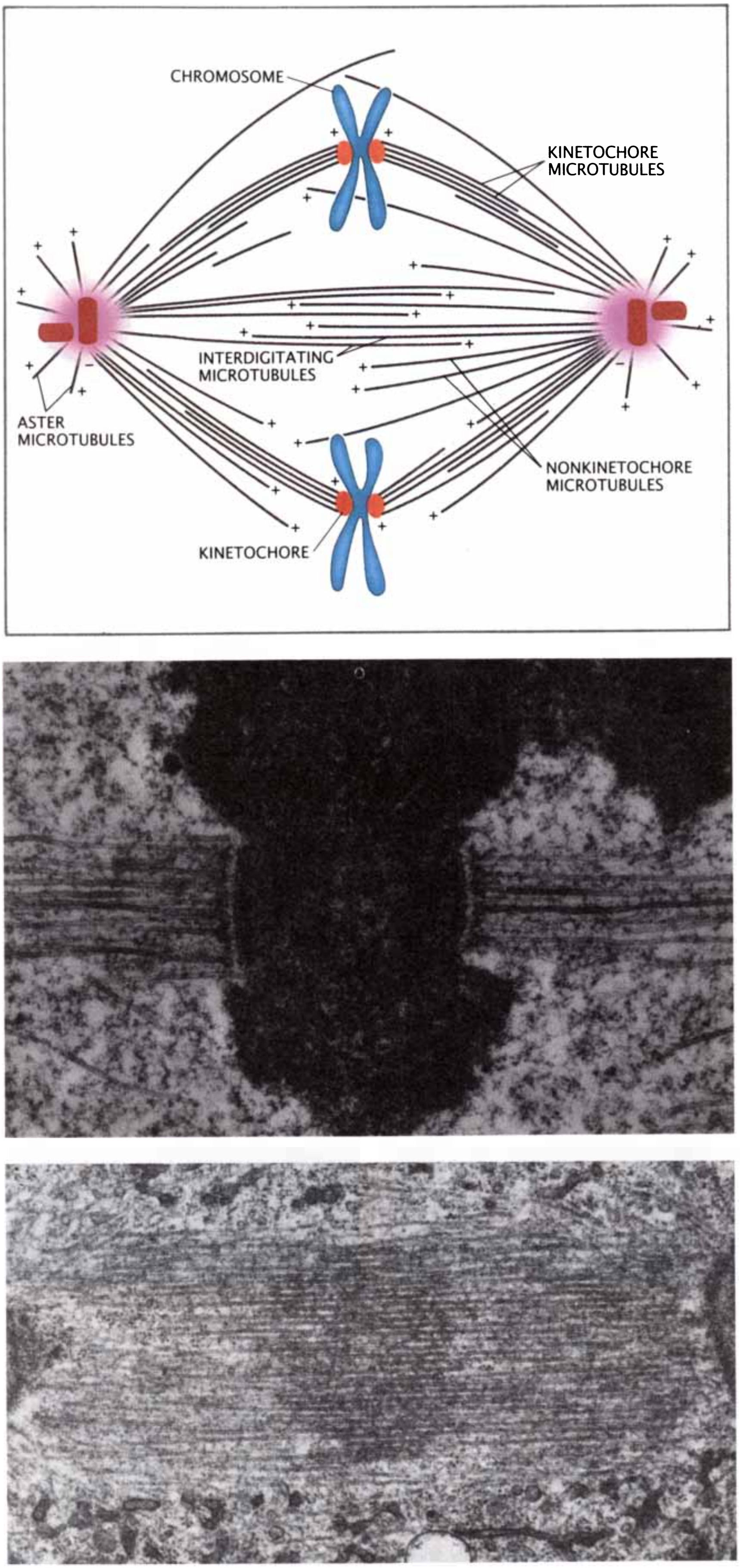


MICROTUBULE is an inelastic, tubelike polymer built from subunits consisting of the protein tubulin (*left*). Tubulin subunits are themselves formed from two distinct protein molecules—alpha- and beta-tubulin (here represented by different colors). Because the subunits are always arranged head-to-tail, the microtubules are polar: their ends are structurally—and, it turns out, functionally—different. One end, the so-called plus end, gains and loses subunits faster than the other (the minus) end. Workers can now distinguish the plus and minus ends by various methods. For example, they can induce the microtubules to "sprout" hook-shaped sheets of tubulin (*right*). If the hooks curve clockwise as they come off the original microtubule, the viewer is looking at the microtubule from the plus toward the minus end.

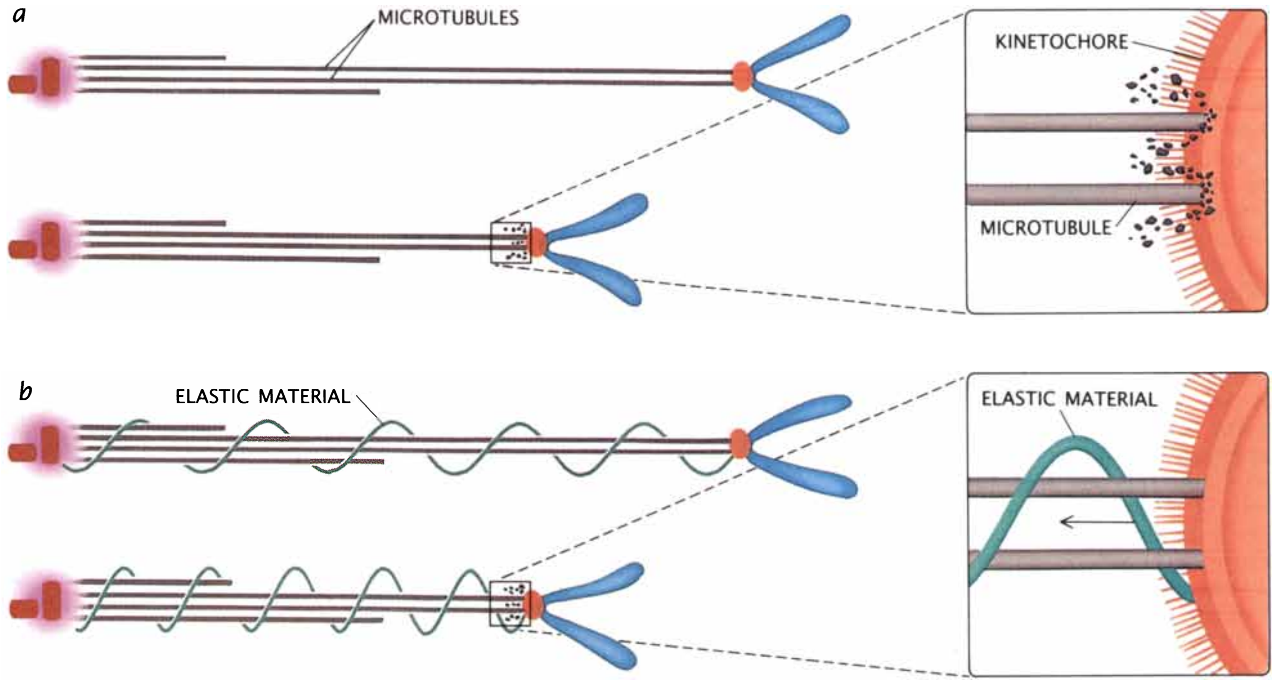
tension exerted when sister kinetochores on a single chromosome are pulled in opposite directions. If the sister kinetochores of a single prometaphase chromosome become linked by microtubules to the same centrosome, the chromosome will spontaneously detach from the spindle. If, however, a microneedle is i
a chromosome and pulled so as to exert the tension normally provided by microtubules from the opposite centrosome, the chromosome does not detach from the spindle. Remarkably, its two kinetochores remain linked to the single centrosome until the experimenter tires or the cell enters anaphase. The fact that the attachment of chromosomes to spindle fibers is stable only under load helps to assure that one chromatid in a pair will migrate to each pole of the mitotic spindle.

The cell usually accomplishes the stable, bipolar attachment of chromosomes to the spindle well before all the chromosomes congregate in the metaphase array at the spindle midplane. How, then, do the kinetochore fibers attached to opposite centrosomes manage to bring a chromosome to the spindle midplane?

Gunnar Östergren of the University of Uppsala in Sweden proposed in the early 1950's that a cell can adjust the magnitude of the forces acting on a kinetochore so that the net force is proportional to the length of the kinetochore-bound fibers. The idea is



COMPLETED SPINDLE of a typical animal cell in metaphase includes different categories of microtubules, classified according to where their ends terminate (*top*). There are two major classes. One class consists of microtubules having one end bound to a kinetochore; many but not all kinetochore microtubules are also bound at their opposite end to a centrosome. Usually a number of microtubules bind to the same kinetochore, as is evident in the top electron micrograph, made by Matthew J. Schibler of the La Jolla Cancer Research Foundation. The other major class consists of microtubules that grow from a centrosome but do not bind to a kinetochore. These nonkinetochore fibers include microtubules that interdigitate at the spindle midplane, as well as shorter fibers and a few that float free. The nonkinetochore microtubules of some microorganisms are particularly well organized, as is evident in the bottom micrograph. Many animal spindles also include radially oriented "aster" microtubules. The fast-growing (plus) ends of the spindle microtubules are located away from the centrosome that initiated the microtubules.



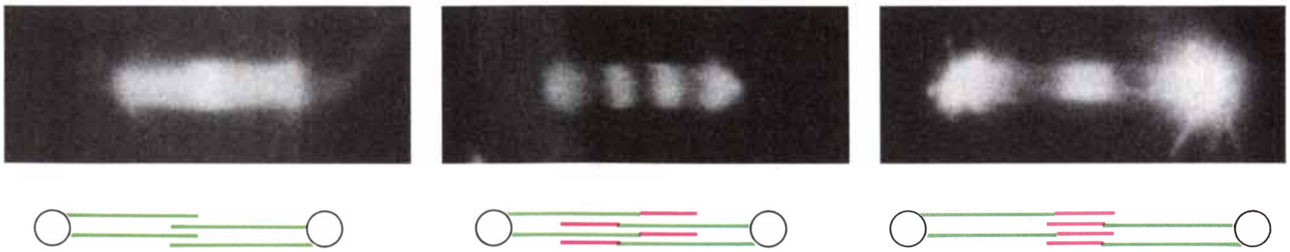
HOW CHROMOSOMES MIGRATE toward the centrosomes (the spindle poles) during anaphase A is a subject of much speculation. The chromosomes were once thought to be reeled in by the kinetochore microtubules as the microtubules shortened at their centrosome end. Actually, the microtubules lose subunits from the kinetochore end. A kinetochore may actively disassemble bound microtubules and then repeatedly grasp the remaining parts of one or more shortening microtubules as if the fibers were lifelines (*a*). It is also possible that an as yet unidentified elastic substance is stretched between the kinetochore and a pole, in which case the microtubules might restrain the elastic substance from pulling on a chromosome (*b*). Depolymerization of the microtubules would then permit the elastic substance to draw the genetic material poleward.

attractive, in part because it is consistent with the finding that during prometaphase long kinetochore microtubules shorten, and the shorter fibers coming from the opposite centrosome lengthen, until the fiber lengths eventually equalize.

Some evidence obtained by Thomas S. Hays and Dwayne A. Wise in Edward D. Salmon's laboratory at the University of North Carolina at Chapel Hill indicates that Östergren's hypothesis is correct. Further, Hays has extended the work, showing that the strength of the pulling force is also proportional to the number of microtubules attached to a kinetochore. More microtubules are associated with greater force.

What agent does the pulling, and by what mechanism is the strength of the force coupled to the length of the kinetochore fibers? One possibility is that the microtubules do the pulling in response to forces exerted on them by bound motor proteins; longer fibers and larger numbers of fibers would exert more force than shorter and fewer microtubules because they can bind more force-producing proteins. There are other logical explanations, however. We would also like to know what controls the assembly and disassembly of kinetochore microtubules as the chromosomes move toward the spindle equator.

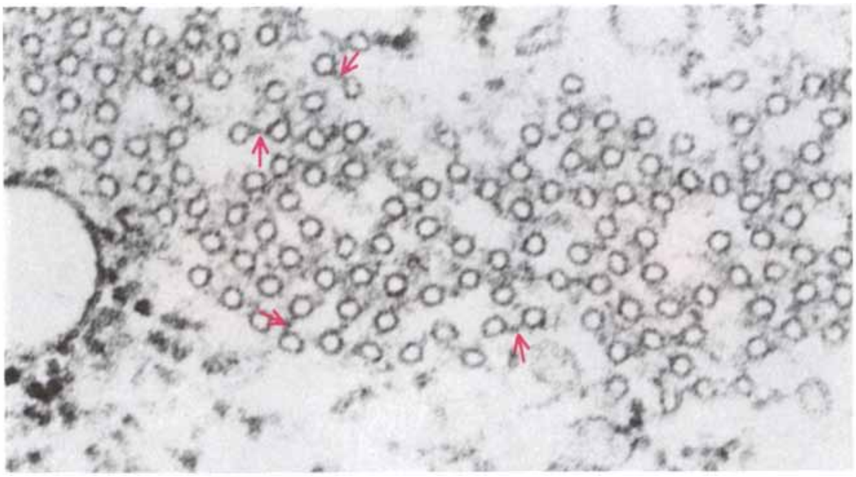
Regardless of how chromosomes find their way to the metaphase position, once they are there, their chromatids are in the correct arrangement for unimpeded movement to their respective poles. Indeed, the metaphase



ELONGATION of the spindle during anaphase B is thought to be accomplished by the growth of interdigitating microtubules at the plus ends and by the subsequent sliding of the lengthened fibers. Hirohisa Masuda and W. Zacheus Cande of the University of California at Berkeley have collected evidence of both processes. They first stained the native microtubules of an isolated spindle with a fluorescent dye (*top left and map below*); the stain is darkest at the center because the interdigitating fibers overlap there. They then added tubulin labeled with a different dye. A photomicrograph (*center*) showing only the newly incorporated tubulin (*red in maps*) reveals increased overlap, indicating that the fibers grew by adding the new tubulin to their plus ends. (Aster microtubules are also stained, but the finding is irrelevant to spindle elongation.) Next, they added adenosine triphosphate, a fuel known to be required for spindle elongation (*right*). The size of the overlap zone then became smaller, and the spindle as a whole elongated—results that can be explained only by the sliding of the labeled ends.

alignment is so important to the segregation process that when individual chromosomes dawdle in getting to the spindle equator, the onset of anaphase is usually delayed. Apparently some cellular mechanism monitors the arrival of the chromosomes at the equator and delays the onset of anaphase until any renegades take up their proper positions.

The separation of sister chromatids, which ends metaphase and initiates anaphase, is one step in mitosis that is apparently independent of spindle activity: it takes place in some cells even when the spindle is absent. Some evidence suggests that the trigger is an increased concentration of calcium ions released from membrane-bound vesicles in the nuclear region.



INTERDIGITATING MICROTUBULES (*small circles*), shown in cross section, are linked to one another in the zone of overlap by cross bridges (*indicated by arrows*). The bridges, which can be visualized in electron micrographs, may be part of the machinery that pushes the two halves of the mitotic spindle apart late in mitosis.

Once the separation of the chromatids is complete, anaphase proceeds, in two phases. First, in what is called anaphase A, the kinetochore microtubules shorten as the chromosomes (the newly separated chromatids) migrate toward their respective poles. Then, in anaphase B (which may overlap the first phase to an extent), the spindle elongates.

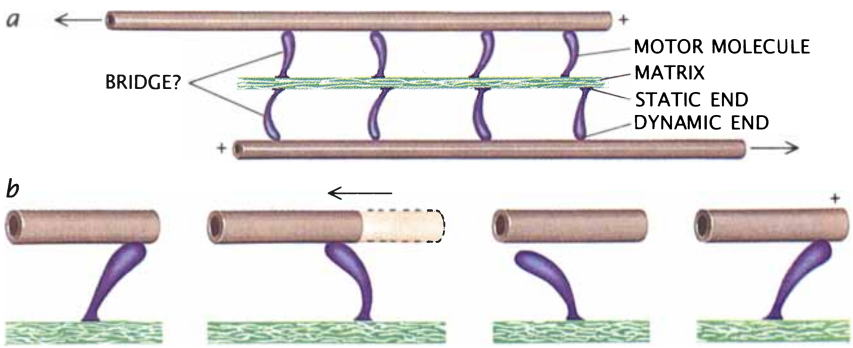
Gary G. Borisy's laboratory at the University of Wisconsin at Madison and Kirschner's and Nicklas's laboratories have each found evidence that during anaphase A the kinetochore microtubules shorten by the loss of subunits at their kinetochore end; apparently the chromosomes are not, as had been thought, reeled in passively by the shortening of the fibers at the centrosome. Indeed, the fact that the chromosomes remain attached to the microtubules as the polymers disassemble suggests that the kinetochores are quite active and might contribute to microtubule disassembly.

What interactions between the kinetochore and the kinetochore microtubules enable chromosomes to move toward their respective centrosomes? One possibility is that something, perhaps the kinetochore itself, exerts a compressive (subunit-removing) force on the microtubules—possibly as a consequence of being pulled by motor molecules that are attached to the kinetochores and that move down the microtubules toward the centrosome end. The chromosomes would then remain attached to the shortening microtubule clusters by repeatedly grasping the remaining parts of one or more microtubules as if they were lifelines. If the kinetochore is indeed as active as the model suggests and exerts force on the microtubules during anaphase, the finding would suggest that earlier—in prometaphase—the kinetochore may also contribute to chromosome motion, helping to propel the chromosomes toward the spindle equator.

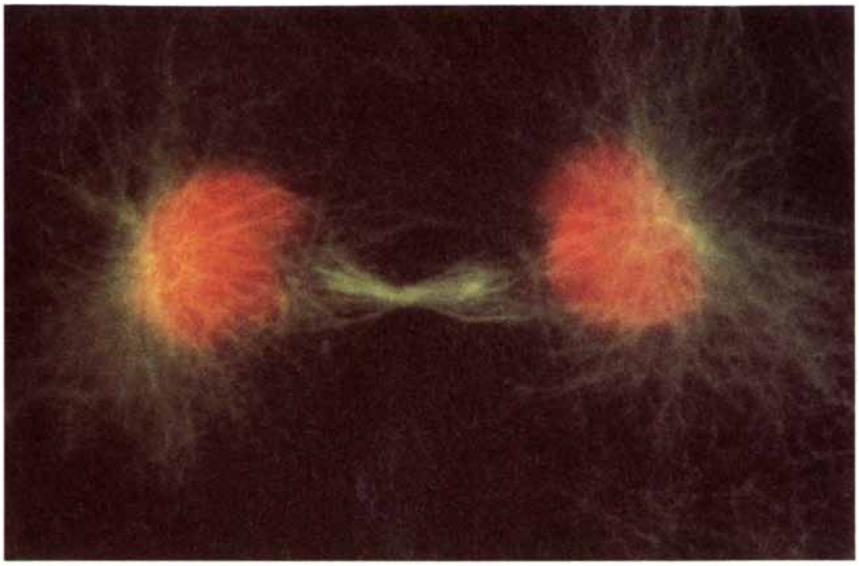
An alternative explanation for chromosome movement in anaphase suggests that an elastic component of some kind stretches between the kinetochore and the pole and that the kinetochore microtubules serve simply to restrain this pulling motion. As the microtubules shortened, then, they would permit the elastic force to pull the chromosomes poleward. There is no evidence that such an elastic system exists, however.

The mechanism by which the spindle poles and the attached chromosomes are pushed far apart during anaphase B is better understood. In at least some organisms, it is clear that the interdigitating fibers elongate at their plus (centrosome-distant) ends; then motor molecules in the midplane of the spindle, where the interdigitating microtubules overlap, push the fibers from opposite centrosomes away from one another, so that the fibers slide toward the opposite poles. This motion forces the centrosomes (and therefore the chromatids attached to them) quite a distance apart.

We gained our first inkling that a sliding mechanism might exist when we were immersed in electron microscopic analyses of spindle structure in the 1970's with Pickett-Heaps and Tippit. By examin-



SLIDING of interdigitating microtubules during anaphase B might be accomplished by so-called motor molecules, which are thought to be components of a matrix that has been identified in the zone of microtubule overlap. The model shown here (*a*) suggests that the motor molecules reach out to the microtubules and push the fibers toward the periphery of the cell (in the plus-to-minus direction). That is (*left to right in detail, b*), the motors are thought to attach to the microtubule surface, push the fiber poleward, detach from the surface and reattach at a position nearer to the plus end of the microtubule. In electron micrographs two nearby motor molecules acting on microtubules from opposite poles could look like a single bridge.



KANGAROO CELL in telophase includes two complete nuclei in the process of returning to the interphase condition; only the spindle fibers (*green*) and the DNA (*red*) are visible. The cell has virtually completed cytoplasmic cleavage. The fluorescent bar of microtubules visible in the center of the image is actually in the isthmus that still connects the sister cells. Ladinsky made the photomicrograph.

ing the structure of nonkinetochore microtubules, we noticed that the spindle microtubules from opposite poles looked as if they had slid apart between the end of anaphase A and the end of anaphase B. We saw similar patterns in spindles from mammalian cells, the planktonic algae known as diatoms and the slime mold *Dictyostelium discoideum*. In each instance, we also found that interdigitating microtubules had increased in length during anaphase B.

The evidence that the force-generating mechanism for spindle elongation resides in the spindle itself is based largely on work done in the mid-1980's by W. Zacheus Cande of the University of California at Berkeley and his colleagues. They isolated spindles from diatom cells and added the biological fuel adenosine triphosphate (ATP), whereupon the half spindles (each consisting of a centrosome and its associated fibers) slid apart. If the spindles did not include their own sliding forces, the addition of the ATP would have had no effect.

Linda Wordeman in Cande's laboratory further showed that proteins located in the zone of overlap had to be activated by the binding of a phosphate group from ATP in order for the spindle to elongate. Cande and his colleague Hirohisa Masuda filled in another piece of the anaphase-B process by showing that the elongation of the interdigitating microtubules in the diatom occurs by subunit addition at the plus (centrosome-distant) end.

How is it that the interdigitating fibers elongate at a time when the rest of the spindle fibers are disassembling, and how is the sliding accomplished? One plausible answer to the first question has to do with the fact that during anaphase the interdigitating microtubules become swathed in the area of overlap by a protein-containing matrix. This matrix may well confer an assembly advantage on the fibers. For instance, by binding to the microtubules, molecules of the matrix might stabilize them, much as kinetochore binding seems to do.

The sliding of interdigitating microtubules, which our colleague William M. Saxton has now demonstrated in living mammalian cells, might be generated in several ways. All the mechanisms proposed are still a mixture of fact and speculation, however. One model proposes that the pushing motors are anchored in the matrix of the overlap region. From there they bind to a microtubule and push it in the plus-to-minus direction. Such activity would push the interdigitating fibers in opposite directions, toward the cell periphery. Bridges between interdigitating microtubules can be seen by electron microscopy and may consist of the postulated motors [*see illustrations on preceding page*].

What, in summary, do the collected findings say about the normal functioning of the mitotic spindle? The spindle turns out to be a surprisingly dynamic structure. Its microtubules are nucleated at the centrosomes and then grow and disassemble rapidly as mitosis progresses. Most of the growth and disassembly apparently takes place at the end of the microtubule away from the pole. Although microtubules all have the same polarity, the fibers behave differently depending on the structures in the spindle to which they bind. If they associate with the kinetochore region on chromosomes, they participate in the movement of chromosomes during prometaphase and anaphase A. If they interact with microtubules from the opposite pole, they form a bundle of interdigitating microtubules that holds the poles apart. By polymerization and sliding, these interdigitating fibers also appear to elongate the spindle during anaphase B and thus to increase the segregation of chromosomes.

The solution to the question of what controls the movements of chromosomes during mitosis will ultimately depend on an understanding of how the various molecules in the spindle interact. Biochemists and geneticists are making progress on this question and are now engaged in an intensive effort to identify the molecular components of the spindle that are essential for its function. Two microtubule-associated motor molecules have been found in dividing cells. They travel along microtubules and bring about the movement of neighboring structures—at least in vitro. It remains to be seen whether these enzymes participate in mitosis in the living cell. At the moment, it is equally possible that enzymes still to be discovered are the mitotic motors we seek. We are hopeful that the identity of these important molecules will be established within the next few years.

FURTHER READING

BEYOND SELF-ASSEMBLY: FROM MICROTUBULES TO MORPHOGENESIS. Marc Kirschner and Tim Mitchison in *Cell*, Vol. 45, No. 3, pages 329-342; May 9, 1986.

CHANCE ENCOUNTERS AND PRECISION IN MITOSIS. R. Bruce Nicklas in *Journal of Cell Science*, Vol. 89, Part 3, pages 283-285; March, 1988.

CELL MOVEMENT, VOLUME 2: KINESIN, DYNEIN, AND MICROTUBULE DYNAMICS. Edited by Fred D. Warner and J. Richard McIntosh. Alan R. Liss, Inc., 1989.

ENZYMES FOR MICROTUBULE-DEPENDENT MOTILITY. J. Richard McIntosh and Mary E. Porter in *Journal of Biological Chemistry*, Vol. 264, No. 11, pages 6001-6004; April 15, 1989.

MITOSIS: MOLECULES AND MECHANISMS. Edited by J. S. Hyams and B. R. Brinkley. Academic Press, Inc., 1989.

DON'T EVEN THINK OF BUYING A TREADMILL... CONSIDER THE OVERWHELMING ADVANTAGES OF NORDICTRACK

Better exercise.

NordicTrack simulates the world's best exercise, cross-county skiing. This smooth, total-body motion exercises all major body muscles including arm and shoulder muscles that are neglected in walking or running.

Non-jarring.

NordicTrack's smooth, efficient skiing motion has none of the jarring motions which can damage joints and ligaments.

Safer.

NordicTrack is human powered and has no electric motors or high-speed belts that require constant concentration or that can throw off a user or child.

Easier speed adjustment.

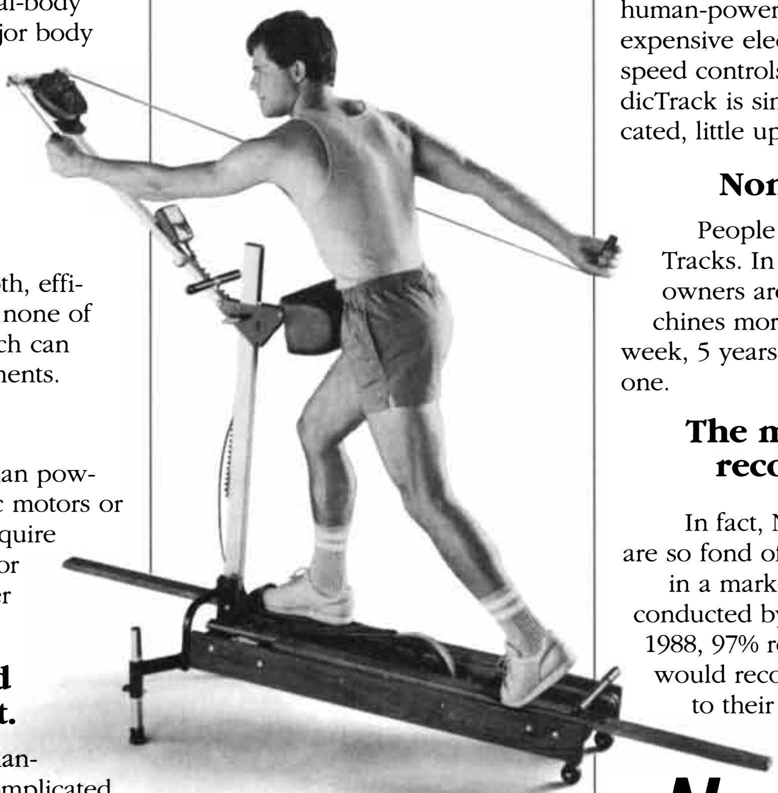
NordicTrack is human-powered and has no complicated and slow-to-react electric speed adjustments to be dealt with for each speed change.

Quieter.

NordicTrack's smooth skiing motion is inherently quiet and has none of the pounding foot step noise associated with running on a treadmill.

No electric cord or outlet required.

Put a NordicTrack wherever you want. You are not constrained by outlet location and you won't trip over the cord.



Costs far less for equal quality.

NordicTrack's efficient and human-powered design needs no expensive electric motors and speed controls. And, because NordicTrack is simple and uncomplicated, little upkeep is required.

Non-boring.

People love their NordicTracks. In fact, 7 out of 10 owners are still using their machines more than 3 times a week, 5 years after purchasing one.

The machine you'll recommend.

In fact, NordicTrack owners are so fond of their machines, that in a market research study conducted by Maritz Research in 1988, 97% responded that they would recommend NordicTrack to their family and friends.

NordicTrack
A CML Company

FREE BROCHURE & VIDEO

Call Toll Free **1-800-328-5888**

In Canada 1-800-433-9582

❑ **YES,** please send me a free brochure

❑ **ALSO** a free video tape ❑ VHS ❑ BETA

Name ______________________

Street ______________________

City ____________ State ________ Zip ________

Phone ( ) ______________

NordicTrack • 141 Jonathan Blvd. N. • Chaska, MN 55318

320J9

The Stanford Linear Collider

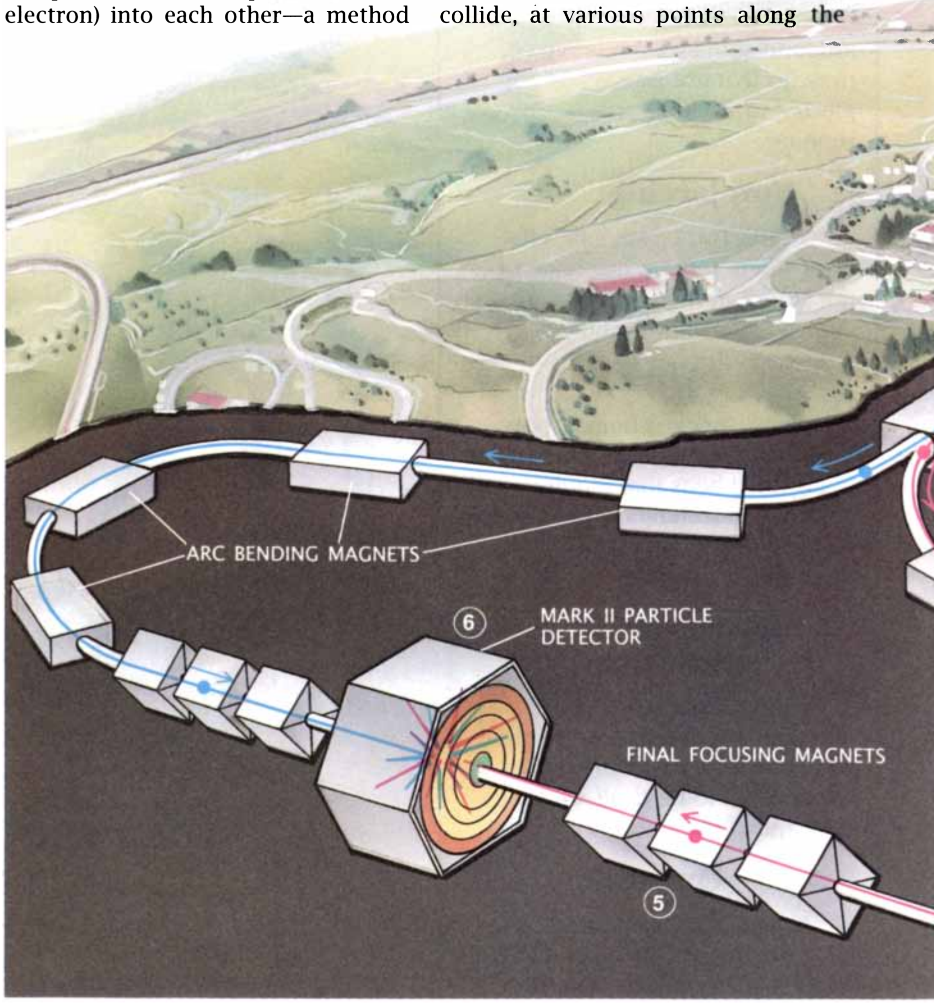
The world's first linear collider is up and running. Stanford's "Z^0 factory" allows physicists to measure the mass and lifetime of the Z^0 mediator of the electroweak force with unprecedented precision

by John R. Rees

Early in the morning of Tuesday, April 11, the sun was starting to burn the fog off the grassy hills around the Stanford Linear Accelerator Center (SLAC) as my colleagues on the night shift started home. Minutes before, unknown to any of us at the time, a surge of energy had suddenly pulsed through the three-story-high, 1,800-ton hulk of iron that forms the shell of the Mark II detector. The event passed in less than an eyeblink. It was not until the following morning that Barrett Milliken, a postdoctoral fellow at the California Institute of Technology, noticed something unusual as he was poring over computer data stored from the previous day. Two spindly jets of particles had sprayed out from the center and struck the detector, depositing some 65 billion electron volts of energy. The brief pulse, Milliken realized, showed the unmistakable features of a Z^0—a particle that "carries" the weak nuclear force, one of the fundamental forces of nature. By noon, the news had flashed around the world: we at SLAC had finally reached the goal that had eluded us for almost a year.

What was remarkable about the April event was not so much the observation of the Z^0 (pronounced "Z zero" or "Z naught"), which had already been discovered six years ago, as the fact that the particle had been created in a machine, the Stanford Linear Collider (SLC), that was of an unprecedented design. The SLC collides high-energy beams of electrons and positrons (the antiparticles of the electron) into each other—a method that has proved exceptionally fruitful in the study of fundamental interactions of matter. Yet ever since the machines were first built in 1960, electron-positron colliders had always consisted of two particle beams coursing in opposite directions along a circular track; the beams intersect, or collide, at various points along the

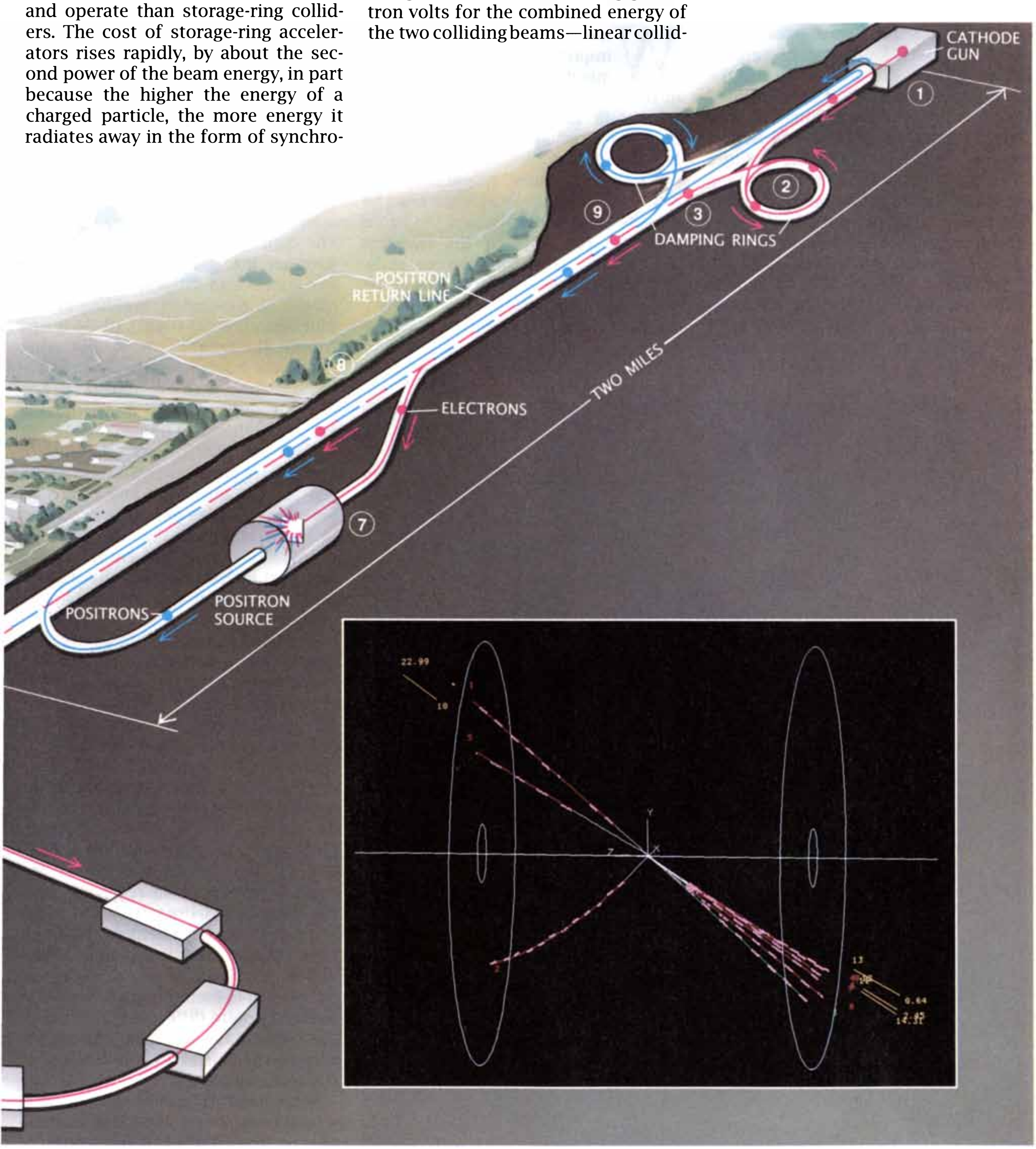
JOHN R. REES might be suspected of less than sterile detachment in describing the Stanford Linear Collider, since he directed the construction of the machine. Rees has been in the colliding-beam business from its inception: his Ph.D. from Indiana University in 1956 coincided with the first practical proposals to build colliding-beam storage rings. Before becoming involved with the SLC, Rees worked with Burton Richter, director of the Stanford Linear Accelerator Center, to design, build and run two colliding-beam storage rings at SLAC: SPEAR, the storage ring with which Richter made his Nobel prize-winning discovery of the psi meson, and the 15-GeV PEP storage ring.



STANFORD LINEAR COLLIDER accelerates electrons (*red*) and positrons (*blue*) in a two-mile-long linear accelerator, or linac, and then steers them into a head-on collision at a combined energy of about 100 gigaelectron volts (GeV). A cathode fires two bunches of electrons in succession (*1*). The bunches are accelerated to 1 GeV and then are condensed in a damping ring (*2*). The damped bunches are injected into the linac, where they are joined by a damped bunch of positrons (*3*). The leading electron loop, producing showers of new particles. In a break with tradition, the SLC aims two linear beams at each other.

Strong motives impelled the Stanford team to choose the route of innovation. For one thing, linear colliders promise to be less expensive to build and operate than storage-ring colliders. The cost of storage-ring accelerators rises rapidly, by about the second power of the beam energy, in part because the higher the energy of a charged particle, the more energy it radiates away in the form of synchrotron radiation when it is forced to follow a curved path. The cost of linear accelerators, in contrast, increases only linearly with the beam energy. At low energies, storage rings are more economical, but at sufficiently high energies—above around 100 gigaelectron volts for the combined energy of the two colliding beams—linear colliders become the less expensive of the two. (One gigaelectron volt, or GeV, is 10^9 electron volts.)

An equally powerful motive was the desire to build a "Z^0 factory," a facility at which the Z^0 particle can be studied

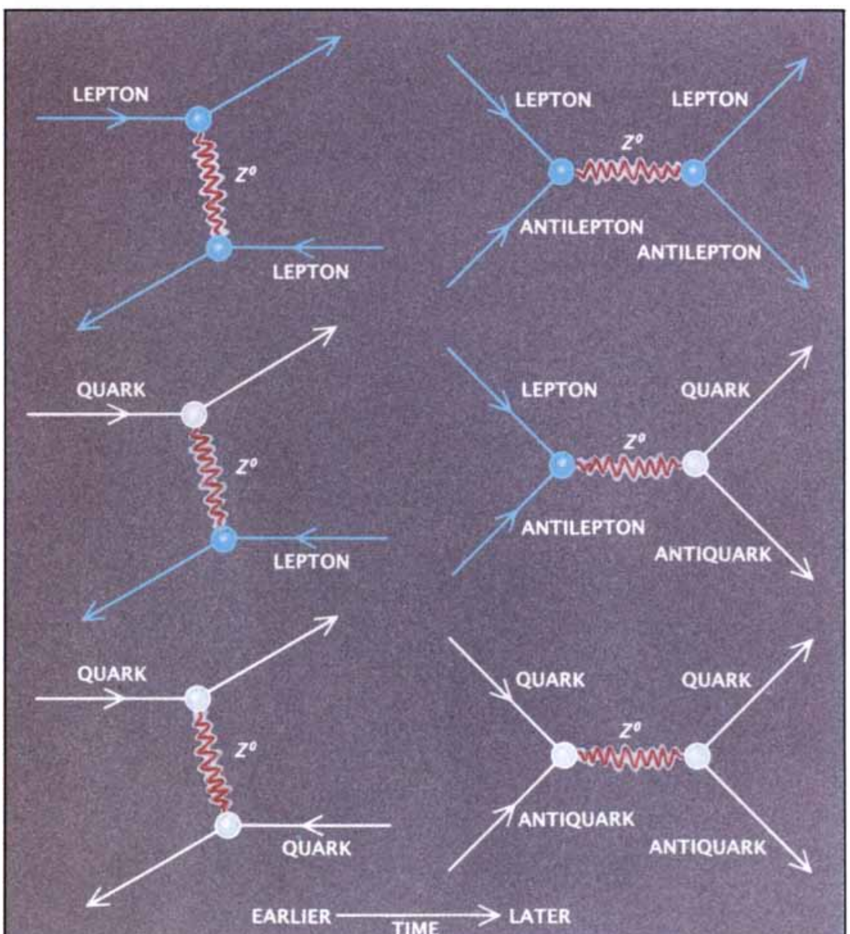


bunch and the positrons are accelerated to the end of the linac, where they are deflected into two large arcs (*4*)—electrons to the left and positrons to the right. Magnets guide the beams to the final focus, where the beams are squeezed down to a diameter of a few microns (*5*). The beams collide inside the Mark II particle detector (*6*). Meanwhile, the trailing electron bunch is diverted into a target to produce a shower of positrons (*7*), which are returned to the head of the linac (*8*), damped and injected into the linac at the same time as a new group of electrons (*9*). The first Z^0 particle seen at the SLC (*inset*) immediately decayed into a quark and antiquark, which produced two hadron jets that struck the end caps (*white disks*).

in detail. This particle was discovered in 1983 at CERN, the European laboratory for particle physics near Geneva, and the study of its basic properties, particularly its mass and lifetime, is one of the chief goals of particle physics today. In 1983 engineers at CERN broke ground for their own Z^0 factory, the Large Electron-Positron (LEP) Collider, a conventional storage ring 27 kilometers in circumference. In the meantime, physicists at SLAC decided to convert the two-mile-long Stanford linear accelerator, or linac—the largest in the world—into a linear collider that could produce Z^0's. By souping up an existing linac, rather than building new linacs from scratch, the SLAC group hoped to have the world's first Z^0 factory—and at relatively low cost.

No linear collider had ever been built, and the urge to be the first to do so was great. With the attraction of innovation came the other side of the coin: risk. We at SLAC chose to venture into uncharted territory where we knew we would face our share of surprise and disappointment, but in the end, the gamble paid off. More than 200 Z^0 particles have been detected at the SLC, and more continue to be churned out regularly.

What is the Z^0, and why is it important? To answer that question, one needs a picture of the current model of physical matter and forces. All of the matter in the universe is constructed out of a handful of building blocks. That is what physicists—or most physicists, anyway—believe. Chunks of matter can be subdivided and subdivided yet again down to the level of the building blocks, but there the process ends. The elementary particles cannot be subdivided further. (That is why they are elementary.) Elementary particles come in only two kinds: quarks and leptons. Quarks stick together in groups of three to form neutrons and protons. Neutrons and protons (which are known as hadrons) stick together in bunches to form atomic nuclei. To a nucleus is joined a swarm of electrons—which are leptons—to give rise to an atom. (Other leptons include muons and neutrinos.) The process goes on: atoms join together to form molecules, and molecules make up gases, liquids and solids.



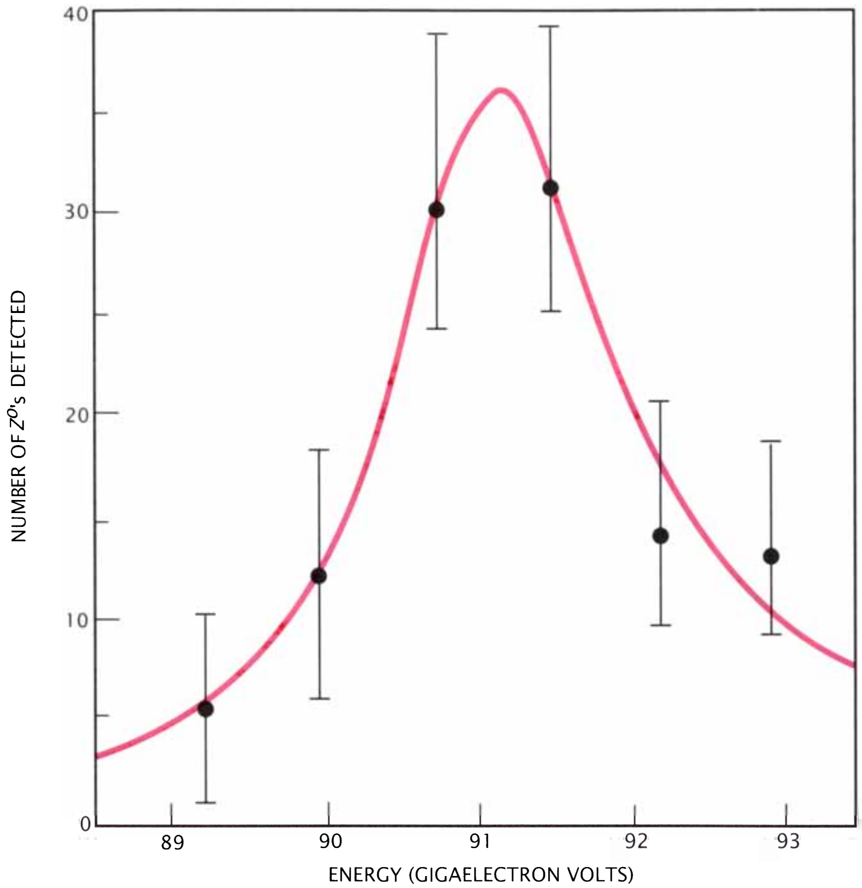
SIX WEAK INTERACTIONS involving the Z^0 mediator are depicted in these Feynman diagrams. Distance is represented on one axis and time on the other. Theory predicts that if the processes in the left column can take place, so too can those in the right column. Collisions between a lepton and antilepton (the two upper processes in the right column) can produce either a lepton and antilepton or a quark and antiquark. The SLC is designed specifically to enable both kinds of interactions to be studied.

There are remarkably few elementary particles: only six quarks and six leptons, according to current theories. (Actually, current theories also call for another particle, the Higgs boson, but one can explain elementary particle interactions without referring to it.) The quarks and leptons are grouped together into three "generations" of two quarks and two leptons each. This makes 12 elementary building blocks (not counting the Higgs), or 24 if one counts their antiparticles. But what makes them stick together? Bricks are of little use without mortar. It has been as difficult to find a satisfying answer to that question as it has been to find out what the complete set of elementary particles is. Since the 19th century the attraction and repulsion between particles has been described in terms of forces that the particles exert on one another. A complete theory must include not only the elementary particles but also the mechanism of their interactions.

The theory of special relativity stipulates that a physical effect cannot move faster than the speed of light. Physicists therefore developed quantum field theory, in which the force between two particles arises from the exchange of a "mediator" that "carries the force" at a finite speed: one of the particles emits the mediator; the other absorbs it. The mediator propagates through space and, for a brief time, is not lodged with either the sending particle or the receiving particle. These mediators have the same properties the elementary particles have—mass, electric charge, spin—and so physicists often call them particles as well, even though their role in nature is quite different from that of the elementary particles. The mediators are the mortar that binds together the building blocks.

A complete theory of matter, then, would consist of a list of the elementary particles and a full description of all the ways the particles can interact through the mediators. Such a theory would explain the occurrence and behavior of matter—at least in principle. The laws would be known exactly, although in many real-world situations it might not be possible to solve the mathematics or to carry out all the calculations that would be required to understand the behavior of the system. (This may seem odd, but it is often the case in physical sciences that the laws are known but computational capacities are not up to the task of finding solutions for real-world problems. For example, although the thermodynamic equations that govern the earth's atmosphere are known, there is not enough computing power to predict the weather accurately.)

Physicists believe they are on the verge of having a complete theory of matter. The theory accounts very well for the forces that hold together nuclei, atoms and molecules. It is so successful and so widely accepted that it is simply called the Standard Model. The only force the model fails to encompass is gravity, and theorists are working hard to rectify that omission. Meanwhile, the business of particle physicists is to verify the existing theory as far as it goes. Among other tasks, they must measure the properties of each elementary particle and mediator. It is in measuring the properties of the Z^0 mediator that the SLC has a seminal contribution to make.



Z^0 RESONANCE measured at the SLC indicates that the particle's mass is 91.2 GeV. The curve was provided by Jonathan Dorfan of the Mark II detector group at SLAC.

Three kinds of mediators are now known to exist: photons, gluons and weak-force mediators. They correspond to three of the four forces in nature. Photons mediate the electromagnetic force; gluons mediate the strong nuclear force, which binds quarks together; weak-force mediators mediate the weak nuclear force, which underlies radioactive decay. There may be a fourth kind of mediator, the graviton, for the gravitational force, but it has not yet been observed.

Photons and gluons have been the subject of experiments for years, and a good deal is now known about them, but the weak-force mediators are still terra incognita. Their existence was verified only in 1983 at CERN. There are three weak-force mediators: W^+, W^- and Z^0. None of them is understood in detail because they are hard to study with the accelerators that preceded the SLC.

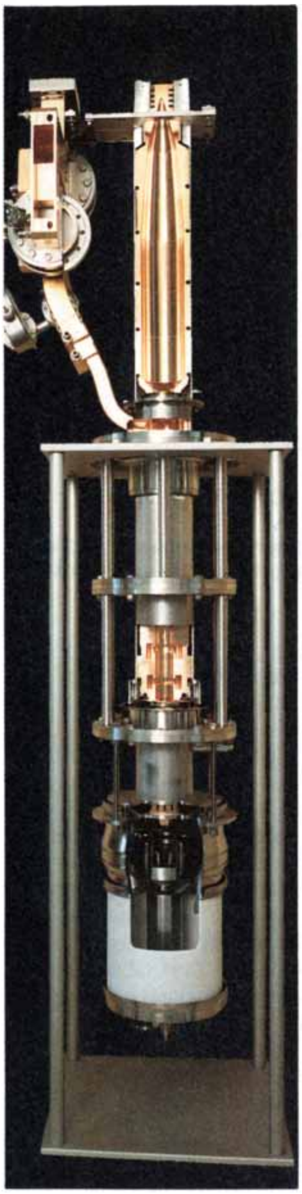
To understand why that is so, one must know how the mediators are produced in the first place. If an elementary particle and its antiparticle—for example, an electron and a positron—pass close to each other, or in essence collide, one of them can emit a Z^0, and the other can absorb it. The Standard Model says something else can happen as well. If the particles are sufficiently energetic, one of them, say the electron, can emit a Z^0, and then the positron can absorb the electron: the particles annihilate each other, leaving the Z^0 momentarily free. Afterward, the Z^0 must decay back into a pair of elementary particles, such as an electron and positron or a quark and antiquark.

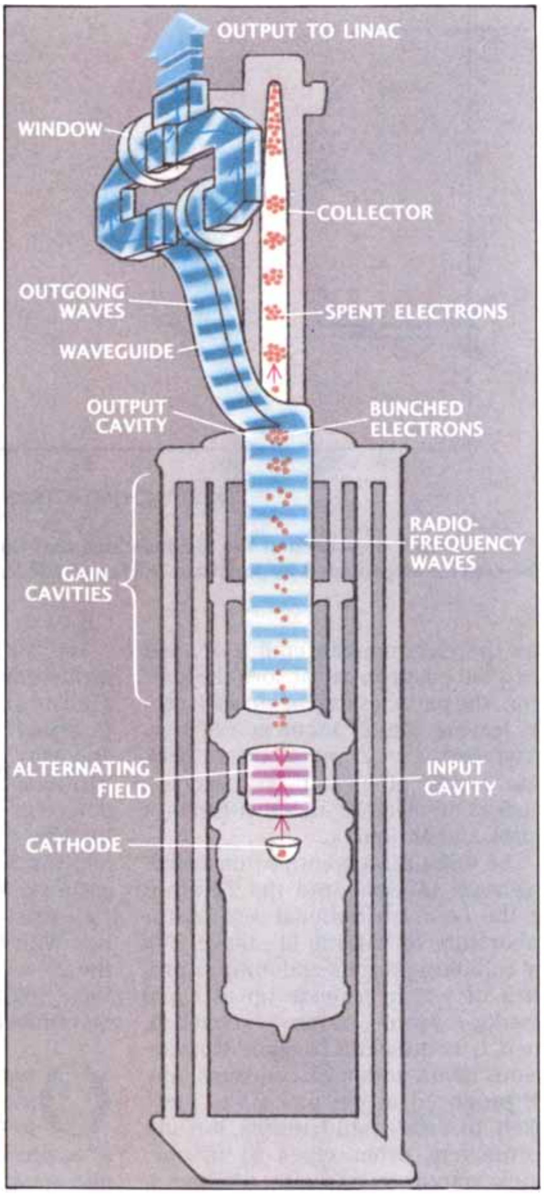
The CERN proton-antiproton colliding-beam machine and the Tevatron at the Fermilab National Accelerator Laboratory in Batavia, Ill., make Z^0's by colliding protons and antiprotons, each of which is made up of three quarks. A proton-antiproton collision, then, is tantamount to many simultaneous quark-antiquark collisions. Any Z^0 produced in the process is most likely to decay into hadrons, but unfortunately, other types of interactions among quarks also produce a flood of hadrons. As a result, these machines not only produce Z^0's but also initiate many other reactions that mimic the decay modes of Z^0's and thereby mask the presence of genuine Z^0 events. Consequently, of the Z^0's produced at CERN and Fermilab, only a tiny fraction—those that turn into leptons—can be identified.

The SLC, in contrast, collides two elementary particles: an electron and a positron. It is designed specifically to produce copious numbers of Z^0's and very few extraneous events, so that almost every Z^0 produced can be detected—in all of its decay modes. Particle physicists need to study a large number of Z^0's in order to get an accurate measurement of the particle's mass, lifetime and other properties. With sources like the SLC and LEP, the Z^0 will be the first of the weak-force mediators to have its properties determined comprehensively.

One of the primary goals of the SLC experimental program is to determine the mass of the Z^0 as precisely as possible. The mediators are among the basic components of the Standard Model, and so their masses are fundamental constants—that is, the masses cannot be calculated from the model but instead must be measured directly. How is the mass determined? As it turns out, the process of creating the mediator is "resonant": the closer the total energy of the two colliding particles is to the rest mass of the Z^0, the higher the probability of creating a Z^0. Experimentalists collide particles at various energies around the expected Z^0 rest-mass energy and plot the number of observed Z^0 particles. The result looks like a hump, with the peak occurring at 91.2 GeV, the actual rest mass of the Z^0 measured at the SLC [*see illustration on preceding page*].

The shape of the Z^0 resonance curve is important for another reason: the width of the hump is a measure of how many different families of elementary particles exist. Why is that? As I explained previously, the electron-positron annihilation leaves a naked Z^0, which then decays after the briefest of lives—about 10^{-25} second. If one imagines that the Z^0 is a bucket full of water, then the more holes are punched in the bucket, the faster the bucket will empty. Similarly, the more different ways there are for the Z^0 to decay, the sooner it will decay. Thus, by measuring the Z^0's lifetime, one knows how many families of elementary particles exist. Now, I originally referred to the width of the resonance curve; how is that related to the lifetime? Here one must appeal to the uncertainty principle of quantum mechanics. The principle states that the width of the resonance—the uncertainty in the Z^0's energy—is reciprocally related to the lifetime of the decaying particle.



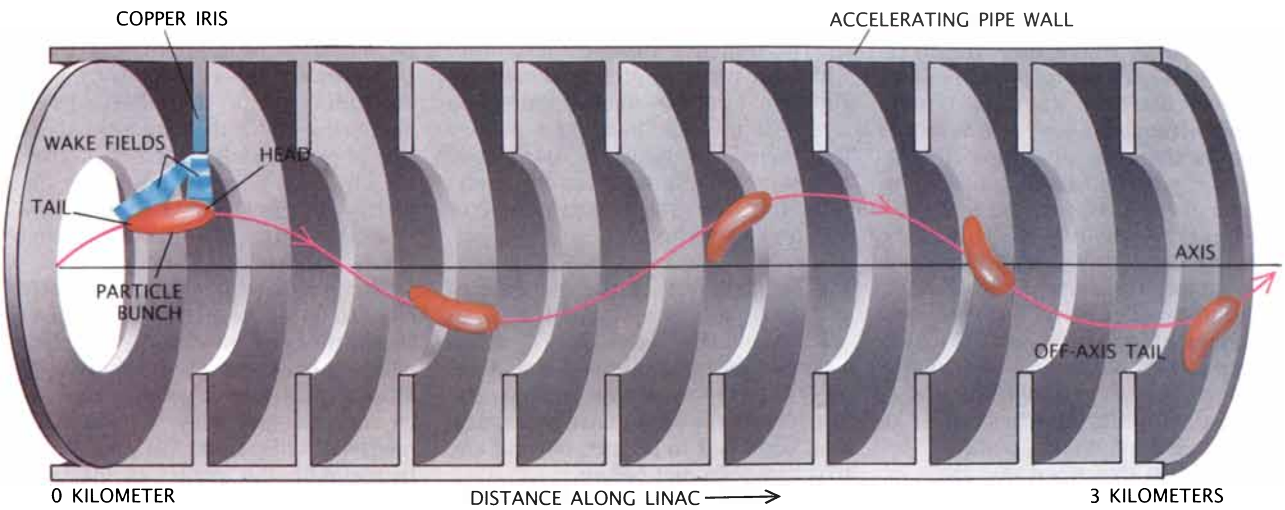


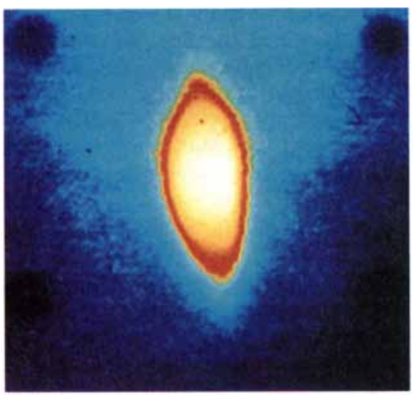
67-MEGAWATT KLYSTRON, the most powerful in the world, supplies the power to accelerate the particles in the SLC. Electrons are ejected from the cathode, focused into a beam and injected into the input cavity. An alternating electric field in the cavity wall causes some of the electrons to speed up and others to slow down. Fast electrons catch up with slow ones and form a dense bunch just as they strike the output cavity. The waves of bunched electrons generate radio-frequency waves, which are diverted into a waveguide. Spent electrons are dumped into the collector. The radio-frequency power is split into two channels and passed through alumina ceramic windows to reach a waveguide that directs the power into the linac.

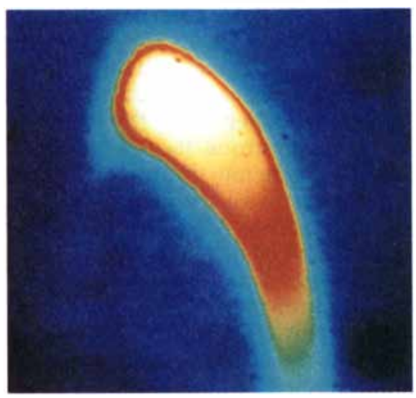
What does the width of the resonance curve imply for the Standard Model? The model calls for six quarks and six leptons. The Z^0 cannot decay into just any arbitrary pair of particles, however. The combined mass of the resulting particles cannot be greater than the mass of the Z^0; otherwise, the process would violate the conservation of energy. It is already known that 11 of the 12 elementary particles have masses low enough to be produced in pairs by the Z^0. On that premise, the theory predicts that the resonance curve will be 2.5 GeV wide. If the width were greater, it would imply that there are other types of quarks and leptons, beyond those currently encompassed by the Standard Model, forming a fourth generation of elementary particles. Some physicists think the neutrinos affiliated with a fourth generation could account for the "missing mass" in the universe [see "Particle Accelerators Test Cosmological Theory," by David N. Schramm and Gary Steigman; SCIENTIFIC AMERICAN, June, 1988]. As this article goes to press, this is still an open question.

The SLC, then, offers rich experimental opportunities. It provides the first chance to study every decay path of the Z^0. This is something that cannot be done with a proton-antiproton collider because of the heavy background signals. The SLC will also enable the first complete tally to be made of all of the elementary particles that have masses less than half of that of the Z^0. And the SLC will also furnish a plentiful source of the Z^0 decay products, thereby providing an excellent laboratory in which to study those particles themselves. An experimental program to exploit all these features promises to keep the SLC busy for years to come.

The physics that will be done in conjunction with the SLC is certainly important: it is the same prize physicists in Europe will be going after with the LEP Collider. But there is another significant story in the SLC: as the first linear collider built it represents a triumph of accelerator







WAKE FIELD EFFECT arises because an electron bunch induces electrostatic fields in the walls of the accelerator pipe, which in turn can perturb the electrons in the wake of the beam (*top*). If the bunch is traveling precisely on the centerline, the wake field effect will cancel itself out through symmetry, but if the beam is off center, the effect will throw the tail out of alignment. Television screens show transverse sections of a well-focused beam (*bottom left*) and of a beam distorted by the effect (*bottom right*). The beam center appears white; the edges appear in color. The beam is about 250 microns in diameter.

physics—the physics of building machines that can attain the extraordinary conditions in which exotic particles can be forged.

The simplest kind of linear collider would consist of two linacs aiming beams of particles at each other—a feat likened to colliding two bullets by firing two rifles at each other over a great distance. In many ways, a linear collider is much more difficult to build than a storage-ring accelerator of equivalent performance. In a storage ring, particles whirl around the ring repeatedly and have about 10,000 chances a second to collide with particles moving in the opposite direction. Particles in a linear collider, on the other hand, have only one chance each time the machine fires. The SLC can pulse about 100 times a second. To make up for the lower frequency of beam collisions, the bunches of particles must be made 100 times more dense than the beams in a storage-ring collider. To do this, we had to concentrate about 50 billion particles into each bunch and focus it down to a diameter of only a few micrometers. I shall say more about this later.

At the SLC this difficulty was compounded by the fact that we had piggybacked the new design onto a 20-year-old machine. That original linac had been designed to perform at less stringent specifications than those required for a linear collider. What is more, using the linac meant we would begin by accelerating both the electrons and the positrons in the same direction and then depend on a tricky maneuver to bring the beams into collision. What we did was to split the beams with a simple dipole magnet at the far end of the linac, thereby turning the electrons to the left and the positrons to the right. The split beams were then guided around two large arcs that come into contact at the tips, like a giant pair of tongs.

The electron beams are simple to make: we pulse a large cathode for a brief instant. A short stream of electrons issues forth and enters the upstream end of the linac where, through the action of accelerating electromagnetic fields, the electrons are quickly gathered into short, intense bunches. The positron beam is harder to produce: a bunch of high-energy electrons is fired into a target, and positrons are winnowed from the resulting shower of particles.

Although both the electron and positron bunches have the requisite number of particles, they are too diffuse. If we let these bunches collide, they would not produce many interactions. It would be as though we were firing two shotguns at each other instead of rifles and the shot had scattered over a broad area. To make the bunches sufficiently dense, we need to slim down their transverse dimensions so that the particles within them will be much closer together. This condensing process is done inside two small storage rings, or "damping" rings. The particle bunches are accelerated to about 1 GeV and then injected into the rings, where the transverse width of the bunches is shrunk to a sufficiently narrow dimension. The process is similar to cooling a gas: as the particles throw off synchrotron radiation in a damping ring, the average distance between them decreases.

The electrons and positrons remain in the damping rings for a fraction of a second. The bunches—still at an energy of 1 GeV—are then injected back into the linac, where they are accelerated to higher energies as they are propelled down its length by powerful radio-frequency waves; each bunch surfs on the crest of a wave. By the time the particle bunches reach the far end of the linac, they are at the desired energy—about 50 GeV—and are diverted into the two large arcs and brought into collision. The final leg of the journey takes each bunch through a system of magnetic lenses, which squeezes the bunches down to the smallest transverse dimension possible, just before the bunches collide at the so-called interaction point.

I have just described one cycle of the SLC's operation. In practice the accelerator repeats this cycle many times a second. In order to make the single linac serve multiple purposes, we resort to some fancy choreography. In each cycle, the cathode is fired twice in rapid succession, thereby producing two bunches of electrons. The bunches proceed down the linac and are joined by a bunch of positrons (the reader will shortly see where the positrons came from), and the three bunches, spaced out along a distance of about 20 meters, enter the damping rings for a fraction of a second. The bunches are then extracted and fed one after the other into the linac, with the positrons in the lead. About two thirds of the way down, a fast-firing pulsed magnet kicks the trailing electron bunch out of the linac. That bunch is fired into a target to produce positrons. The remaining electron bunch continues down the linac, together with the positron bunch, and the two bunches finally collide. Meanwhile, the newly minted positrons are conducted back to the beginning and injected into the linac just as two fresh electron bunches start their journey. Now the machine has come full circle. The cycle is repeated 60 or 120 times a second.

The builders of the SLC had to face many new challenges in accelerator physics. To attain acceleration to high energies in a linac, and at the same time achieve stable operation with a dense beam, new standards had to be met, beyond any that had been met before.

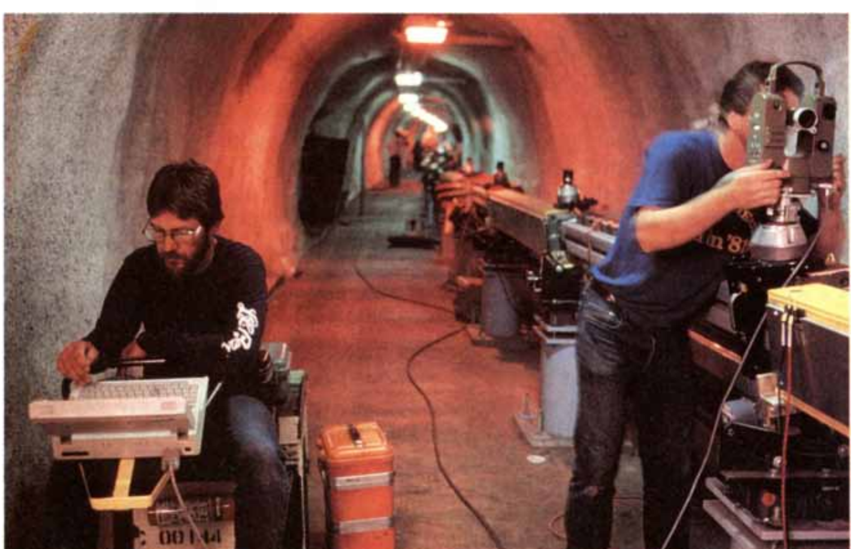
First, the SLAC linac had to be pushed to raise its top energy from 30 to 50 GeV, which required higher accelerating electric fields than any attained previously in a linac. The two-mile-long vacuum pipe in which the beam is accelerated was easily capable of sustaining the higher fields. The problem lay in the klystrons, the six-foot-high tubes that generate the radio-frequency power to drive the fields. Even before the SLC was built, the SLAC klystrons produced the highest peak power—35 megawatts—of any klystron in regular production in the world, but it was not powerful enough. SLAC physicists had to design a new klystron that could produce 67 megawatts.

Klystrons are notoriously difficult to make, and there were a large number of failures. We needed hundreds of klystrons, and yet at first only a dismal 30 percent of the ones that were made actually worked. Another serious problem was that the alumina ceramic windows through which the generated power escapes from the klystron waveguide into the linac kept shattering under the explosive liberation of power. Early efforts to improve the klystrons were disappointing, and many observers feared that the SLC would never reach beam energies high enough to produce Z^0 particles.

One by one we overcame the obstacles. The window problem was solved by splitting the power into two channels and releasing it through two separate windows: each window now had to contend with only half as many megawatts of power. Arduous efforts to maintain the quality of materials and workmanship enabled the SLAC team to boost the manufacturing yield to 85 percent. By the time the SLC entered its first trials, it was equipped with enough of the new klystrons to reach the energies needed to produce Z^0's. The new klystrons proved to be extraordinarily serviceable, and by the beginning of this year, SLAC had built about 500 of them and had an ample supply of backups.

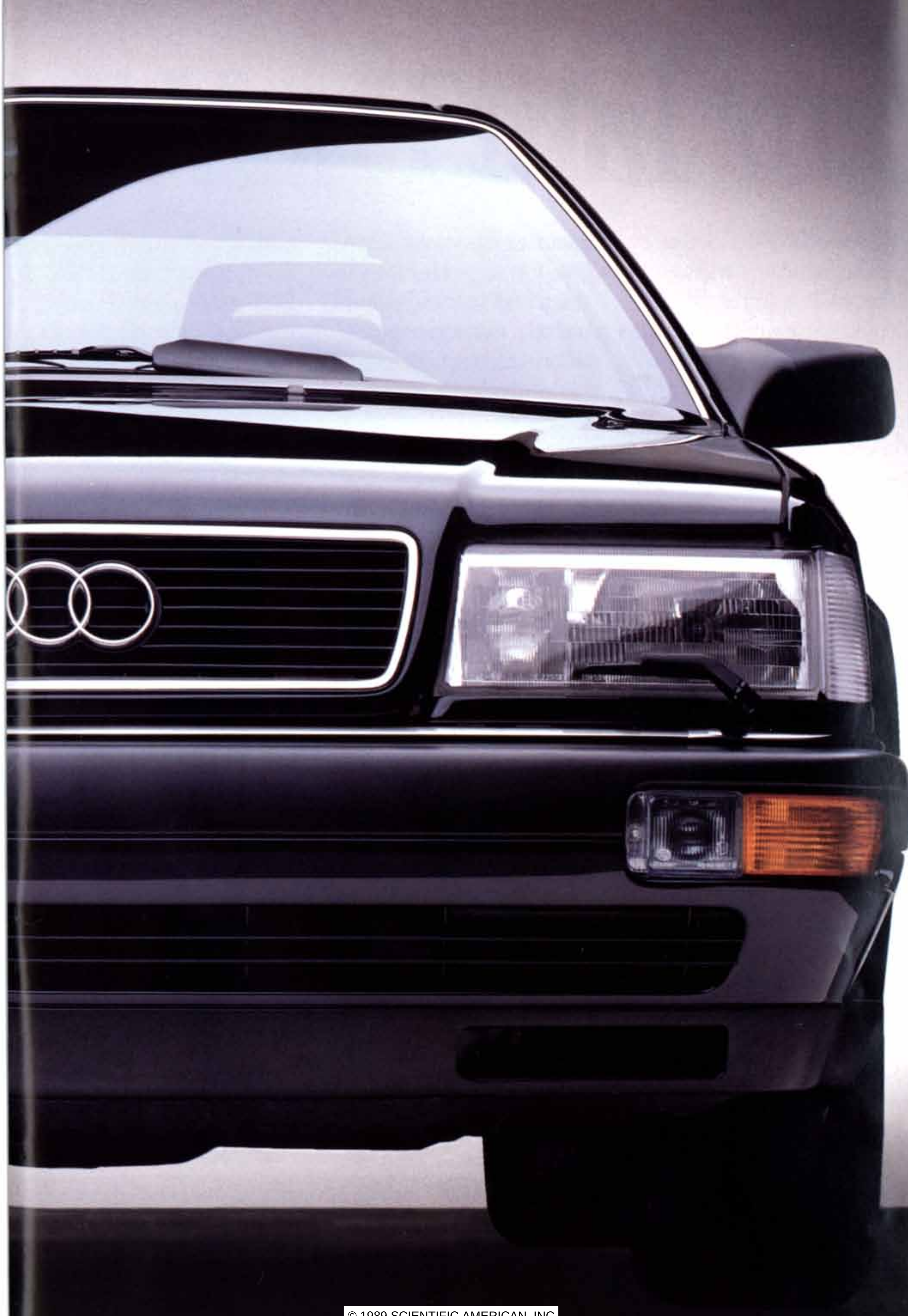
A second major challenge was to create beams of electrons and positrons of unprecedented density. At the interaction point, the beam's cross-sectional diameter is only a few microns, and the particle density is almost as great as that of molecules in a gas at room temperature and pressure. That may sound commonplace—and indeed, for ordinary gas molecules, it is—but for electrons and positrons it is extraordinary. Beams of such charged particles are usually a thousand times less dense, because they are produced in a hot state; the electrons are boiled off a cathode, and the positrons are created in a high-energy interaction. As a result, they careen back and forth across their intended path in a state of great disarray.

To compress the disorderly beams, the designers of the SLC proposed a new kind of device, the damping rings I mentioned earlier. As the particles radiate away energy in the damping rings, their back-and-forth motions are reduced, or damped. Designing these rings was a new art. The designers had particular difficulty devising the "kicker magnets," which apply a sharp magnetic pulse to eject particles out of the rings. In order to reach the peak cycle rate of 120 cycles a second, there must be two bunches in the damping rings at any given time. Since each bunch makes a complete circuit in only 800 nanoseconds (800 times 10^{-9} second), the challenge is to create pulses sharp enough to kick one



MAGNETS in the SLC arc are aligned to within 100 microns to ensure precise control of the beams as they are collided. An engineer measures the magnet positions with a surveying instrument while a colleague records the data on a laptop computer.


Introducing the Audi V8 Quattro.



Take The Wor

Maybe you've decided to reward yourself with an ultra-luxury sedan. (Who would say you're not entitled?) But that doesn't mean you've lost your enthusiasm for revolutionary ideas.

Like the notion that a car of this class should reflect your intelligence, not just your wealth.

Or that power is only a virtue when it's properly distributed.

Or that driving a luxury car should be more than a spectator sport.

Audi agrees. And is taking such revolutionary beliefs to the streets. In the form of the new V8 Quattro.

A car that translates Audi's 79 years of unconventional automotive thinking into unequalled technology. And uncompromised performance.

Because its stunningly advanced V8 engine and automatic all-wheel Quattro® drive form a potent combination of power and control. And that exclusive combination yields a most exclusive kind of exhilaration.

Pulling away, you immediately feel the power. Thanks to an astonishingly advanced V8. (So sophisticated, two competing design groups were involved in creating it.) The *leichtmetall* block weighs a mere 62 pounds—yet it reaches a top track speed of 146 MPH.

That familiar stretch of coastal highway still feels fresh. Because the four-speed automatic transmission offers a choice of three Quattro-linked driving programs: "Sport" for when you want to attack the road; "Efficiency" for go-with-the-flow driving; and "Manual" for the option of hands-on, clutch-free gearing down a steep mountain pass.

An exhilaration that must be measured over time and over distance to be evaluated fairly. So, for the next few minutes, take the metaphorical wheel. And take a drive into the future of the ultra-luxury sedan.

From the moment you pull away, you sense the authority of this innovative powerplant. Constructed of *leichtmetall,* it is the world's most compact V8. Yet its 3.6 liters produce 240 HP. And its 32 valves explain the engine's willingness to rev freely. The power output, the throttle response, the smoothness all suggest a nearly ideal design configuration is at work.

You feel the surge, but hear only silence. As the venerable *Financial Times* of London (which knows the difference between noise and news) discovered: "It is quiet. At 120 mph on the

d's First Fully Autc

autobahn, the loudest noises to be heard inside this large long-legged car were some wind roar around the mirrors and whine from the tyres..."

So, you're propelled past the wide suburban lawns quietly. And quickly. You go from 0 to 60 in 8.7 seconds. Swift, by any standards. (Especially on the German fast track where it competes; in fact, at top speed, you're likely to see cars like the BMW 735i and the Mercedes 420 SEL in the rearview mirror.)

Soon, the winding coastal highway becomes your plaything, as you set the Electronic Driving Program to "sport" mode. Result? The automatic four-speed ZF® gearbox recalibrates instantly; the performance is brisk and bracing.

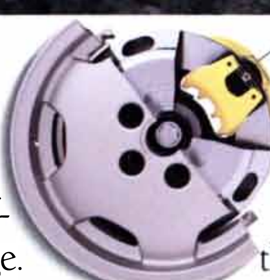
But power alone is not sufficient challenge. After all, August Horch (one of Audi's founders) developed Germany's first eight-cylinder engine back in 1926. To Audi, the issue is *usable power.*

So your new V8 distributes its power in a revolutionary way. Thanks to the latest evolution of all-wheel Quattro drive.

How? Three separate differentials work in concert. Traction is continually monitored. Wheel slip is detected within 1/50th of a second. And the system compensates accordingly. You enjoy special confidence, because it widens your margin of performance.

Consider that stretch of snow-covered, half-plowed road up ahead. In

Conditions look a little dicey? Don't worry. The first automatic all-wheel Quattro drive extracts maximum traction out of every road surface. Comprised of three separate differentials, it detects wheel slip instantly. And reassigns power accordingly.



Another Audi first: internal brake calipers. Front discs deploy calipers at the *inner* radius – for greater braking surface, better heat dissipation, less brake fade.

omatic Quattro Dri

the blink of an eye, the Quattro drive can shift 100% of the power from front wheels to rear—and back again. If necessary, its rear differential can even transfer as much as 80% of that considerable power from side to side. Just as startling, it seamlessly choreographs such surefootedness with the security of anti-lock braking. And on a road like this, that might make all the difference.

Of course, Quattro is also a fair weather friend—with similar advantages on dry pavement. Need evidence? How about the string of sunny-day upsets it's been compiling on the SCCA Trans-Am and IMSA-GTO circuits? Or its effortless way with a curve? Or its laser-like precision through the corners?

Even after the sun goes down, an Audi V8 still shines. Halogen headlamps use dual paraboloid reflectors for an intensified beam; lenses are specially faceted for a more focused "throw." Inside, backlit instrumentation soothes the eyes.

As you're beginning to discover, in the Audi V8 Quattro, *you* determine where the road will take you.

Unfortunately, in this case, it's taking you right into stop-and-go traffic. Fortunately, you have the assurance of another Audi first: *internal* brake calipers. By turning traditional brake design on its head, the Audi V8 provides a 17% larger brake surface. Which means less heat buildup, less brake fade and more peace of mind.

In fact, you take a moment to consider your good fortune. Surrounded by burled walnut trim. Ensconced in heated leather sport seats. Refreshed by an Electronic

Your favorite switchback and your favorite tape. That's entertainment. The Audi/Bose® Premium Music System is so acoustically precise, even the upholstery material is a factor in its final equalization. An integrated microphone in the driver-side pillar permits "hands-free" operation of the standard cellular telephone.

e For A Drive.

nate Control capable of ewing the air every 27 onds. Serenaded by an di/Bose® Music System. d protected by a driver- e Supplemental Restraint tem. You sit in sensible ilence. The cellular phone

Down the road, the Audi Advantage provides unequalled coverage, including three years or 50,000 miles of scheduled maintenance at no charge.

ows for "hands-free" eration. The on-board to-Check monitors 13 al operating conditions. e seats offer 10-way ustment.

You drive on. Alert, are and at ease. Even as y turns to night. And pan freeways turn to coun- lanes and mountain itchbacks.

Though the road twists and turns like a Slinky,® your V8 Quattro is undaunted. The newly improved suspension (with its telescoping gas-charged shocks) shrugs off the road's worst. Truth is, here is where the Audi V8 excels: turning raw power into a refined asset and stepping lightly where lesser cars might trip. In short, as *Car and Driver* said, it "provides what can only be called a superior driving experience."

As the miles click by uneventfully, you take solace in this Audi's farsighted design. Shaped from 100% galvanized steel, its body offers a full decade of rust protection.* Equipped with self-diagnosing computers, its engine stores service data for easy maintenance. And protected by the Audi Advantage, its owner enjoys years of worry-free driving.

But while the rational reasons for driving the V8 Quattro are legion, the true reward is ultimately emotional. You simply *feel it.*

And, as you pull into your drive, you can be forgiven your small measure of smug satisfaction.

After all, at this exact moment, in this exact car, life is good.

Audi

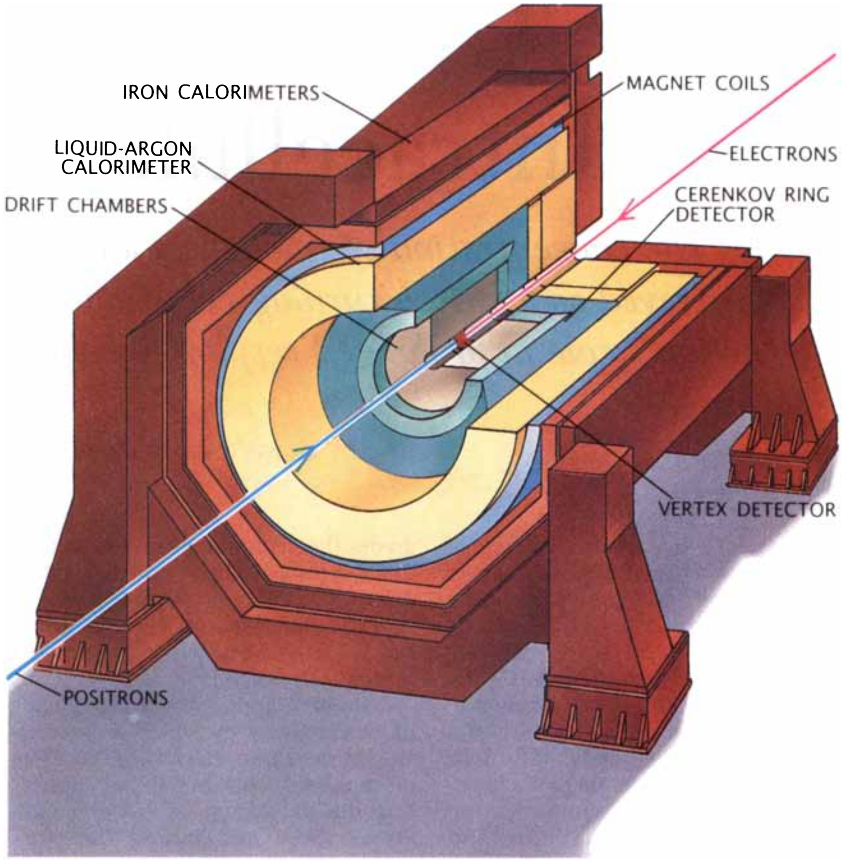
The Alternate Route.


The new Audi V8 Quattro.
It is a sanctuary for enthusiasts.
A costly combination of rare wood, rich leather and raw excitement.
But, like most things of true value, well worth it.
For a closer look, call 1-800-FOR-AUDI.

bunch out without interfering with the other bunch. To succeed, the pulse must be turned on and off in about 50 nanoseconds—the time it takes for light to fly a distance of 50 feet. The SLC engineers are still working to bring the kicker magnets up to that goal. Meanwhile, the machine is operating at 60 cycles a second.

The third challenge was to keep the beams compressed as they fly down the linac and bend through the arcs. The linac environment tends to disrupt the bunches. Particles at the front of a bunch induce electromagnetic fields called wake fields in the walls of the accelerating pipe, and these fields in turn buffet the back end of the bunch, which causes it to spread laterally. It is as though the bunch were whipping its own tail. Wake fields become seriously disruptive if the bunches are not exactly on the centerline of the pipe. Because the pipe is not geometrically perfect, the only practical way to control the wake field effect is to learn by trial and error to steer the beam in such a way that the effect is as small as possible. The task was complicated by the fact that in order to save on the cost of drilling tunnels through the Stanford hills the arcs were made to hug the terrain.

Finally, all of the SLC's parts had to operate with extraordinary stability and precision. Computers control thousands of magnets and power supplies, some of them to within one part in 10,000. Maintaining such high tolerances simultaneously for all of the parts was a formidable challenge. If adequate performance standards could not be met, the beams would not collide. After all, the beams at the point of collision are about the diameter of a spider's thread. They must be well aimed.



STANFORD LARGE DETECTOR shown in this cutaway view will be installed in the SLC in 1990. The 4,000-ton detector will surround the collision site in all directions, which will enable workers to characterize every particle track. The vertex detector will allow the original position of each particle track to be determined with high accuracy. Other detectors are layered like tree rings around the final focus.

The various systems in the collider were all in place by the spring of 1987. In the first trials, the job of forming the fine bunches and guiding them to the collision point proved to be taxing—but that was to be expected, because such precise navigation of a particle beam along a beam-transport channel had not been attempted before. Other complications in running the SLC came somewhat as a surprise. The computer control system proved more difficult to tame than we had expected. And a severe heat wave in the summer of 1988 led to power brownouts and thermal prostration of the equipment, which kept us on the ground for weeks.

The effort to make the SLC work continued for about two years, and more often than not the thousand-strong SLAC staff worked in shifts around the clock. Finally, shortly after sunrise on April 11 of this year, the massive Mark II detector found the first Z^0. The mediator revealed itself by decaying into a quark and antiquark that were transformed almost immediately into two showers of hadrons. That event marked the first time physicists had seen a Z^0 do the thing it is most likely to do—that is, turn into hadrons.

As the SLC began producing Z^0's, the LEP Collider, which provided few new technical challenges, was racing to completion. The different technologies for the SLC and LEP imply that the two machines will have different future capabilities. The LEP can be pushed to higher intensities and energies, which will enable physicists to search in higher-energy realms for such particles as the Higgs boson. The SLC, on the other hand, can produce beams of particles that all have the same spin and should enable experimenters to gain more useful information from certain experiments. Yet although the SLC was completed earlier, the CERN machine promises in the long run to be the more prolific source of the Z^0. In the end, the SLC's greatest significance will be in having proved a new accelerator technology, one that many physicists think is essential if particle physics experiments are to proceed into higher-energy realms than any attainable by the largest and most costly electron-positron storage-ring colliders now envisioned.

FURTHER READING

PARTICLE STORAGE RINGS. Gerard K. O'Neill in *Scientific American*, Vol. 215, No. 5, pages 107-116; November, 1966.

THE HIGGS BOSON. Martinus J. G. Veltman in *Scientific American*, Vol. 255, No. 5, pages 76-84; November, 1986.

INITIAL MEASUREMENTS OF THE Z BOSON RESONANCE. G. Abrams et al. in *Physical Review Letters*, Vol. 63, No. 7, pages 724-727; August 14, 1989.

The Cycling of Calcium as an Intracellular Messenger

The concentration of calcium in a cell has generally been portrayed as a switch turning cellular processes on and off. But the ion's role in prolonged responses belies the traditional model

by Howard Rasmussen

Two of the more remarkable events in the course of evolution were the development of the exoskeleton of mollusks and, hundreds of millions of years later, the bony endoskeleton of higher animals. Each development represented a new biological use for calcium. Calcium salts in the form of shell, bone and tooth are familiar materials of bioarchitecture; they are visible signs of the importance of calcium in the growth and function of organisms.

There is less general awareness, however, of an older and more pervasive role of the calcium ion: within a wide variety of animal cells, calcium serves as an almost universal ionic messenger, conveying signals received at the cell surface to the inside of the cell. The calcium ion is involved in such diverse processes as the regulation of muscle contraction, the secretion of hormones, digestive enzymes and neurotransmitters, the transport of salt and water across the intestinal lining and the control of glycogen metabolism in the liver.

Whereas the calcification that produces bone involves the ordered deposition of large amounts of salts, the intracellular messenger function involves minute flows of calcium ions across the membranes of cells. In fact, calcium ions can carry out their informational role only at very low and tightly controlled concentrations, because higher ones are detrimental to normal cell function.

Cells have a simple but elegant set of mechanisms by which they regulate intracellular calcium levels. The mechanisms work mainly by controlling the movement of calcium ions across three membranes: the plasma membrane, which surrounds the cell; the inner membrane of the mitochondrion, a cell's energy-producing organelle; and the membranes of compartments that contain reserves of calcium ions, which are called the sarcoplasmic reticulum in muscle cells and calcisomes in nonmuscle cells. Although the concentration of calcium inside a cell remains fairly constant, the flow across the plasma membrane (the amount of influx and efflux) can vary significantly.

Recently it has become clear that such calcium-ion "cycling" across the plasma membrane is part of a complex chain of events by which cells generate sustained responses to stimuli in their environment. Calcium's role in sustained cellular responses, such as the secretion of insulin or the contraction of the smooth muscle surrounding the blood vessels, has historically been more elusive than its part in transient responses, such as the contraction of skeletal muscle. My colleagues and I have been able to piece together a picture of how cycling across the plasma membrane mediates sustained cellular responses. We have already found similar mechanisms operating in three disparate cell systems. Our findings have led us to a more sophisticated understanding of calcium as an intracellular messenger than we had even five years ago.

HOWARD RASMUSSEN is professor of internal medicine and of cellular and molecular physiology at Yale University School of Medicine. He has had a lifelong interest in the human metabolism of calcium and has published hundreds of articles and a book on the calcium-messenger system. Rasmussen earned an M.D. at Harvard University in 1952 and a Ph.D. at Rockefeller University in 1959. He worked at the University of Wisconsin and the University of Pennsylvania School of Medicine before moving to Yale in 1976.

The sensitivity of a cell to very small changes in calcium concentration reflects the ion's very low concentration within the cell. The concentration of calcium ions is generally 10,000 times greater in the fluid surrounding the cell than in the intracellular fluid, or cytosol. Maintenance of the concentration difference depends on two features of the plasma membrane: its low permeability to calcium and the presence of membrane-bound "pumps" that drive calcium out of the cell, against the concentration gradient. At resting conditions the rate of calcium-ion leakage or influx into the cytosol is balanced by a similar rate of pump-driven efflux of calcium ions.

The Classic Calcium Signal

When a cell is stimulated by an extracellular signal, channels in its plasma membrane open and allow calcium ions to enter at from two to four times the normal rate. Such channels allow calcium but no other ions to flow into the cell cytosol. Some channels open when a neurotransmitter changes the voltage difference that normally exists across the cell membrane; others open when a hormone or neurotransmitter interacts with a cell-surface receptor that is linked to the channels.

The traditional view of calcium as an intracellular messenger is fairly straightforward [see "The Calcium Signal," by Ernesto Carafoli and John T. Penniston; SCIENTIFIC AMERICAN, November, 1985]. Stimulation by a hormone or a neurotransmitter increases the calcium-ion concentration in the cytosol as calcium channels open in the plasma membrane or as calcium is released from the sarcoplasmic reticulum or from calcisomes. When the concentration rises, calcium-binding proteins in the cytosol, such as the specific receptor calmodulin, attach to calcium ions; the calcium-protein complexes then interact with other proteins in the cell to alter their functions. When the calcium concentration in the cytosol falls again, the ions dissociate from the receptor proteins and the system turns off.

In this scenario, calcium acts as a simple on-off switch that conveys information from the cell surface to the cell interior. Calcium does in fact serve as such a switch in several transient cellular responses, including the secretion of neurotransmitters by nerve cells and the contraction of skeletal and cardiac muscle cells. In each case, the rise in calcium-ion concentration in the cytosol initiates the response, and the fall in calcium-ion concentration terminates it.

A First Order of Complexity

While the preceding model of calcium's messenger action was being developed, a similar model was proposed for another messenger molecule known as cyclic adenosine monophosphate (cyclic AMP, or cAMP). It was thought that the synthesis of cAMP at the plasma membrane and its breakdown in the cytosol represented an on-off switch much like variations in calcium-ion concentration. At first the two switches were thought to operate independently.

Now, however, biologists believe that cAMP and calcium usually work together to regulate cell behavior. For example, cAMP can control the rate of calcium cycling across the plasma membrane, and calcium can regulate the enzymes responsible for the synthesis and destruction of cAMP. A single hormone acting by way of a single receptor can cause a simultaneous increase in calcium-ion influx and cAMP production.

Finally, both calcium and cAMP exert many of their cellular effects by controlling the activity of a particular class of enzymes called protein kinases. Protein kinases catalyze the transfer of phosphate groups from a molecule called adenosine triphosphate (ATP) to other proteins. The addition of a phosphate group alters protein function; indeed, widespread protein phosphorylation is thought to underlie the changes in cell behavior induced by some extracellular signals.

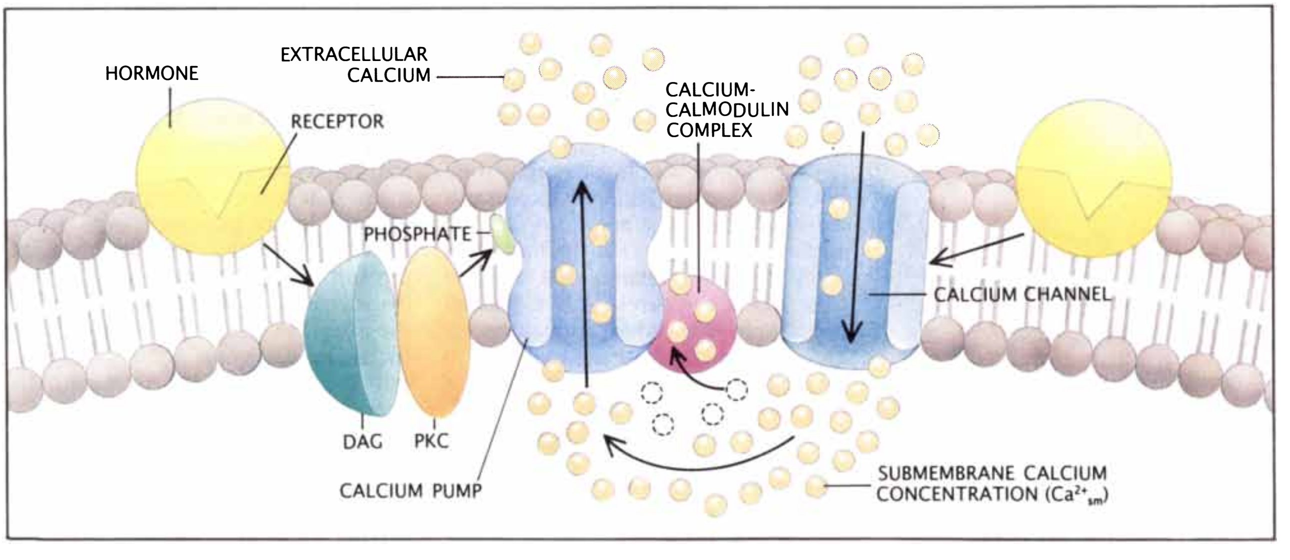
The picture of the individual messenger functions of calcium and cAMP has changed as well. The view that calcium ions and cAMP serve as simple switches is not borne out in all contexts. In particular, that paradigm cannot account for sustained cellular responses to the sustained presence of an extracellular messenger.

My colleagues and I have focused on the role of the calcium ion in sustained cellular responses. We have investigated three such responses: the secretion of the hormone aldosterone, which regulates potassium metabolism, by the cells of the adrenal glomerulosa (part of the adrenal gland); the secretion of insulin by the beta cells of the islets of Langerhans in the pancreas; and the contraction of the smooth muscle cells surrounding the trachea and many blood vessels. In spite of marked differences in the nature of these responses and the extracellular signals that elicit them, we have found that the calcium ion carries out its role as messenger in much the same manner in all three cell types.

A Second Order of Complexity

Until several years ago, biologists thought that sustained, calcium-dependent cellular responses resulted from sustained rather than transient increases in the concentration of calcium ions in the cytosol. This idea was not based on direct measurements of calcium. Instead it was founded on the belief that calcium acted in a manner similar to cAMP, which undergoes a sustained increase in concentration in response to an appropriate stimulus. Yet when the cytosolic concentration of calcium ions was actually measured in stimulated cells, it was found that although the calcium-ion concentration rose as predicted, it did so only transiently and fell back to its base level within a minute or so even when the cells continued to respond for hours.

This paradox forced us to reconsider the classic view of calcium-messenger function. Accordingly, my colleagues Itaru Kojima, Kumiko Kojima,



AUTOREGULATION of the calcium flowing across the plasma membrane is achieved by the control of calcium pumps. The entry of calcium through membrane channels (*right*) increases when a hormone interacts with its receptor. The rise in submembrane calcium concentration (Ca^{2+}_{sm}) stimulates the activity of a calcium pump (*left*) by activating the calcium-binding protein calmodulin and the enzyme protein kinase C (PKC). In this way calcium efflux balances calcium influx. The rise in submembrane calcium caused by such "cycling" constitutes a new kind of calcium messenger.

William J. Apfeldorf and Paula Q. Barrett further analyzed the alterations in calcium-ion metabolism induced in the cells of the adrenal glomerulosa by the hormone angiotensin II—the trigger for aldosterone secretion. We learned that whereas in such cells angiotensin II does cause only a transient rise in the cytosolic calcium-ion concentration, it also causes a sustained twofold increase in the influx of calcium ions.

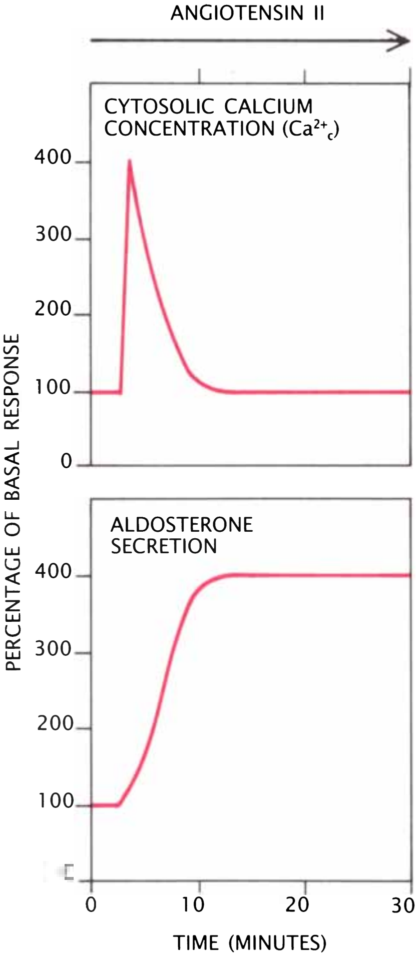
This finding presented a second paradox. It was generally thought that a sustained increase in calcium influx would give rise to a sustained increase in the cytosolic calcium-ion concentration, but clearly that was not the case. It was also assumed that a sustained increase in calcium influx would lead to an increase in the total calcium in a cell, but we found no such increase. We therefore concluded that during the sustained phase of the adrenal-cell response, angiotensin II causes a sustained increase in calcium cycling across the plasma membrane.

The molecular basis for this remarkable ability of the plasma membrane resides in the properties of the calcium-ion pump. It turns out that this pump is activated by a complex of calmodulin and calcium. When cytosolic calcium levels rise, the complex interacts with the pump to enhance both its efficiency and its sensitivity to calcium ions. The efficiency of the pump is further enhanced when it is phosphorylated by a calcium-activated protein kinase called protein kinase C (PKC). The stimulation of the pump by the calcium-calmodulin complex and calcium-activated phosphorylation enables calcium efflux to compensate for the increased influx.

Having discovered the change in the calcium-cycling rate and defined the mechanisms by which it occurs, my colleagues and I explored the possibility that the cycling serves as a messenger during the sustained cellular response. We found that if we blocked the increase in calcium cycling, the response was transient rather than sustained. In other words, calcium cycling is critical to maintaining a sustained response. We concluded that such cycling acts as a messenger because it leads to a change in the calcium-ion concentration in a restricted part of the cell—the "submembrane" domain within or just beneath the plasma membrane.

The next question we wanted to answer was just how this messenger acts. We found that simply increasing calcium influx was not enough to induce a response. Thus, an increase in calcium cycling is a necessary but not sufficient condition for the induction of a sustained response. It became clear that a calcium-sensitive, plasma membrane-associated "transducer" is also required to read the calcium-ion message and convert it into a form that can affect the rest of the cell. Several of these transducer-molecules have been discovered in adrenal cells and other cell types, but the one of present interest is protein kinase C—coincidentally, the same enzyme that regulates the activity of the calcium-ion pump.

The activation of PKC, which enables it to act as a transducer for the sustained, submembrane calcium signal, is linked to the turnover of a class of molecules called inositol phospholipids. The metabolism of these molecules is regulated by a certain class of hormones and neurotransmitters [see "The Molecular Basis of Communication within the Cell," by Michael J. Berridge; SCIENTIFIC AMERICAN, October, 1985]. When such an agent binds to its receptor, an enzyme linked to the receptor catalyzes the breakdown of phosphatidylinositol 4,5-biphosphate (PIP_2), which is a component of the cell membrane. The reaction yields inositol-1,4,5-triphosphate (IP_3) and diacylglycerol (DAG). Released into the cytosol, IP_3 induces the liberation of calcium ions from the calcisomes. Calcium ions released by this mechanism cause a transient increase in the calcium concentration in the cytosol. The rise promotes the formation of calcium-calmodulin complexes, which go on to assist in the phosphorylation of a specific subset of proteins by activating certain protein kinases.



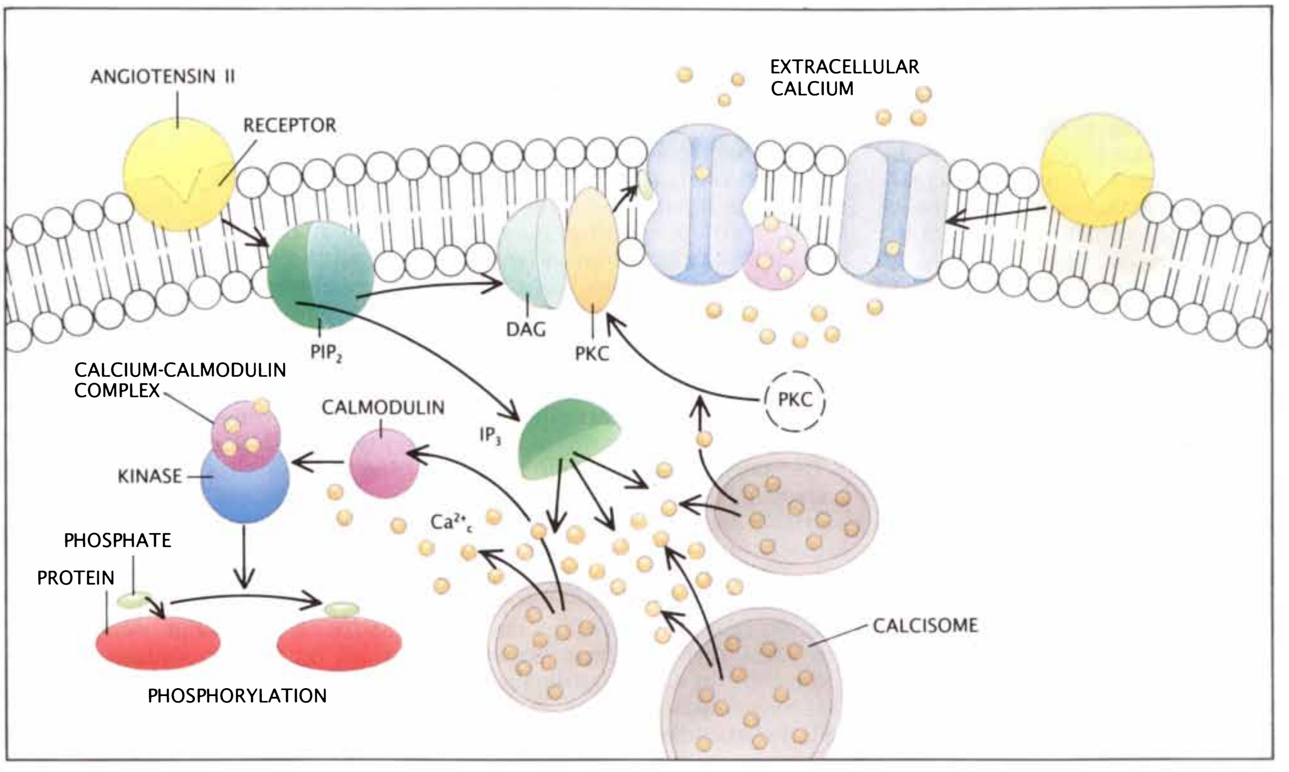
DISSOCIATION with time of changes in calcium-ion concentration (*top*) and cell response (*bottom*) is not predicted by the classic model of calcium-messenger action. These data describe the secretion of the steroid hormone aldosterone from cells of the adrenal glomerulosa (part of the adrenal gland) in response to the hormone angiotensin II. The concentration of calcium in the cytosol, or intracellular fluid, spikes one minute after angiotensin II is added, but aldosterone secretion lasts for more than 30 minutes.

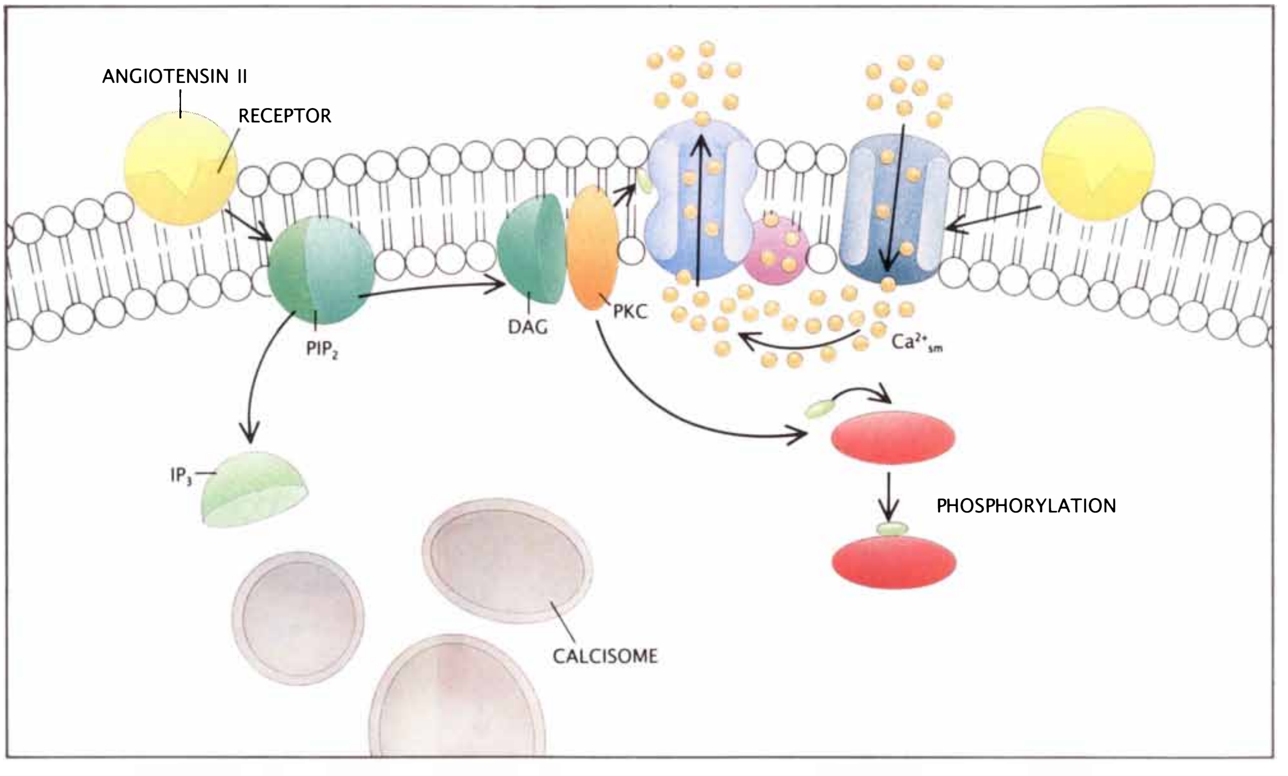
A Two-Pronged Response

The calcium transient and DAG, the other product of PIP_2 breakdown, together cause PKC to associate with the plasma membrane. Unlike IP_3, DAG remains in the membrane; as long as the DAG content of the membrane stays high, PKC remains associated with the membrane as well. The transient release of calcium ions from the calcisomes and the migration of PKC from the cytosol to the plasma membrane are the hallmarks of the initial stage of a sustained cellular response to an extracellular signal.

Receptor activation also causes the twofold increase in calcium-ion influx seen at the plasma membrane. The mechanism by which this increase occurs and the type of calcium channel that opens vary from one cell type to the next. It is not yet clear whether the calcium influx increases as a direct result of receptor activation or as a result of a signal generated by PIP_2 hydrolysis. It is clear, however, that the increase in calcium influx, and therefore in calcium cycling, is a critically important messenger during the sustained phase of the response.

It is thought that the cytosolic form of PKC is relatively inactive. When PKC associates with the plasma membrane, however, it comes in contact with phospholipids (the major components of the membrane) that increase the enzyme's maximal rate of activity by a factor of from 25 to 30 and its sensitivity to calcium ions by a factor of 100 or more. This calcium-sensitive,





ACTIVATION of an adrenal glomerulosa cell by angiotensin II illustrates how calcium acts as a messenger in two different ways during two temporally distinct phases of a prolonged cellular response. In the initial phase (*top*), the binding of an extracellular signal to its receptor prompts the breakdown of the membrane component PIP_2 into IP_3 and DAG. IP_3 causes the release of calcium ions from intracellular compartments called calcisomes, resulting in a transient rise in cytosolic calcium (Ca^{2+}_c). The ions bind to calmodulin, and the calcium-calmodulin complex activates protein kinases (enzymes that transfer phosphate groups to proteins). The phosphorylated proteins initiate the cellular response, which in this case is the secretion of aldosterone. The calcium ions released from the calcisomes and the increase in DAG also prompt the enzyme PKC to associate with the membrane. In the sustained phase (*bottom*), angiotensin II increases calcium cycling across the membrane, and the resulting rise in the submembrane concentration of calcium (Ca^{2+}_{sm}) activates the membrane-associated PKC. This activation brings about the phosphorylation of a different set of proteins that sustain aldosterone secretion.

plasma membrane-associated form of PKC is what acts as a transducer during the second phase of the sustained cellular response. It is the target of the localized change in the concentration of calcium ions brought about by increased calcium cycling. The increase in the submembrane concentration of calcium ions somehow increases the rate at which PKC helps to phosphorylate other proteins.

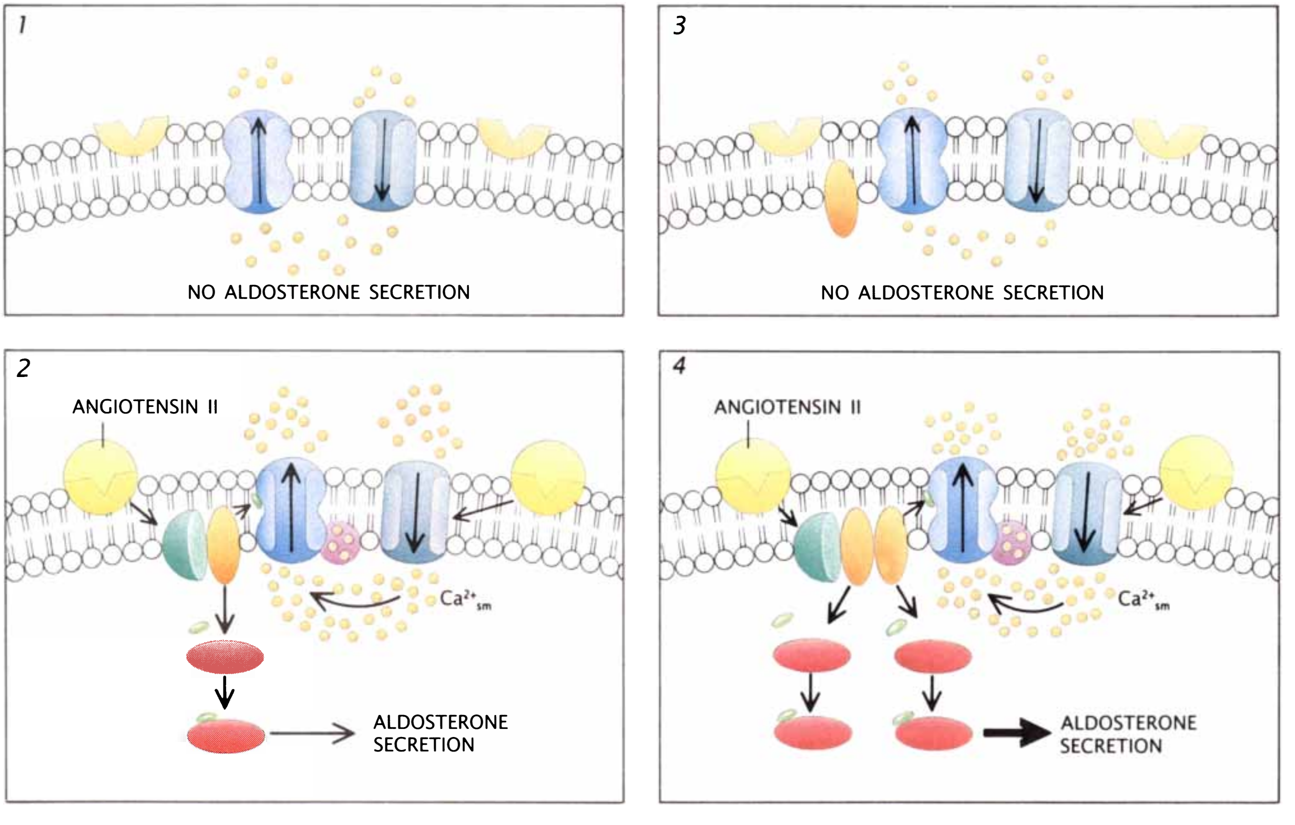
The key feature of the model that has emerged from our studies of a sustained response in adrenal cells is the operation of two temporally distinct branches of the calcium-messenger system: a calmodulin branch, active during the initial phase of response, in which the transient, IP_3-induced rise in the cytosolic concentration of calcium acts on calmodulin-dependent protein kinases to alter the phosphorylation of one subset of cellular proteins; and a PKC branch, in which the rise in calcium concentration in the submembrane domain acts on plasma membrane-associated PKC to alter the phosphorylation of a different subset of cellular proteins involved in mediating the second, sustained phase of the cellular response.

This two-branch model of calcium-messenger action seems to account not just for the secretion of aldosterone from adrenal cells in response to angiotensin II but also for the stimulation of insulin secretion from beta cells and the contraction of smooth muscle. In those systems, however, the operation of the calcium-messenger system depends heavily on the activity of the cAMP-messenger system.

Additional Complexity

The interactions between calcium and cAMP in the regulation of insulin secretion are too complex to be addressed fully in this article. I shall limit my discussion to the action of acetylcholine, a neurotransmitter that attaches to specific receptors on beta cells and thereby brings about in those cells the same signaling events that angiotensin II evokes in adrenal cells. In beta cells, however, the situation is more complex than it is in adrenal cells, because both the intracellular content of cAMP and the extracellular concentration of glucose determine how effective the signals acetylcholine generates will be.

The intracellular content of cAMP can be increased by the binding of certain hormones to specific receptors on the beta cell. One such hormone, called gastric inhibitory peptide (GIP), is released from intestinal mucosal cells in the course of food intake and digestion. When my colleagues Walter S. and Kathleen C. Zawalich and I studied the combined effects of acetylcholine and GIP on insulin secretion from isolated islets in the laboratory, we found that, as we had anticipated, acetylcholine stimulates the breakdown of PIP_2, and GIP boosts the production of cAMP. The remarkable observation was that the effect of such stimulation on insulin secretion depended on the extracellular glucose concentration. For example, when the glucose concentration was similar to that present in the blood before a



"MEMORY" effect occurs when adrenal cells are exposed to successive doses of angiotensin II in the test tube. In the basal state (*1*), calcium cycling is slow, submembrane calcium concentration is low and PKC is not associated with the plasma membrane. The first exposure to angiotensin II (*2*) causes PKC to move to the membrane and increases calcium cycling; these changes in turn increase PKC activity, protein phosphorylation and the secretion of aldosterone. When angiotensin II is removed (*3*), calcium cycling decreases, but PKC remains associated with the membrane. With subsequent exposure to angiotensin II (*4*), more PKC moves to the membrane, enhancing protein phosphorylation and aldosterone secretion. Thus, the persistent association of PKC with the plasma membrane gives rise to a type of memory of angiotensin II stimulation.

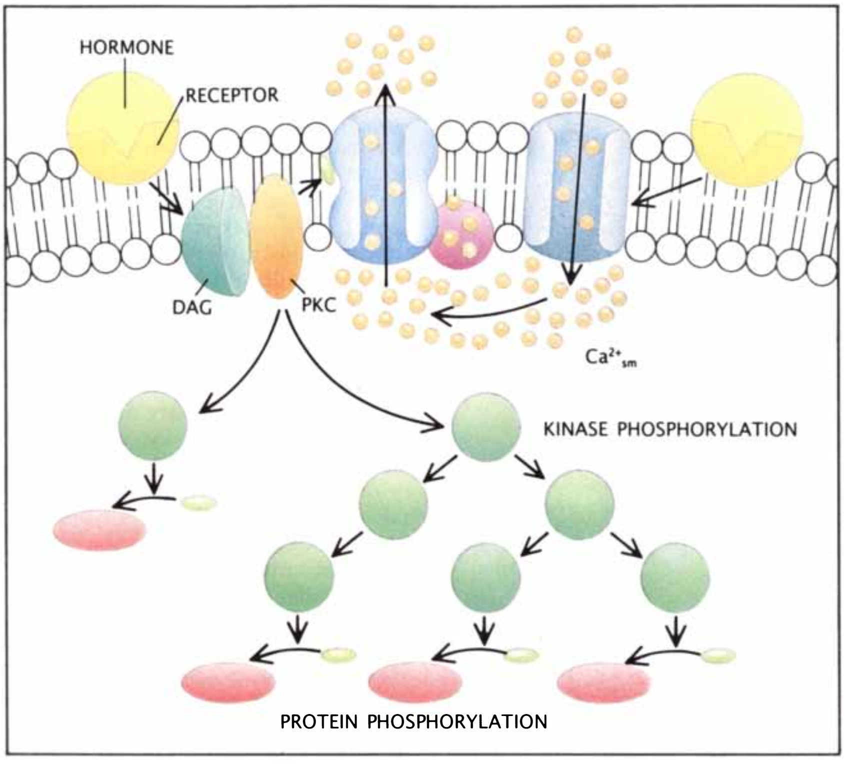
meal, combined acetylcholine and GIP increased insulin secretion only briefly and to a small extent. If, however, the islets were exposed to a glucose concentration about 50 percent higher, which is similar to the one present in the blood from 30 to 60 minutes after a meal, the same combination of acetylcholine and GIP caused a significant and sustained increase in insulin secretion.

The results from these studies and from studies in other laboratories led us to conclude that the different concentrations of glucose altered the effects of acetylcholine and GIP on calcium influx. In the case of the lower glucose concentration, acetylcholine and GIP had little or no effect on calcium influx, but at higher glucose concentrations, they stimulated the influx by way of a specific type of voltage-dependent membrane channel. In the beta cell as in the adrenal cell, a sustained increase in calcium influx (and hence calcium cycling) is essential for a sustained cellular response.

Glucose acts as a conditional modifier of beta-cell responsiveness by controlling the electrical potential of beta-cell membranes. When blood glucose is low, the membrane potential is high, and voltage-dependent calcium channels in the cell membrane stay closed even when acetylcholine initiates a depolarization—a reduction in voltage across the membrane. When blood glucose is higher, however, the membrane becomes partially depolarized: it is poised so that the additional depolarization caused by acetylcholine induces channels to open, letting calcium into the cell.

Furthermore, the increase in cAMP concentration that results from stimulation by GIP also causes certain latent calcium channels to become sensitive to voltage, so that an increasing number of channels respond when an appropriate change in membrane potential occurs. Thus, the rate of calcium-ion influx—and hence of calcium cycling—is increased by two different mechanisms when acetylcholine and GIP act in concert at the proper glucose concentration.

From a physiologic point of view, the control of the calcium-ion influx rate in this way provides a fail-safe mechanism to prevent an inappropriate secretion of insulin when the blood glucose is low—before a meal, for example. A similar fail-safe system operates in adrenal glomerulosa cells. A major effect of the aldosterone they secrete is to lower the concentration of potassium in the blood. A rise in extracellular potassium-ion concentration partially depolarizes the adrenal-cell plasma membrane; the system is thereby poised so that a particular type of voltage-dependent channel opens when angiotensin II is present. When blood potassium falls to a low value, the plasma membrane becomes hyperpolarized, the channels are closed and angiotensin II cannot open them. Consequently, angiotensin II cannot elicit a sustained increase in calcium influx or in aldosterone secretion when blood potassium is low. This fail-safe mechanism prevents secretion of aldosterone when an increase in the plasma concentration of the hormone could have lethal consequences.



KINASE "CASCADE" may explain the ability of PKC activation to alter the phosphorylation of proteins in remote parts of the cell even though the enzyme remains at the plasma membrane. The cascade would begin with PKC activation; PKC could phosphorylate and thereby activate other kinases (*green*), which would in turn activate still other enzymes to modify the function of many different target proteins (*red*).

Altered Responsiveness

Whether the calcium signal acts alone or in concert with cAMP, the two events of critical importance in maintaining a sustained cellular response are the association of PKC with the plasma membrane and an increased rate of calcium-ion cycling across the membrane. In the case of angiotensin II's activation of adrenal glomerulosa cells, those two events take place simultaneously. In certain circumstances, however, the events are temporally dissociated.

For example, in adrenal cells the membrane association of PKC and the increased calcium cycling can become uncoupled to produce a kind of cellular "memory." If isolated adrenal cells are perfused in the laboratory with angiotensin II for three periods of between 15 and 20 minutes, separated by intervals of similar duration, the aldosterone secretion that occurs during each successive exposure is higher than the secretion during the preceding one. Clearly, the adrenal cells "remember" their previous exposure to angiotensin II.

That memory is transient; the longer the interval between exposures to angiotensin II, the less striking the increase in the secretion of aldosterone. The basis of this phenomenon seems to lie in the fact that the PKC associated with the membrane does not dissociate immediately when the angiotensin II signal is terminated. Additional exposure to angiotensin II not only reactivates the PKC that is still associated with the plasma membrane but also recruits additional PKC molecules to the membrane. The result is an enhanced response on reexposure to angiotensin II.

Another example of the same persistent association of PKC with the plasma membrane comes from the action of acetylcholine on beta cells. Recall that when levels of blood glucose are low, acetylcholine stimulates the breakdown of PIP_2 but does not cause a significant increase in insulin secretion because it does not increase the rate of calcium cycling. What could be the role of the acetylcholine signal at that time? My colleagues the Zawalichs and I think we have an answer.

We propose that by stimulating a transient, IP_3-induced elevation of cytosolic calcium concentration and the generation of DAG, acetylcholine brings about the translocation of PKC to the plasma membrane of the beta cell. Because calcium cycling has not increased, the PKC is membrane-associated but is not activated; it is available, however, to become activated when the small, postprandial increase in glucose concentration occurs and leads to a calcium influx. Acetylcholine, then, acts to prepare beta cells to respond to a postprandial rise in blood glucose concentration with a greater release of insulin than the hormone would otherwise bring about.

Kinase Cascades

My colleagues and I extended our investigation of sustained cellular responses by looking at the contraction of smooth muscle, one of the most common tissues in the human body. Smooth muscle—which unlike skeletal muscle is not under voluntary control—is a key component of the walls of the trachea and bronchi and of the blood vessels, ureter, stomach, intestine and uterus. Susanna S.-C. Park, Yoh Takuwa, Grant G. Kelley, Hermann Haller and I have focused our studies on the smooth muscles of the trachea and carotid arteries in the cow. Acetylcholine induces a rapid and sustained increase in the contraction of tracheal muscle, and the extracellular signal histamine induces the same kind of response in the carotid artery muscle.

A model nearly identical to the one developed for angiotensin II action in adrenal cells seems to account for the action of both signals. The initiation of contraction is brought about by a transient rise in cytosolic calcium, which stimulates the calmodulin-dependent enzyme myosin light-chain kinase. A transient increase in the extent of phosphorylation of the myosin light-chain protein follows, which initiates a rapid but transient contractile response. At the same time, the rise in cytosolic calcium (along with the DAG produced by PIP_2 breakdown) also induces PKC to associate with the membrane, as it does in adrenal cells. During the sustained phase of muscle contraction, a rise in calcium-ion concentration in the submembrane domain of the cell activates the plasma membrane-associated PKC, so that a number of proteins become and remain phosphorylated, prolonging contraction.

Two of the high-molecular-weight proteins that are phosphorylated during smooth muscle contraction, namely desmin and caldesmon, have shed light on a question presented by our new model: How does plasma membrane-associated PKC exert its effects on proteins at distant locations in the cell? Desmin and caldesmon are two such proteins, localized in domains of the cell that are remote from the site of PKC action. Given their intracellular location and their associations with complex, highly organized macromolecular structures, it is unlikely that either of these proteins shuttles between its particular domain and the membrane region where PKC operates.

Yet our own studies and those of David R. Hathaway and his colleagues at the Indiana University Medical Center show that both proteins do in fact become phosphorylated during the sustained phase of tracheal or carotid artery smooth muscle contraction. In the test tube PKC will phosphorylate desmin and caldesmon directly. Close examination reveals, however, that the site of this test-tube phosphorylation is different from the site of phosphorylation during the contraction of intact muscle in response to treatment with either acetylcholine or histamine.

We derive two conclusions from these results. First, many intracellular proteins may be potential substrates for a given protein kinase, but in the cell they do not serve as such because they reside in a subcellular domain different from the one in which the kinase normally functions. Second, the phosphorylation of desmin and caldesmon in the stimulated muscle is probably achieved by a protein kinase other than PKC.

Because PKC does seem to trigger the phosphorylation, we postulate the presence of protein kinase "cascades" in which one or more of the substrates of PKC are themselves protein kinases. Kinase phosphorylates kinase until eventually one of the kinases in the cascade phosphorylates, say, desmin or caldesmon. Workers have already seen such a cascade in the action of insulin on its target cells.

A New Calcium Messenger

In our studies of calcium-messenger function in sustained cellular responses, we have found that, contrary to the classic model of calcium as messenger, a rise and fall in the cytosolic concentration of calcium ions appears to operate as an intracellular messenger only during brief cellular responses or during the initial phases of sustained responses. During the sustained phase, a calcium signal is generated in a restricted region of the cell membrane by an increase in the rate of calcium cycling across the membrane. This submembrane calcium signal acts on calcium-sensitive, plasma membrane-associated transducers to generate other signals. By and large, it is messengers generated by the transducers—rather than calcium or the transducers themselves—that convey information from the cell surface to the cell interior.

Much remains to be learned about kinase cascades and about the separate control of calcium cycling and the plasma membrane association of PKC. Yet already a growing awareness of this new type of calcium-ion messenger has contributed to research on associative learning [see "Memory Storage and Neural Systems," by Daniel L. Alkon; SCIENTIFIC AMERICAN, July]. Because it touches on insulin secretion and blood-vessel constriction, research on the messenger should illuminate the sequence of events leading to diabetes and high blood pressure as well.

FURTHER READING

CALCIUM AND cAMP AS SYNARCHIC MESSENGERS. Howard Rasmussen. John Wiley & Sons, Inc., 1981.

INTRACELLULAR CALCIUM: ITS UNIVERSAL ROLE AS REGULATOR. Anthony K. Campbell. John Wiley & Sons, Inc., 1983.

THE CALCIUM MESSENGER SYSTEM. Howard Rasmussen in *New England Journal of Medicine*, Vol. 314, No. 17, pages 1094–1101; April 24, 1986 and Vol. 314, No. 18, pages 1164–1170; May 1, 1986.

A SPATIAL-TEMPORAL MODEL OF CELL ACTIVATION. Daniel L. Alkon and Howard Rasmussen in *Science*, Vol. 239, No. 4843, pages 998–1005; February 26, 1988.

PHOSPHORYLATION OF CALDESMON IN ARTERIAL SMOOTH MUSCLE. Leonard P. Adam, Joe R. Haeberle and David R. Hathaway in *Journal of Biological Chemistry*, Vol. 264, No. 13, pages 7698–7703; May 5, 1989.

SCIENCE AND BUSINESS

Light Talk

U.S. and Japanese compete to put optical fibers in the home

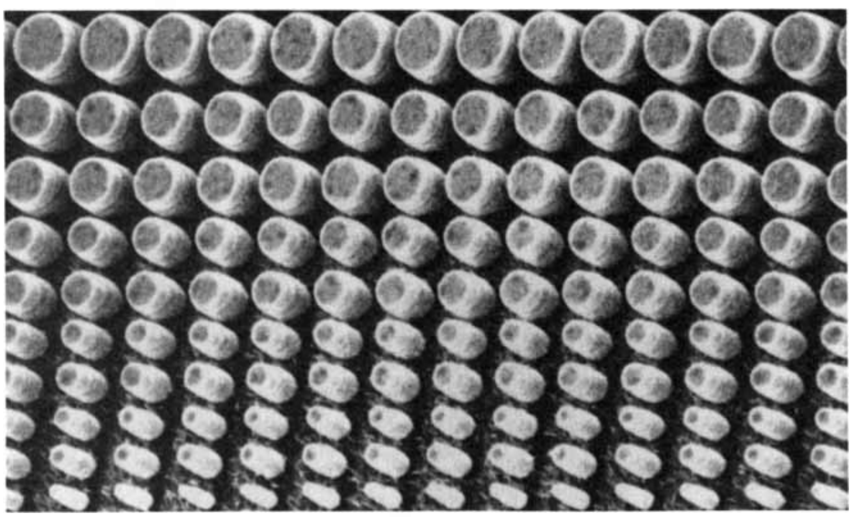
Fiber-optic cables are finally coming home. During the past few years telecommunications companies have stretched such cables across continents and under oceans to carry data over long distances. Regional phone companies have linked central switching offices with fiber cables. Now the local phone companies and equipment manufacturers are working together to string fiber all the way to consumers' doors.

Marshall McLuhan's global village awaits. The optical links, able to carry many times as much data as copper cables, should make vast amounts of information, pictures and sound available at home. For telecommunications equipment manufacturers, fiber to the home means a vast new market totaling at least $100 billion in the U.S., according to a report last year by the National Research Council. Global villages, however, demand low-cost, reliable photonics components—devices that replace electrons with photons and are the lifeblood of optical communications networks. These include optical-fiber cables, laser-based transmitters and receivers and a collection of devices for tightly packing signals, regenerating them and so on.

Homespun fibers, live-in robots, green profits, costly capital

Yet U.S. companies, including AT&T, still lag in manufacturing, says Stewart D. Personick of Bell Communications Research, who contributed to the National Research Council's photonics report. As a result, they may find themselves scrambling for a share of the market. "There are five or six big Japanese companies all going after the fiber-to-the-home market," says David V. Lang, a director at AT&T Bell Laboratories. "We can't take on Japan by ourselves."

No one questions American prowess in photonics research. U.S. engineers invented much of the early technology and hold critical patents on lasers and optical fibers. Current U.S. research efforts are among the best in the world. At a meeting this summer in Japan, a team of engineers from Bell Communications Research and AT&T Bell Laboratories astonished their international colleagues by packing two million lasers on a tiny substrate. Although strictly experimental, the technique for building such chips may be an important step toward the better integration of electronics and optics.



MICROSCOPIC LASERS, from one to five microns in diameter, were chiseled from a multilayered semiconductor substrate by a team of workers from Bell Communications Research and AT&T Bell Laboratories. A million of the lasers occupy a square centimeter on a chip.

In past years AT&T has parlayed its strength in optical fibers into a strong business in long-distance equipment. Apart from laying down long-distance cables in the U.S., AT&T has won contracts to run a second set of optical cables across the Atlantic and Pacific oceans in the early 1990's. Japan's telecommunications giant, NTT, has also selected AT&T to build a very high speed (2.5 gigabits per second) long-distance network in Japan. But AT&T faces much stiffer competition in the manufacture of other photonics components. "There are literally dozens of equipment vendors," says Bill Judge, a manager with NYNEX. "It's anybody's ball game if they can bring down the cost."

Cost is at the crux of the fiber-to-the-home industry. Until this year, the regional Bell companies had been reluctant to construct fiber networks to homes because it was too expensive. Investments in long-distance fiber lines could be justified on the basis of the additional capacity such networks would provide. But the local phone companies are discouraged from raising rates simply to cover the cost of wiring with fiber. They must, therefore, find ways to lay fiber networks that are no more expensive than copper ones from the start, Judge says. Now, as the price of fiber has dropped, regional phone companies have initiated about 20 fiber-optic pilot projects for homes.

Already, the local phone companies say, a diverse range of manufacturers have offered full fiber-to-the-home networks. At the top of the list have been AT&T and Canada's Northern Telecom. Southern Bell used both companies in its four pilot fiber networks for transmitting voice. But as Southern Bell gears up to invest heavily in fiber to the home, it is looking beyond these suppliers, says Robert A. Morrow, a manager at Southern Bell. He says that contracts are being finalized with four manufacturers. "If you can come in at a certain price, we'll buy from you," he emphasizes.

NYNEX turned to a small company called Raynet, based in Menlo Park,

SCIENCE/SCOPE®

A small satellite earth station, capable of simultaneously receiving voice, video, and digital data information, allows companies to reduce their communications costs. The Hughes Aircraft Company Personal Earth Station™ uses patented demand-assigned, packet-switched transmission techniques, in which different types of information are transmitted over a single communications channel as small, high speed data packets. Other types of communications systems require expensive, separate channels for voice, video and digital data information. The earth stations are part of Hughes' Very Small Aperture Terminal network, which provides end-to-end satellite communications for private business data networking and videoconferencing.

Application of integrated circuit processing technology may lead to higher density packaging. For the U.S. Navy's VLSIC Packaging Technology (VPT) program, Hughes will apply its high-density multichip interconnect (HDMI) technique which uses integrated circuit processing technology to build the substrate circuitry in a hybrid package. The polymide dielectric used in the HDMI process is considered excellent for high-frequency device applications. The packaging technique is aimed at meeting the need for higher density hybrid microelectronics to take advantage of the next generation of integration using Very High Speed Integrated Circuits (VHSIC-II) chips.

A faster, deeper running, more accurate torpedo will counter enemy submarine threats through the year 2000. The Mk-48 Advanced Capability (ADCAP) wire-guided torpedo, under production at Hughes for the U.S. Navy, uses an advanced digital-based active sonar, an inertial measurement unit, a signal processor and two computers to home in on fast moving targets in shallow water, high seas, strong thermal gradients, even under ice. A long, thin communications wire between the torpedo and submarine serves as a real-time relay for changes in the torpedo's attack functions, improving the probability of interception.

Image processing computers will perform 100-billion operations per second in a package slightly larger than a tuna fish can. Under development by Hughes for the U.S. Air Force, the new 3-D Computer consists of a number of integrated circuit wafers stacked like records in a juke box. Each wafer contains a large array of processors with each processor connected directly to its four neighbors. Using a unique technology, signals are passed vertically through each wafer. Adjacent wafers are connected via "microbridges," permitting the distribution of signals to all other cells in a vertical column. The computer's signals are thus moving in three dimensions instead of two. This ultra-fast supercomputer will be used in space-based missions and applications like image and radar signal processing.

Engineers and scientists are eligible for approximately 100 Hughes Fellowships awarded for the pursuit of Master's and doctoral studies in Engineering and Science. All Fellows work full time at Hughes during the summer, with Work-Study Fellows working part time during the academic year and Full-Study Fellows attending classes full time. Fellows receive full academic expenses plus stipends for studies at approved universities. Additionally, Hughes offers a two-year, entry-level rotation program that enables qualified BS and MS graduates to diversify their engineering experience. For more information contact the Hughes Corporate Fellowship Office, Dept. S2, C1/B168, P.O. Box 45066, Los Angeles, CA 90045-0066. U.S. citizenship may be required. Equal Opportunity Employer.

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

HUGHES

Subsidiary of GM Hughes Electronics

Food with thought.

A great deal of care, thought and preparation goes into our international First and Executive Class cuisine. With continental and regional dishes such as this Beef Tenderloin Careme, we've recently won our second National Restaurant Award in the last three years. Judge for yourself. Once you try food like this, you won't give anyone else's a second thought. To fly First or Executive Class to Europe or Asia, call your travel agent or Northwest at 1-800-447-4747.

NORTHWEST AIRLINES

Calif., when it began its first fiber-to-the-home project this summer. Raynet offers an unusual topology that it says is cheaper than AT&T's architecture. In AT&T's design, a central office serves as a hub for a collection of remote terminals, which in turn serve an array of curbside optoelectronic boxes, in which the optical signals are switched. These then become the hubs that are linked, via fiber-optic spokes, to each of the households in the network.

In contrast, Raynet has designed a bus architecture. A fiber-optic cable runs through an area; hanging off that bus are as many as two dozen links to curbside nodes, each of which is then connected to three or four houses. The topology reduces the number of optoelectronic devices needed. Moreover, whereas AT&T architecture brings fiber to every doorstep, Raynet relies on high-speed copper cables to link the nodes to the individual homes. Over a few hundred yards, Judge says, copper cables can transmit data at about 1.5 megabits per second. Eventually, as prices for photonics devices drop, he says, dedicated fibers will replace these copper cables.

As AT&T works to make headway in the fiber-to-the-home market, it will rely increasingly on its Solid State Technology Center, which opened last year in Pennsylvania. The purpose of the center is to push devices faster from the idea stage to manufacturing. Although the lab is exploring everything from new photonics materials to optical-component subsystems, there is a strong emphasis on economics. "The old Bell system made everything it needed in a controlled way so it built the best system," Lang says, regardless of the cost. "Now our challenge is to come up with the best cost-to-performance ratio." *—Elizabeth Corcoran*

Homebody

The "father" of robotics plots its future

When Joseph F. Engelberger quit Unimation in 1984, he announced he was retiring from the robotics industry and bought a sailboat. Sailing lasted about two months. Now the man often credited with sparking the industrial robotics industry wants to put a robot in every home. "I want it to scrub the floor, cook the meals, get outside and wash my car and handle security and all those things," he says enthusiastically.

Engelberger's love affair with robots began more than 30 years ago. Back then his ideas met with skepticism and apathy; robot fever finally took hold in the late 1970's, only to fade quickly. Recently, however, the Robotic Industries Association has begun predicting a comeback for U.S. industrial robots: sales reached almost $290 million in the first half of 1989 as compared with about $330 million for all of last year.

ROBOT POSES WITH ITS CREATOR, Joseph F. Engelberger. The robot—an early model—was designed to deliver dinner trays directly to hospital patients. Engelberger is working on larger robots that will wash and vacuum floors. Photograph by Wayne Sorce.

As Engelberger seldom tires of telling, the robotics industry began at a cocktail party in Westport, Conn., one evening in 1956. George C. Devol, an inventor with a string of patents, was pitching his latest ideas about an industrial robot. Devol had not yet built a robot, nor had he had any luck trying to convince "an awful lot" of corporations to invest in one. Yet he caught the imagination of Engelberger, then a company manager.

"I suppose in the haze of alcohol it looked like a better idea than it really was," Engelberger says. As a student at Columbia University, Engelberger had been thrilled by Isaac Asimov's tales of robots. Now he was hooked. Automated machines dedicated to a single task already existed. Devol envisioned building programmable "pick and place" machines that could undertake a variety of tasks.

After touring several factories, Engelberger and Devol concluded that the automobile industry was ripe for robots. Building cars was a steady process that involved heavy machinery and so was well suited for strong, albeit slow, one-armed robots. Most important, men worked in multiple shifts. "We were trying to replace labor on economic grounds," Engelberger says.

By about 1959 Engelberger and Devol secured enough funding to build their first robot, which was delivered to a General Motors plant in New Jersey in 1961. The Unimate, as Devol named it, was a hydraulic-powered robot that ran a die-casting machine. To "learn" its task, the robot was led painstakingly through the process so it could make a digital magnetic recording of what it did as it went.

Orders came slowly and often from surprising quarters. Many early customers were what Engelberger now calls "irrational buyers," people curious about the new tools and eager to try them out. In 1962 the chairman of Pullman Railcars, charmed by the idea of robots, contributed $3 million to the newly formed Unimation in exchange for 51 percent of the company. Unimation's first major success came in 1966 when General Motors bought 66 robots to do spot welding in its new Youngstown, Ohio, plant.

Although colleagues recall that Engelberger made technical contributions to robot design, he became more widely known as a showman of robots. Unimates made their debut on a Johnny Carson show in 1967 by putting a golf ball into a cup and conducting the orchestra. A Unimate showed up in a beer commercial. "At that time, it was very hard to get enough attention," Engelberger says, "so you did gags." On the Dean Martin show, "I was asking, 'When do I get to tell people what the robot really does?' and they said, 'Oh, we can't give any boosts on our show.'" So the robots did tricks, Engelberger recollects.

Engelberger helped to lay the foundations for Japan's robotics industry as well. In 1967 that government flew him to Japan first-class to give a presentation on robots. "Where I had trouble getting eight or 10 people to listen to me in the U.S., they brought 600 people in for a lecture," Engelberger says. (The audience consisted of engineers and senior executives.) By the end of the trip, Engelberger had agreed to license Unimation's technology to Kawasaki Heavy Industries. "The license was a good idea," he argues. Others have raised doubts, particularly now that Japan dominates the field. But Engelberger counters

that the Japanese copied the robots built by other small U.S. companies.

Unimation took 14 years to become profitable, but by the early 1980's it had cornered more than 30 percent of the robotics market and had placed more than 7,000 robots in factories. Its Puma line, which had rotational joints and could assemble parts or weld metal seams, rapidly won acceptance. Robots from the U.S., Japan and Europe were selling briskly. Former Unimation engineers were starting their own companies. "I would be baffled by an average [industrial] growth rate of less than 35 percent annually through the decade," Engelberger wrote in *Industrial Robot* magazine in 1983.

Yet trouble was on the way. The rosy estimates of the future market attracted too many manufacturers, says Maurice J. Dunne, a retired Unimation vice-president. According to Engelberger, the large manufacturers that entered the industry were willing to absorb losses to win business. When General Motors co-founded GM Fanuc Robotics (with the Japanese manufacturer Fanuc), Unimation lost 60 percent of its business, Dunne adds.

Westinghouse Electric bought Unimation in 1983 for $107 million. But within a year, Unimation and the industry were in a tailspin. Critics have charged that Unimation failed to keep pace with technology; they argue it should have replaced the Unimate's hydraulic-power system with electric motor drives and added more microelectronics. Engelberger counters that the early Unimates were well equipped for many tasks and the company's West Coast division employed advanced electronics. (The division is now a freestanding company called Adept Technology and is a leading robot maker for the electronics industry.) After about six years Westinghouse absorbed some of Unimation and sold the remaining pieces.

Engelberger, meanwhile, was exploring mobile robots with tactile and vision sensors. He founded Transitions Research Corporation in Danbury, Conn., and began work on a robot that could do housekeeping chores.

TRC has several prototypes now; two experimental "HelpMate" robots scoot around a local hospital fetching dinners for patients. The boxy robots even use the elevators. At the TRC office, a similar robot practices vacuuming the floor of a mock living room. Another aspires to scrub floors. "The dream, back to Asimov," Engelberger says, "is that a machine is most robotic when it is most human." —*E.C.*

Good Intentions

Earning money the socially conscious way

Want to better the environment? You could recycle your trash—or, some fund managers suggest, you could put your money in socially conscious investments.

Investing in companies that boast strong records of environmental protection and show other signs of social consciousness is a growing trend, says Gordon Davidson, executive director of the Social Investment Forum in Boston, an association for investment managers. In early September the forum established a project—the Coalition for Environmentally Responsible Economies (CERES)—devoted to fostering investments in companies with sound environmental policies.

Even though social investments amount to about $450 billion, they represent less than 10 percent of the more than $5 trillion invested in U.S. corporate stocks and bonds. About 80 to 90 percent of all social investments simply avoid companies that invest in South Africa, Davidson observes. The $350-million Calvert Social Investment Fund in Bethesda, Md., automatically rejects companies involved with nuclear energy, South Africa and weapons systems, then looks for companies that pass minimum standards in such areas as environmental safety and employee relations.

Still, "there's no perfect company," says Andrew Rubinson, an assistant portfolio manager at the Parnassus Fund in San Francisco. For instance, many high-technology companies in Silicon Valley treat employees well and donate to charities—but have soiled environmental records. On the flip side: not all "environmental technology" funds invest in socially responsible companies, experts say.

Getting a clear picture of the behavior of a large corporation can be difficult, too. Fund managers at the U.S. Trust Company of Boston, who oversee the Calvert Social Investment Fund, often call environmental groups to assess companies' records. A handful of firms, such as Franklin Research & Development in Boston, publish reviews of companies. CERES has set some broad environmental standards.

A company's good behavior does not necessarily mean it will produce good returns, says W. Scott Klinger, a vice-president at Franklin Research. But in the hands of an adept manager, social funds can earn as much as conventional ones do, he adds. Investors, moreover, may have dual motives for investing in such funds. "Some are altruistic," says David J. Schoenwald, vice-president of the New Alternatives Fund in Great Neck, N.Y. "Other investors think they'll earn great returns in the '90's," he adds. They may have a point: under the current administration, corporations with faulty environmental records may have to pay huge clean-up bills that will hurt their profitability. As a result, cleaner companies may indeed turn out to be good investments. —*E.C.*

THE ANALYTICAL ECONOMIST

The cost of capital

Imagine two manufacturers—one in the U.S., the other in Japan—each weighing whether to add a production line. The expansion would cost each company $100 million this year but in the long run could help the company to grab a larger market share. The U.S. managers argue that the investment must pay for itself in four years; the Japanese managers give it eight. As a result, the U.S. company decides the deal is too risky and forgoes the investment even as its Japanese competitor proceeds.

Was the U.S. firm shortsighted or economically prudent? Ceteris paribus, or all else being equal, as economists are fond of saying, both firms may have chosen wisely. The difference, say economists and business leaders, is that U.S. companies perceive the "cost of capital" to be higher than do their competitors overseas.

Why should capital cost more here than there? In broad terms, the cost of capital is the expense of securing the funds used to buy facilities and equipment. A more precise definition is the one employed by George N. Hatsopoulos, chairman of Thermo Electron, and Stephen H. Brooks, an economic consultant, who together did a landmark study on the subject. They describe the cost of capital as the rate of return (accounting for inflation, depreciation and taxes) that companies calculate they must earn on an investment in order to justify spending the money in the first place.

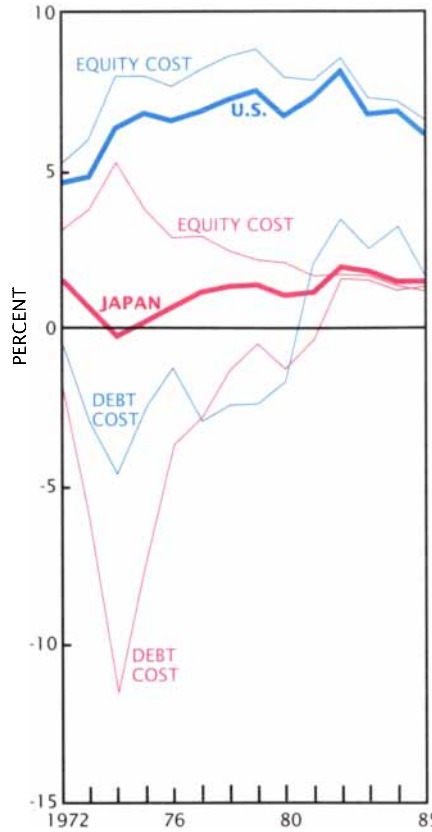
Capital comes in two basic flavors: debt (such as bank loans and corporate-issued bonds) and equity (such as stock). Varying the proportions of debt and equity in financing a deal changes the cost—and the risk—of the project. The trade-off: from the vantage of U.S. managers, equity is less risky, but debt is cheaper in comparison. (Debt payments are tax-deductible; dividends paid to stockholders are not.) Debt has typically been cheaper in Japan as well, but managers there may not perceive debt financing to be particularly risky, economists say.

What governs the costs and risks of these types of financing? The cost of debt reflects prevailing interest rates. These rates rise and fall: banks' lending rates vary with the tides of the economy, whereas bond yields largely reflect Wall Street's assessment of a company's creditworthiness. The current U.S. prime lending rate is about 10.5 percent as compared with less than 5 percent in Japan; highly rated, or triple-A, corporate bonds have annual yields of almost 9 percent in the U.S. and 5 percent in Japan. Accounting for inflation rates, however—currently about 5 percent in the U.S. and less than 3 percent in Japan—shrinks the differences, observe Hatsopoulos and Brooks, as well as Albert Ando and Alan J. Auerbach, economists at the University of Pennsylvania.

Yet debt is risky: a company that neglects to pay its debt holders may be sued or liquidated. Moreover, Wall Street analysts, who rate a company's creditworthiness in great part by its ability to pay off its debt, downgrade highly leveraged companies.

Calculating the cost of equity is trickier. Although many companies have evolved elaborate equations, Hatsopoulos asserts that most managers (including himself) rely on "gut intuition." Essentially, managers must estimate how much of an increase in future corporate earnings and dividends stockholders will demand in return for depressed earnings and dividends now. The value of the increased future earnings and dividends is discounted by combining the risk involved, the current returns and the returns investors could earn elsewhere. So the cost of equity represents the rate at which future earnings and dividends are discounted. Hatsopoulos and Brooks estimate Japanese equity costs are roughly one third of U.S. costs; applying a different approach, Ando and Auerbach find small differences.

U.S. managers' efforts to maximize the value of their stock account for much of the disparity, Hatsopoulos argues. Because stockholders have no guarantee they will see their money again, they demand the prospects of



CORPORATIONS' COST OF CAPITAL is the weighted average of the expense of borrowing money (debt) and of issuing stock (equity). Dark lines show the combined cost for the U.S. and for Japan, based on studies by George N. Hatsopoulos of Thermo Electron and Stephen H. Brooks, an economic consultant.

high future returns to compensate for their risk. Company managers also worry about how the financial markets will react to their decisions. Professional financial investors who manage large pension funds are obligated to maximize value. That can lead them to seek short-term returns from their investments. "Managers are captive to investors," Hatsopoulos declares.

Unlike U.S. corporations, Japanese companies do not have to realize returns of a certain percent in the short term, explains Katsumi Shimizu, a senior vice-president at Daiwa Bank Trust Company. Japanese companies have traditionally paid negligible dividends to investors, in part because often more than half of a company's stock was held by a bank or another corporation, Shimizu says. "Since companies don't have to worry about takeovers, they can concentrate on corporate strategy," he adds.

Adding up the costs and risks of debt and equity financing, then, U.S. managers have traditionally chosen to rely on costly equity while their Japanese counterparts have turned to debt. From 1967 to 1983, U.S. companies had an average ratio of debt to market value of 26 percent, Ando and Auerbach report. During the same period Japanese companies had a debt-to-value ratio of 63 percent.

The bottom line, Hatsopoulos and Brooks say, is that the cost of capital is almost three times higher in the U.S. than in Japan. They fix much blame on differences in capital-gains taxes. In the U.S., investors pay the same proportion of taxes on capital gains as they do on income from interest; in Japan, investors pay very low capital-gains taxes and so do not demand future corporate earnings to be as high as U.S. investors do.

Ando and Auerbach also conclude that capital costs have been lower in Japan. Although they do not point to specific causes, they suggest that the capital markets have not behaved in economically rational ways; Japanese investors may not be earning adequate compensation for their risk.

From the vantage of U.S. companies, a lower cost of capital would be beneficial. But there are no quick fixes, economists say. Simply lowering interest rates would discourage foreign investment in the U.S. and dry up needed funding for the federal debt.

Increasing corporate debt also receives mixed reviews. "There's nothing wrong with a lot of leverage," says George J. Kirk, director of strategic studies at Westinghouse Electric. But Auerbach says that if more U.S. companies turn to debt for financing, interest rates on debt will surely rise.

Auerbach suggests that the difference in the cost of capital may narrow as capital markets become more liberalized. During the past decade Japan has begun easing restraints on investments abroad; in 1985, for instance, it began permitting some large institutions to place more of their investments overseas. Other trends may play a role in changing the relative cost of capital as well. As Japanese companies establish more facilities outside Japan, they may find it increasingly expensive to borrow from home: if they earn revenues in a currency that falls against the yen, it will become more costly for the expatriate firm to repay its yen debts.

Lowering capital-gains taxes, Hatsopoulos argues, would help. Yet, he warns, "you'll probably have to wait 10 years before the financial markets force companies to create teams of farsighted managers."

—Elizabeth Corcoran and Paul Wallich

Carbon Monoxide and the Burning Earth

Measurements of atmospheric carbon monoxide from space have found large amounts of the gas in unexpected places. Tropical burning rivals transportation and industry as a source of carbon monoxide

by Reginald E. Newell, Henry G. Reichle, Jr., and Wolfgang Seiler

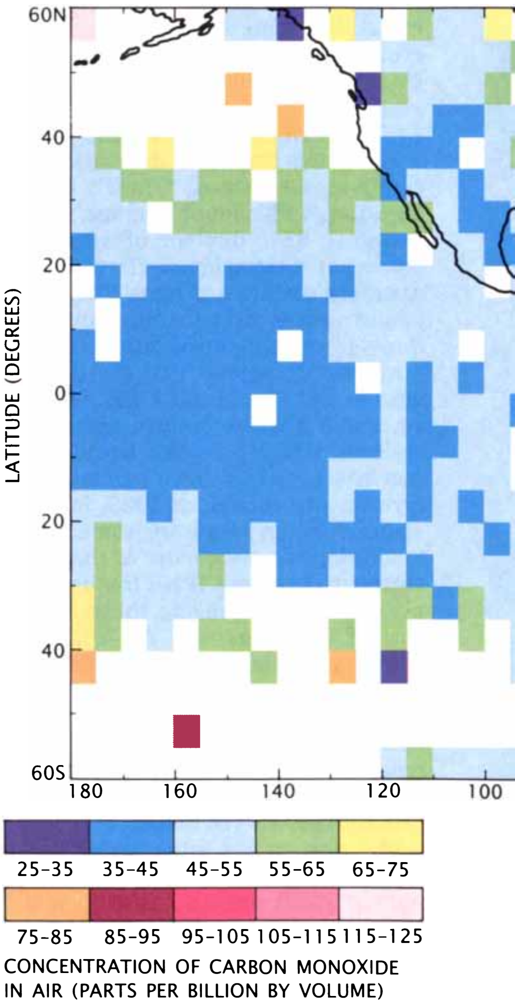
Twenty years ago, a map of the atmosphere illustrating the presumed distribution and flow of carbon monoxide would have placed almost all of it in the Northern Hemisphere. Experts agreed that virtually all carbon monoxide came from burning fossil fuels, and the Northern Hemisphere was home to most of the world's industry and transportation. The map would have shown much of the gas lingering near the ground where it was produced, in the lowest two kilometers of the atmosphere, or boundary layer. Convection might carry some to higher altitudes where it could drift into the Southern Hemisphere; otherwise carbon monoxide stayed in the north.

That map would have been wrong. Industrial smokestacks and automobile tail pipes are not and have never been the only major sources of carbon monoxide—they may not even be the largest sources of this colorless, odorless gas. An instrument that we helped to develop, and that flew twice on board the space shuttle, has provided global "snapshots" of carbon monoxide distributions over a wide swath of the earth. Its measurements and data collected by airplanes and ground stations during the past 10 years have shown clearly that the burning of tropical rain forests and savannas generates at least as much carbon monoxide as the burning of fossil fuels does.

This discovery is alarming, but not because carbon monoxide is poisonous. The gas does bind strongly to the blood's hemoglobin and prevents oxygen from being carried to the tissues. Concentrations of carbon monoxide typical of traffic tunnels or busy streets, around 20,000 molecules per billion air molecules, can dull mental acuity. Yet the concentrations of carbon monoxide over the tropical rain forests are usually hundreds of times lower than that. Instead carbon monoxide has other ominous implications for the environment. First, high levels of carbon monoxide from burning vegetation confirm other evidence that the rain forests are being diminished rapidly. Destruction of the tropical rain forests would probably wreak disastrous change on the climates of these regions and possibly those of the rest of the world. Second, major increases in atmospheric carbon monoxide could encourage the accumulation of pollutant gases such as ozone, which is highly toxic to plants, and methane, which adds to the greenhouse effect and thus may contribute to higher global temperatures.

For many years high concentrations of carbon monoxide in the Southern Hemisphere and tropics seemed unlikely because sources of the gas were thought to be exclusively industrial and automotive, and mainly in the Northern Hemisphere. Until relatively recently, studies of atmospheric carbon monoxide made from the ground or sea level generally supported that idea. For example, in 1969, samples of air collected from a ship by one of us (Seiler) and Christian Junge of the Max Planck Institute for Chemistry in Mainz while

REGINALD E. NEWELL, HENRY G. REICHLE, JR., and WOLFGANG SEILER have collaborated on the study of atmospheric carbon monoxide during the Measurement of Air Pollution from Space project. Newell is professor of meteorology at the Massachusetts Institute of Technology. His main interests are in the large-scale general circulation of the atmosphere and the physics of climatic fluctuations. Reichle is a senior research scientist at the NASA Langley Research Center and has been active in research on the remote measurement of atmospheric properties since 1965. Seiler is the director of the Fraunhofer Institute for Atmospheric Environmental Research in Garmisch-Partenkirchen, West Germany.



sailing the tropical Atlantic Ocean showed higher carbon monoxide values north of the equator. During the same period, measurements made by aircraft flying between Frankfurt and Johannesburg at an altitude of about 10 kilometers showed almost identical carbon monoxide concentrations in the Northern and Southern hemispheres. The results were ascribed to highly efficient mixing of Northern Hemisphere air into Southern Hemisphere air at great heights [see "The Global Circulation of Atmospheric Pollutants," by Reginald E. Newell; SCIENTIFIC AMERICAN, January, 1971].

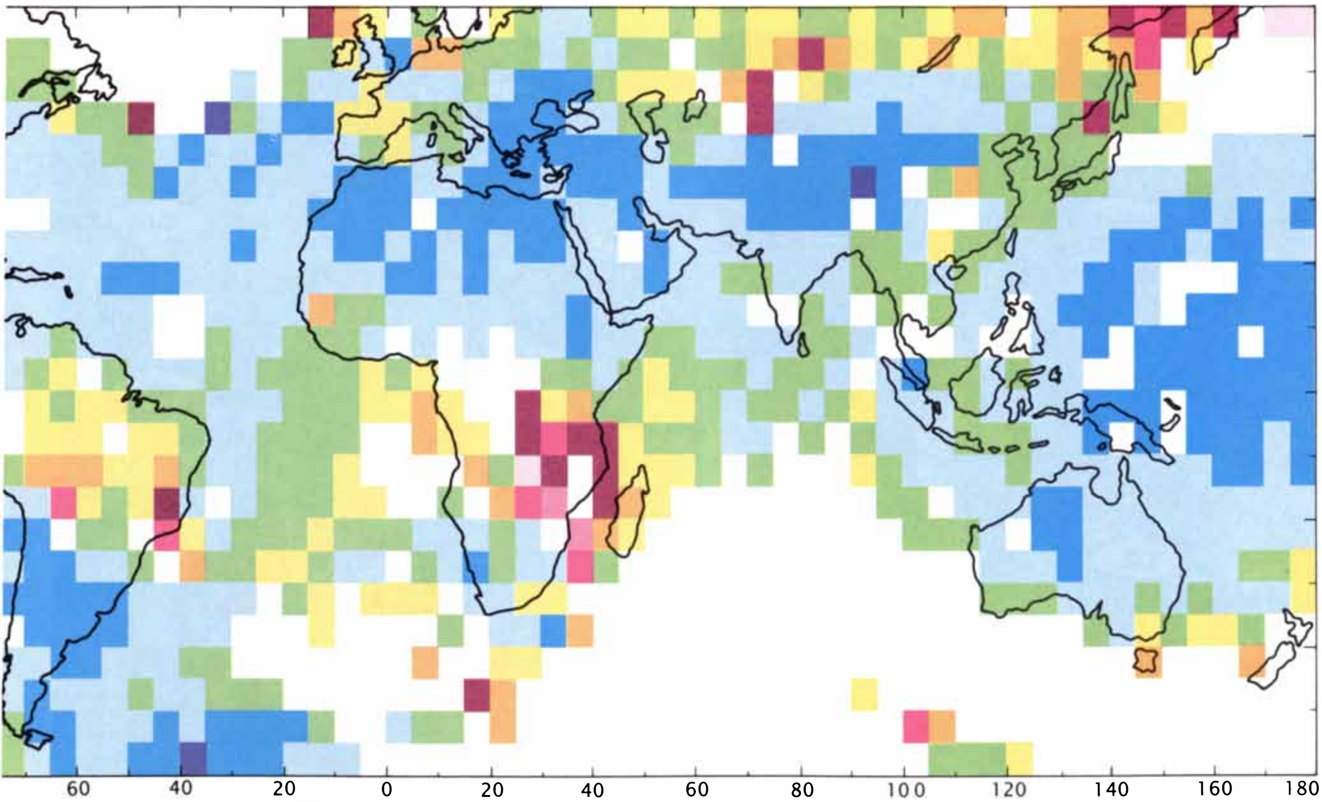
Still, the discovery of significant carbon monoxide levels in the Southern Hemisphere prompted investigations of possible sources other than burning fossil fuels. One focus of attention was atmospheric chemistry involving hydroxyl (OH), a highly reactive radical. Hydroxyl is the product of a reaction between atmospheric water molecules and excited oxygen atoms that are released when sunlight decomposes low-altitude ozone molecules. Its high reactivity makes hydroxyl the most important scavenging gas in the atmosphere, voraciously oxidizing methane and other molecules.

James C. McConnell, Michael B. McElroy and Stephen C. Wofsy of Harvard University suggested in 1971 that when hydroxyl oxidizes methane in the atmosphere, it triggers a series of reactions that yield an abundance of carbon monoxide. Methane is distributed in almost uniform concentrations throughout the atmosphere and is therefore plentiful in the Southern Hemisphere. According to the Harvard group's calculations, methane was potentially a greater source of carbon monoxide than fossil-fuel burning.

Carbon monoxide injected into the atmosphere does not stay there indefinitely. After a period ranging from about 10 days to several months, it vanishes. Some of the gas is known to settle to the earth and be absorbed by the soil. By the time of the Harvard studies, Hiram Levy II of the Smithsonian Astrophysical Observatory in Cambridge, Mass., had already shown that hydroxyl could also remove carbon monoxide from the air by converting it into carbon dioxide.

Despite these theoretical advances, the riddle of how carbon monoxide originated and disappeared could not be solved completely without knowledge of the distribution of the gas. Areas of unusually high concentration might offer clues about where carbon monoxide is coming from, and areas of low concentration might reveal where it is leaving the atmosphere. How, though, could extensive and detailed maps of carbon monoxide distribution be compiled? Measurements made solely from the ground or aircraft are impractical because of the large number required. It would take months or years for even a fleet of aircraft to gather enough readings for a detailed map, and such a map would obscure short-lived carbon monoxide patterns.

While the Harvard studies and other work were being carried out, other groups were considering possible ap-



CARBON MONOXIDE-RICH AIR from burning rain forests and savannas rises above tropical regions, according to data collected by the Measurement of Air Pollution from Satellites (MAPS) team. It is now clear that carbon monoxide is not exclusively the by-product of transportation and industry. This map is based on measurements made from the *Challenger* with infrared-sensitive equipment during October, 1984. The carbon monoxide values show ratios of the gas mixing with air at altitudes of from three to 18 kilometers; the plumes are displaced from their points of origin by winds. Each square measures five degrees on a side and is color-coded to show the average of multiple measurements taken within the region.

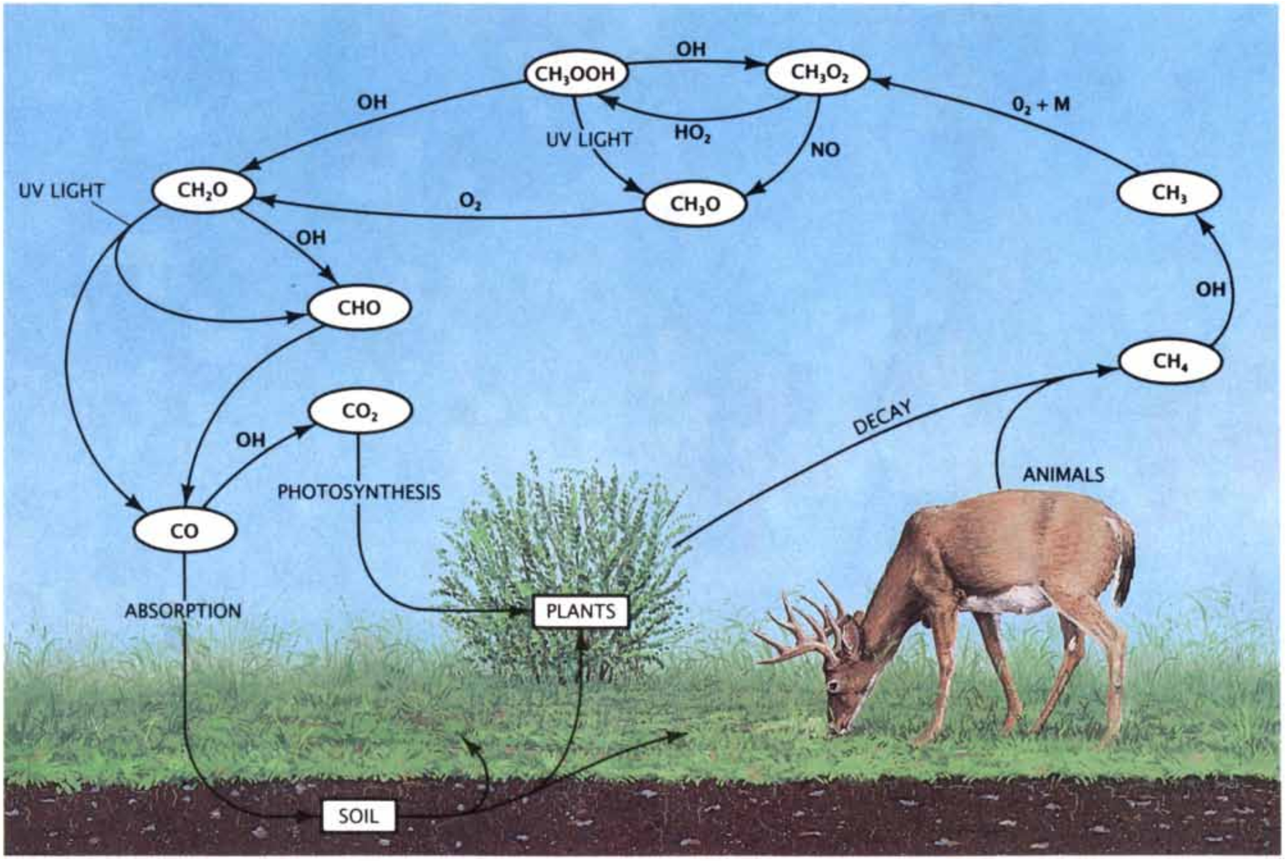
plications of earth-orbiting satellites to global carbon monoxide surveys. Orbiting instruments could compile enough measurements for a map in only a few days, revealing nearly simultaneous conditions at various latitudes and longitudes. Theoretical investigations by Claus Ludwig and his collaborators at Convair Incorporated showed that it would be possible to measure carbon monoxide from a satellite by a technique called gas-filter radiometry. This technique had already been applied by John T. Haughton's group at the University of Oxford to study atmospheric temperature with the *Nimbus IV* satellite.

Various implementations of the gas-filter technique were tested on board aircraft at the Langley Research Center of the National Aeronautics and Space Administration (NASA), and one suggested by Anthony Barringer of Barringer Research, Inc., in Toronto was adopted for development as a candidate satellite instrument. One of us (Reichle) then invited the others (Newell and Seiler) to serve as members of a science team that would guide the development of an experiment to be carried on board the space shuttle. The experiment, called Measurement of Air Pollution from Satellites (MAPS), was proposed during 1976 for one of the shuttle's orbital test flights.

The MAPS gas-filter radiometer is based on the principle that carbon monoxide absorbs infrared radiation at discrete frequencies. This distinctive absorption pattern can serve as a "fingerprint" for detecting and measuring concentrations of the gas in the atmosphere. Carbon monoxide characteristically absorbs infrared energy at wavelengths near 4.67 micrometers (millionths of a meter), although the precise pattern of absorption depends on both the temperature and the pressure of the gas.

The objective lens of the MAPS instrument is aimed toward the planet and collects radiation from the earth-atmosphere system. The radiation is then periodically "chopped" by a rotating wheel, which also introduces flashes of infrared radiation from a temperature-controlled, black aluminum plate. This plate, called a blackbody source, radiates a known spectrum that does not show losses due to absorption at any wavelength. This smooth spectrum serves as a reference for measuring the amounts of radiation absorbed by the atmosphere. The combined beam passes through a filter that screens out all wavelengths except those near 4.67 micrometers. Beam splitters divide this selected radiation and channel it to three photodetectors. One of the detectors is set behind an evacuated, transparent gas cell; it measures alternately the absolute intensity of the infrared signal from the atmosphere and that from the blackbody source. The other two detectors sit behind cells filled with carbon monoxide at different pressures. Three measurements are stored by the instrument's flight recorder: the time-stamped electronic output of the vacuum-cell detector and the two differences between that output and those of the gas-filtered detectors.

The two difference signals represent the degrees of similarity between the changing spectrums of the atmosphere and the constant signals from



CARBON MONOXIDE forms naturally in the atmosphere in a chain of reactions beginning with the oxidation of methane (CH_4) by the hydroxyl radical (OH). Hydroxyl is essential at several steps in the cycle, in both forming carbon monoxide (CO) and converting it to carbon dioxide (CO_2); carbon monoxide competes for hydroxyl with other gases. As levels of carbon monoxide rise because of rain-forest burning and other processes, carbon monoxide uses relatively more hydroxyl.

each of the pure carbon monoxide cells. When the region of the atmosphere being measured is relatively free of carbon monoxide, the differences are large; as the atmosphere becomes enriched in carbon monoxide, the differences shrink. On the basis of these measurements and the known pressures and temperatures inside the gas cells, we can infer the proportion of carbon monoxide in the air, a value termed the mixing ratio.

Because the radiation spectrum of carbon monoxide changes with pressure, each detector responds most strongly to carbon monoxide at a different altitude. The detector set behind carbon monoxide pressurized at 266 millimeters of mercury is most sensitive to gases at altitudes from three to eight kilometers, whereas the one "reading" through carbon monoxide at 76 millimeters of mercury is in effect tuned to the gas at higher altitudes. The unfiltered detector responds most to radiation from the ground. The different response curves for the detectors help us to estimate the height of the carbon monoxide measured during MAPS experiments.

To arrive at carbon monoxide mixing ratios from the detector measurements, we also needed to know about other factors that influence the radiation passing through the atmosphere. We therefore took into account meteorological conditions, the angle of the sun during our observations and estimates of the terrain's light-reflecting qualities. Information from the U.S. Navy Fleet Numerical Oceanography Center was invaluable. We developed atmospheric models that helped us to correct for the presence of water vapor, carbon dioxide, ozone and nitrous oxide, all of which also absorb energy near 4.67 micrometers.

Clouds in the instrument's field of observation can also distort measurements, and so we corrected for them. Our solution during the first shuttle flight of MAPS was to align a camera with the radiometer's sensing axis so that the earth-atmosphere system could be photographed while measurements were made. Later, our colleagues Warren D. Hypes of the Langley Research Center and Barbara B. Gormsen, who was then at the Old Dominion University Research Foundation in Norfolk, Va., inspected these photographs painstakingly for indications of cloud cover and deleted from the MAPS data set any carbon monoxide measurements that had been made through clouds.

The MAPS instrument needed to be tested for sensitivity at both high and low mixing ratios of carbon monoxide in air. For a high mixing-ratio test, we mounted the radiometer on an aircraft and flew it over Lake Michigan to monitor carbon monoxide produced by morning rush-hour traffic in Chicago. Plumes of carbon monoxide from both Chicago and Milwaukee were detected. The Chicago plume showed carbon monoxide levels, averaged over altitude, of 260 molecules per billion air molecules, a high value but not an unexpected one.

It was during what was supposed to be the low mixing-ratio test that MAPS first measured high carbon monoxide levels in remote, undeveloped areas. In the summer of 1979, MAPS was included in an international project, called MONEX, studying the Indian monsoon. On board a NASA Convair 990 aircraft, the MAPS radiometer took readings during long flight tracks across the Arabian Sea at an altitude of roughly 12 kilometers. Its readings were verified by analyses of bottled air samples collected in flight by Estelle P. Condon, who was then at the Old Dominion University.

To our astonishment, these flights detected even higher concentrations of carbon monoxide in the boundary layers over Saudi Arabia and the Ganges Valley in India than were found over Chicago at rush hour. As measured by both MAPS and the bottled air samples, mixing ratios in these regions were over 300 carbon monoxide molecules per billion air molecules. Over the Arabian Sea, much lower concentrations of roughly 80 molecules per billion air molecules were found near the equator, where air from the Southern Hemisphere was entering the monsoon circulation.

As work on MAPS progressed, other evidence continued to accumulate against the old view associating carbon monoxide almost exclusively with developed areas. During the dry season in August and September, 1980, one of us (Seiler) and investigators from the National Center for Atmospheric Research in Boulder, Colo., participated in a study of carbon monoxide and various other gases over Brazil. Aircraft measurements found boundary layer concentrations of up to 400 molecules per billion air molecules over untouched tropical rain forests. Even higher values, some going off the scale of the measuring device, were found near a Brazilian savanna that was being burned.

New theories soon emerged to explain how forests could contribute carbon monoxide to the atmosphere. Inspired by the Brazilian data, Paul J. Crutzen of the Max Planck Institute [see "The Changing Atmosphere," by Thomas E. Graedel and Paul J. Crutzen; SCIENTIFIC AMERICAN, September] suggested that large quantities of carbon monoxide could be generated over completely untouched rain forests through the photochemical oxidation of hydrocarbons other than methane. These hydrocarbons would arise from the resins and oils that are released primarily by trees. Similar conclusions have recently been drawn by Alain Marenco of the Center of Atomic Physics at Toulouse and Jean Claude Delauny of the Atmospheric Physics Laboratory at Abidjan, Ivory Coast, from data obtained over tropical forests in Africa.

Crutzen, working with Seiler, also proposed that burning biomass, such as vegetation being cleared from the land or animal dung being burned as fuel, is a major source of atmospheric carbon. Only some of this carbon takes the form of carbon monoxide; much more becomes carbon dioxide and particulate carbon. By Seiler and Crutzen's calculations, burning biomass annually contributes from two to four billion metric tons of carbon to the atmosphere. Work done by Helene Cachier and her colleagues at the Centre de Faibles Radioactivités in Gif-sur-Yvette has also shown that carbonaceous aerosols from tropical forests add as much fine particulate carbon to the atmosphere as all industrial sources do. Many of these aerosols are produced during the dry season, when most natural and anthropogenic burning occurs.

Against this backdrop of surprising new data and novel theories, the MAPS radiometer flew into space on board the shuttle's second engineering test flight in November, 1981. Because of failures in the shuttle's power and cooling systems, only 11 hours of usable data were collected over two days, representing about 10,000 observations of carbon monoxide at altitudes between three and 12 kilometers.

The observed areas lay in the tropical band between 37 degrees north and 37 degrees south latitude. From the shuttle's altitude of 260 kilometers, MAPS could record the radiation from successive areas of the earth's surface 20 kilometers across. These monitored areas were divided into grid squares measuring five degrees of latitude by five degrees of longitude. We smoothed out small-scale variations in the data by averaging multiple measurements made within each grid square.

When we analyzed the MAPS data in this way, we were startled by the picture of global carbon monoxide distribution that emerged. Least surprising was the observation that the lowest concentrations, around 40 molecules per billion air molecules, came from over the southeast Pacific and Argentina, where winds blow from the west over large tracts of ocean. Since previous studies had discounted the open oceans as major contributors to atmospheric carbon monoxide, these findings agreed with expectations.

We noted higher carbon monoxide values, around 75 molecules per billion air molecules, over the eastern Mediterranean and adjoining landmasses. The same air had previously passed over Western Europe during a time of strong convection: the burning of fossil fuels was presumably responsible for these high readings.

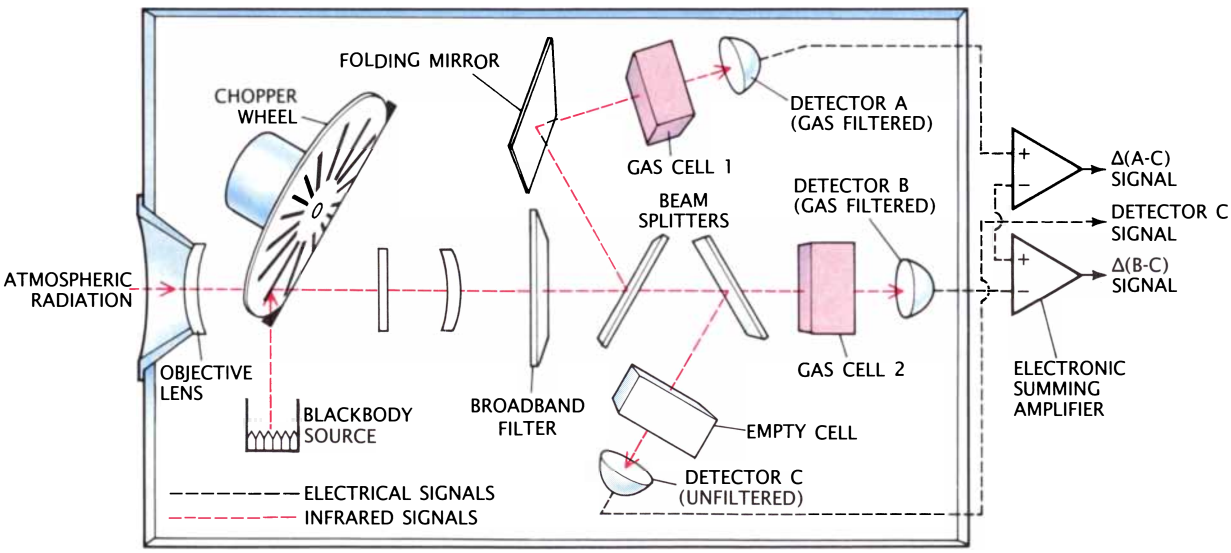
The big surprise, however, was that the highest carbon monoxide readings had been recorded over regions with little or no industry or automobile traffic, many of which lay in the Southern Hemisphere or tropics. Northern South America, central Africa and eastern China all showed readings greater than 100 molecules per billion air molecules. (The highest measured value was found over the Gulf of Guinea along equatorial Africa's western coast, but this may have been a statistical fluke reflecting the small number of usable readings.)

Where was the carbon monoxide over these undeveloped areas coming from? We looked at wind velocity maps and convection reports for the observed regions during November. It became clear that the carbon monoxide-laden air flowing 10 to 12 kilometers above South America and the equatorial Atlantic Ocean had originated in the boundary layer over tropical rain forests. The air over China had blown over rain forests in northwest Burma the day before the sampling was taken. The terrain immediately beneath the measured central African boundary layer was grassland and savanna, and there was rain forest within 500 kilometers.

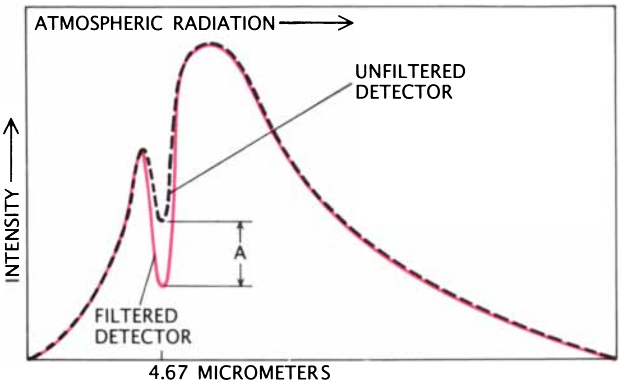
Clearly, something other than industry had to be producing clouds of carbon monoxide in these undeveloped areas. The proximity of rain forests seemed to be a common element; burning vegetation in the savanna might also be involved. In other words, the theories proposed by Crutzen, Seiler and others could account for the data collected by MAPS on its first shuttle flight. We hoped that measurements from a second shuttle flight would confirm and improve on the initial findings.

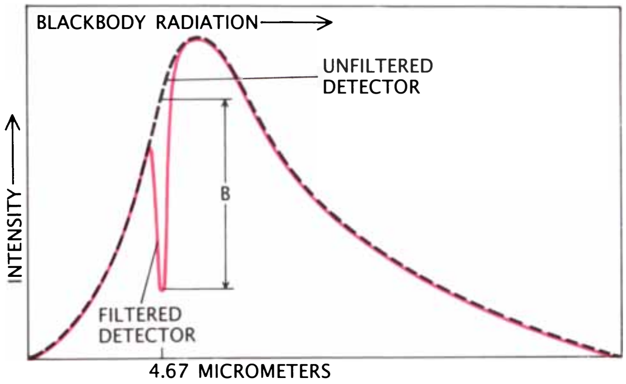
Before the second shuttle experiment, the MAPS instrument was modified to provide a simple system for cloud detection. Now, only one gas cell was filled with carbon monoxide. This change slightly limited our ability to determine the altitudes of carbon monoxide plumes but was acceptable to us. The other cell was filled with nitrous oxide, which mixes with air at an almost constant ratio of 305 molecules per billion air molecules throughout the lowest 12 kilometers of the atmosphere. Like carbon monoxide, nitrous oxide absorbs energy near 4.67 micrometers. With this new configuration, we could measure the mixing ratios of both carbon monoxide and nitrous oxide. Since the mixing ratio of nitrous oxide is nearly constant, any apparent fluctuations that the instrument observed had to be caused by obscuring clouds moving into view. By looking for such fluctuations, we could immediately recognize and discard the accompanying carbon monoxide data, which would also have been distorted. This change in the setup of the instrument eliminated the tedious step of examining photographs for clouds.

The second shuttle flight of MAPS took place in October, 1984. It had originally been scheduled for the early spring of that year, but delays in the shuttle launch schedule forced a postponement. This was unfortunate, because we had hoped that measurements made during the spring would complement the data from November, 1981, and give some



INFRARED RADIATION from the atmosphere enters the MAPS radiometer and is "chopped" by a spinning, slotted wheel that introduces flashes of radiation from a warm blackbody reference. An optical filter screens out all wavelengths except those around 4.67 micrometers, which is the peak absorption band of carbon monoxide. Mirrors direct the beam of selected radiation toward three detectors. The unfiltered detector sits behind an evacuated gas cell and measures the absolute intensity of the radiation. The two other detectors are located behind cells that are filled with carbon monoxide. The output of the unfiltered detector and the differences between its output and those of the gas-filtered detectors are recorded for later study.





MIXING RATIOS of carbon monoxide in the atmosphere can be calculated by comparing the atmospheric and reference spectrums observed by the detectors. A detector equipped with a gas-cell filter of carbon monoxide gives a low, nearly constant signal because the gas cell strongly absorbs almost all the available 4.67-micrometer radiation from any source. At that wavelength an unfiltered detector reports either partial absorption of radiation (when reading the atmosphere) or no absorption (when reading the blackbody reference). The difference *A*, taken when both detectors are observing the atmosphere, is mathematically related to the difference between the concentrations of carbon monoxide in the atmosphere and inside the cell. Difference *B*, which compares the blackbody emission with the cell's absorption, helps to calibrate the system.

indication of carbon monoxide variations throughout the year. Seiler and his colleagues had previously monitored the air from surface stations in both hemispheres and found marked seasonal variation, with peak carbon monoxide levels in the local spring.

In other respects the execution of the second MAPS shuttle flight improved on the first. Because the shuttle's orbit on this trip passed over a greater range of latitudes—from 57 degrees north to 57 degrees south—the observations covered a more extensive geographic range. Over a period of nine days, 86 hours of measurements were collected. These were used to make two maps, each showing carbon monoxide concentration patterns averaged over consecutive four- and five-day periods.

The patterns of carbon monoxide distribution recorded on the October, 1984, flight resembled the one seen in November, 1981. Values exceeding 100 carbon monoxide molecules per billion air molecules were measured over South America, southern Africa, Europe, the U.S.S.R., China, the northern Pacific and the southern Indian Ocean. The lowest values were over the tropical Pacific, the north Atlantic, the Sahara and Argentina.

Photographs taken from space by Kathryn D. Sullivan, a NASA astronaut, confirmed an association between large fires visible from orbit and plumes rich in carbon monoxide. For example, smoke from fires near the mouth of the Zambezi River in Africa was seen blowing inland, carried by an easterly wind. Convection carried this smoke to altitudes of from five to 10 kilometers, where the MAPS radiometer could detect the carbon monoxide it contained.

To calibrate the 1984 MAPS data, Seiler's research group made extensive measurements from an aircraft flying over the Atlantic Ocean during the shuttle flight. Because the instrument that made these measurements had direct contact with the air, it was sensitive to mixing ratios lower than one carbon monoxide molecule per billion air molecules. The measurements were taken at an altitude of about 10 kilometers during flights from Frankfurt to São Paulo and back.

North-south distribution patterns for carbon monoxide measured almost simultaneously by MAPS and the aircraft corresponded extremely well. The aircraft measurements were consistently about 40 percent higher than the MAPS readings; the discrepancy is currently under study.

All the MAPS and aircraft measurements strongly suggest that at least during the Northern Hemisphere autumn and the Southern Hemisphere spring, when the measurements were made, rain forests and savannas rival or surpass fossil fuels as sources of atmospheric carbon monoxide. Both burning and hydrocarbon oxidation contribute to this output. Despite similarities in the highest values recorded around the world, several distinct mechanisms of carbon monoxide generation appear to be at work.

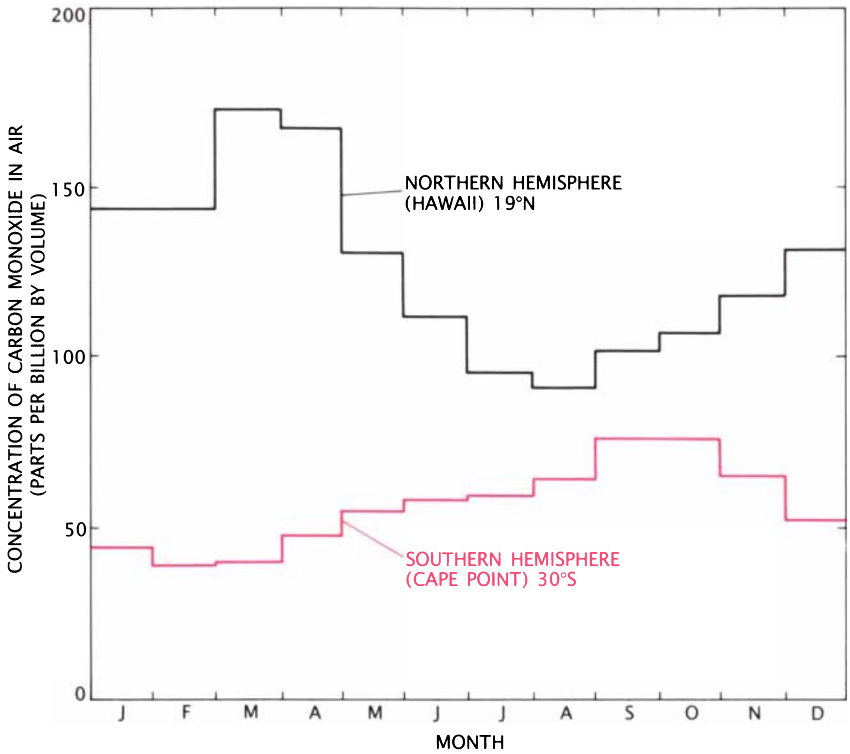
The relative importance of the various mechanisms adding carbon monoxide to the atmosphere depends on location. Oxidation of methane and other hydrocarbons is a major contributor in both hemispheres; releases of the gas from biochemical processes in the soil and vegetation are relatively minor. In general, combustion of fossil fuels is the major source of carbon monoxide in the developed Northern Hemisphere, whereas burning of biomass predominates in the Southern Hemisphere and the tropics.

The amount of carbon monoxide introduced by burning vegetation, much of which results from human activities in developing nations, raises troubling questions. How does the rate at which forests lose carbon through burning compare with the rates at which carbon is incorporated into growing trees by photosynthesis and respiration? What are the environmental consequences?

People in the developing nations rely heavily on wood for fuel; they also burn it to clear land for cattle ranges, agriculture and new settlements. Moreover, the rate of burning is rising because of economic pressures and urgent fuel shortages.

Working from statistical data, Seiler and Crutzen have calculated that from .5 to .75 percent of the tropical forests is lost annually to burning. As the productive base of forestland decreases and burning rates increase, the percentages of forest lost skyrocket. Some logged areas will be allowed to grow back, but forests cleared for agriculture are lost forever.

The disappearance of tropical rain forests, or even substantial decreases in their size, could dramatically alter the climate by changing global patterns of evaporation and heat circula-



SEASONAL VARIATIONS in atmospheric carbon monoxide levels have been measured in both hemispheres. The Northern Hemisphere measurements were made at the Mauna Loa Observatory in Hawaii, the Southern Hemisphere measurements at a Cape Point ground station on the Cape of Good Hope in South Africa. Each monthly average value is based on a continuous carbon monoxide record covering a period of at least five years. In both hemispheres the carbon monoxide mixing ratio reached a maximum in or near the local spring season, possibly because trees were releasing large quantities of hydrocarbons that were then oxidized into carbon monoxide.

tion. Trees return large quantities of moisture to the atmosphere by releasing water vapor from their leaves during the day. When trees are removed, rain runoff increases and less water evaporates from the soil into the air. The evaporation of moisture also moderates surface temperatures by absorbing solar energy that would otherwise heat the soil. As a result of trees disappearing, temperatures would become more extreme.

Nor would all the effects be local. Evaporation is also important because it efficiently conducts solar energy into the atmosphere at high altitudes. Without evaporation, heat from the warm earth is conducted directly into the surface layers of the atmosphere. Water vapor, however, usually rises to heights of from two to eight kilometers before releasing its latent heat by condensing as rain. The injection of heat at these altitudes drives global patterns of weather and air circulation. The consequences of altering these patterns are difficult to predict.

It is also difficult to predict how the large amounts of carbon monoxide produced by biomass burning may change the climate by altering atmospheric chemistry. Hydroxyl reacts readily with carbon monoxide. As carbon monoxide emissions climb, proportionally more of the hydroxyl is consumed by carbon monoxide, and less remains to break down methane and other molecules. This disturbance of the atmosphere's chemical equilibrium may help explain increases in atmospheric methane that have been observed in recent years. Methane, like carbon dioxide, is a greenhouse gas. In the atmosphere, it helps to trap heat that the earth would ordinarily emit back into space. Much of the growing concern that an increasing greenhouse effect might change the global climate has centered on huge increases in atmospheric carbon dioxide; whether increases in carbon monoxide will result in enough methane to seriously compound the problem remains to be seen.

High levels of carbon monoxide also increase the formation of ozone at relatively low altitudes. It appears that this low-altitude ozone is effective at screening harmful solar ultraviolet radiation. It might therefore be helpful in offsetting the effects of decreases in the stratospheric ozone layer, which seems to be diminishing because of reactions with chlorofluorocarbon contaminants. Yet even fairly small increases in ozone concentrations can harm the growth of vegetation. The effects of ozone exposure may already be visible in some hardwood forests that appear to be suffering from slower growth rates.

It is clear, then, that measuring the production, circulation and elimination of carbon monoxide is a matter of vital interest. The MAPS instrument has already proved to be a useful tool for detecting carbon monoxide from burning biomass, fossil fuels and natural sources at high altitudes. Repeating the MAPS experiment during seasons other than the Northern Hemisphere autumn would help to increase understanding of seasonal variations in carbon monoxide throughout both hemispheres. Additional measurements could also fill in gaps in geographic coverage, particularly over the oceans. Further improvements in MAPS could increase its sensitivity to carbon monoxide in the lowest layers of the atmosphere, which would permit more direct and accurate measurements of burning and local carbon monoxide generation.

FURTHER READING

THE CYCLE OF ATMOSPHERIC CO. W. Seiler in *Tellus*, Vol. 26, No. 116, pages 118–135; 1974.

ESTIMATES OF GROSS AND NET FLUXES OF CARBON BETWEEN THE BIOSPHERE AND THE ATMOSPHERE FROM BIOMASS BURNING. W. Seiler and P. J. Crutzen in *Climatic Change*, Vol. 2. D. Reidel Publishing Co., 1980.

TROPOSPHERIC CHEMISTRY: A GLOBAL PERSPECTIVE. J. A. Logan, M. J. Prather, S. C. Wofsy and M. B. McElroy in *Journal of Geophysical Research*, Vol. 86, No. C8, pages 7210–7254; August 20, 1981.

MIDDLE AND UPPER TROPOSPHERIC CARBON MONOXIDE MIXING RATIOS AS MEASURED BY A SATELLITE-BORNE REMOTE SENSOR DURING NOVEMBER 1981. H. G. Reichle, Jr., V. S. Connors, J. A. Holland, W. D. Hypes, H. A. Wallio, J. C. Casas, B. B. Gormsen, M. S. Saylor and W. D. Hesketh in *Journal of Geophysical Research*, Vol. 91, No. C9, pages 10865–10887; September 20, 1986.

THE DISTRIBUTION OF MIDDLE TROPOSPHERIC CARBON MONOXIDE DURING EARLY OCTOBER 1984. H. G. Reichle, Jr., J. C. Casas, E. P. Condon, V. S. Connors, B. B. Gormsen, J. A. Holland, W. Seiler, R. T. Sherrill and H. A. Wallio in *Journal of Geophysical Research*, in press.

Waterweed Invasions

Vast vegetative mats of the two most noxious aquatic weeds plague the world's waterways. Investigations of the water hyacinth and the kariba weed are leading to new programs for weed control

by Spencer C. H. Barrett

Populations of plants and animals ordinarily migrate and multiply slowly over time as a result of the natural interplay among many ecological factors: soils, waters, glaciation, drought, the rise and fall of competitor species. Yet in some regions human activities have quickened the pace of change by dramatically altering the range and abundance of many species. Today plants and animals are shipped, sometimes accidentally, all over the globe to zoos and botanical gardens and for commercial and agricultural purposes. A small minority of the "alien," or introduced, species become ecological as well as economic disasters; although they may have been innocuous in their native region, these species are transformed into aggressive pests or weeds that invade and dominate their new environment.

Nowhere are these biological invasions more evident than in the rivers, lakes and reservoirs of the world. In the past century more than a dozen weed species have laid siege to the world's waterways. Canadian species, such as the pondweed *Elodea canadensis,* have infested canals in Europe. In return native European weeds, such as the Eurasian water milfoil, *Myriophyllum spicatum,* have overgrown lakes in Canada. Such native tropical species as the alligator weed, *Alternanthera philoxeroides,* have clogged irrigation systems in the U.S.; to even the score, such native American weeds as the grass *Echinochloa microstachya* have invaded rice fields in Australia.

Two species of aquatic weeds exemplify the problem such pests can create: the water hyacinth, *Eichhornia crassipes,* and the kariba weed, *Salvinia molesta.* These two species have wreaked havoc on waterways throughout the world, particularly in the tropics and subtropics, where they cause severe hardship and immense economic difficulties. In a single growing season the weeds can reduce a thriving water community to a destructive mass, halting transportation, killing fish and promoting disease.

In an attempt to eradicate the water hyacinth and the kariba weed, farmers, biologists and government officials have launched valiant weed-control programs and have spent millions of dollars on mechanical and chemical remedies. Unfortunately, most mechanical methods cannot destroy plants fast enough, and herbicide controls have harmful side effects on water quality, fish stocks and other elements of the aquatic food chain. However, new programs developed from careful studies of these noxious weeds offer hope.

The water hyacinth and the kariba weed share a feature common to most weeds: the ability to grow and multiply rapidly in habitats that are disturbed by human activity. In recent times the large-scale disruption of natural ecosystems has opened up many new ecological niches for these aquatic weeds. Irrigation schemes, hydroelectric projects and artificial lakes establish ideal environments. In natural waterways the plants prosper on a steadily replenished supply of nutrients that agricultural activities provide by the runoff of fertilizers and the leaching of minerals from soil.

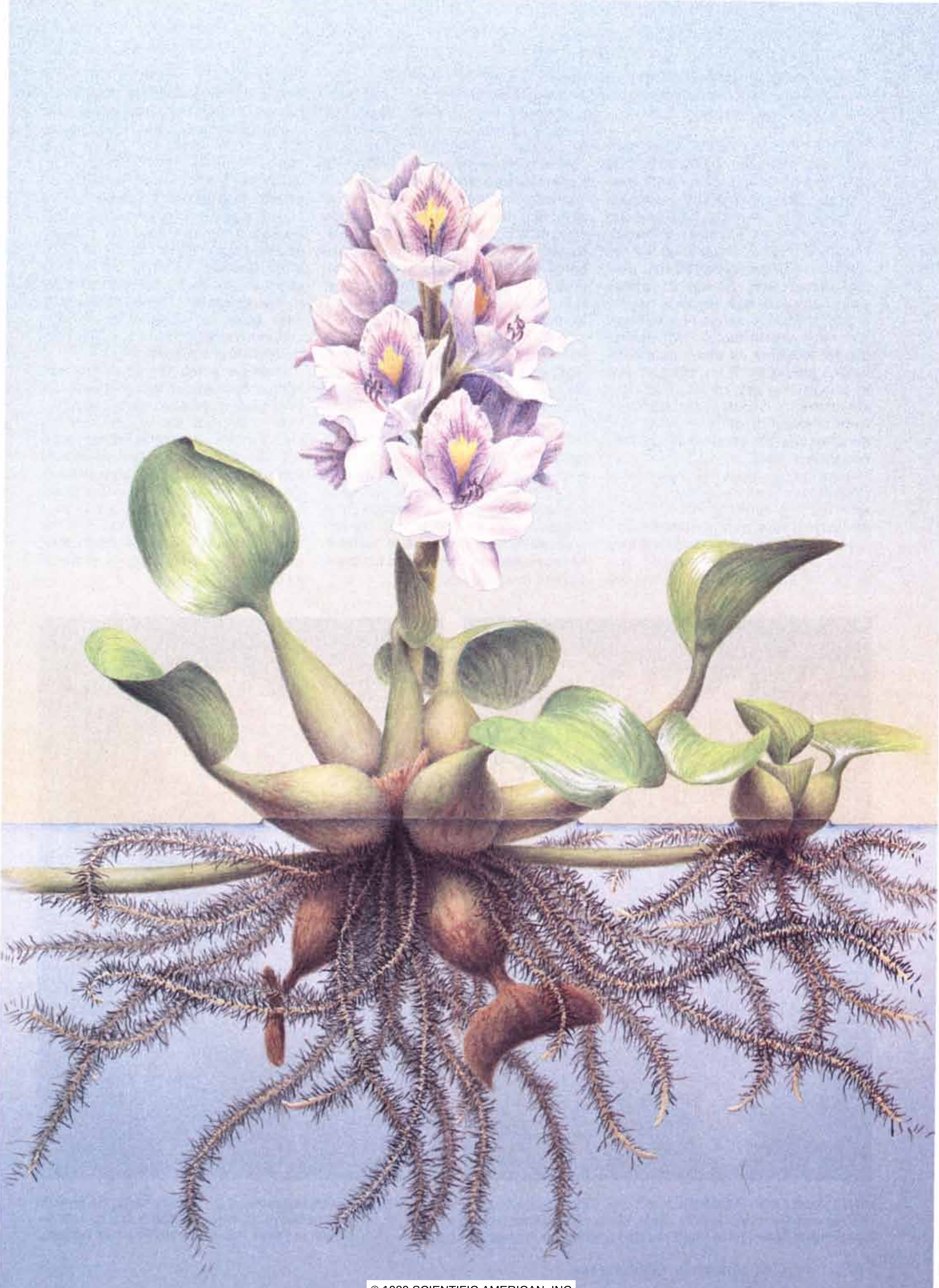
Two special characteristics, high mobility and clonal propagation, have allowed the water hyacinth and the kariba weed to reign over this rich aquatic domain. Their high mobility is made possible by air-filled tissue known as aerenchyma, which gives the stems and leaves of the plants considerable buoyancy. The weeds can therefore float with wind or water currents to unoccupied waters, where they can grow and regenerate.

Clonal propagation helps the weeds to grow rapidly over large areas. Recently the word "cloning" has become a familiar term because of popular interest in molecular biology and genetics. Cloning has similar connotations in botany. Cloning in the botanical sense is the propagation of genetically identical plants by asexual reproduction from some sexually produced ancestor. A clone is therefore a plant produced without fertilization of male and female gametes; it represents an identical genetic copy of the parent plant. This kind of cloning will be familiar to anyone who has grown ornamental plants from cuttings.

The water hyacinth and the kariba weed display a particularly fascinating method of clonal propagation: the plant breaks apart into many separate pieces, each having the potential to grow into a complete organism. As wind or water currents disperse the fragments, colonies can expand rapidly over vast open stretches on the water surface. Freed from competition with other plants and guaranteed almost unlimited space, nutrients and sunlight, the water hyacinth and the kariba weed grow and multiply at an extraordinary pace to achieve some of the highest rates of biomass production recorded in the plant world. Consequently, a single genetic individual

SPENCER C. H. BARRETT is professor of botany at the University of Toronto. His research specialities are the evolutionary genetics of weeds and plant reproductive systems. In 1977 he received his Ph.D. from the University of California, Berkeley. Barrett was recently awarded the E.W.R. Steacie Memorial Fellowship, which will enable him to devote his attention to research on the genetics and evolution of weeds. His first article for *Scientific American* appeared in the September, 1987, issue.

WATER HYACINTH, exported from its native South American region for its striking beauty, now grows uncontrollably over the world's rivers and lakes. In four months two plants can yield 1,200 offspring. The water hyacinth's growth has devastated many aquatic communities.



may be essentially immortal, covering vast areas and experiencing many different environments.

The water hyacinth is usually singled out as the aquatic weed that is most troublesome both ecologically and economically. It invades waterways at rates that have become legendary among biologists and weed-control experts. A group studying the weed in Louisiana reported that during one growing season 25 plants can produce enough biomass to cover 10,000 square meters of water surface with approximately two million plants, weighing as much as a fully loaded jumbo jet. If the plant settles in waters that are enclosed or have slow currents, colonies can coalesce to form continuous mats of living and decaying organic material as much as two meters thick.

Great green mats of water hyacinth fill reservoirs, spoiling water resources; they infest rivers, impeding navigation; they dam drainage channels, flooding lowlands; and they clog pipes, disturbing hydroelectric systems. The mats indirectly deplete the water's supply of dissolved oxygen, thereby asphyxiating fish and phytoplankton. As the weed drives fish away, it jeopardizes human nutrition in riverine communities where fish are the primary source of protein. The water hyacinth also provides an excellent microhabitat for agents of several human diseases, including malaria, encephalitis and schistosomiasis. Although the water hyacinth rarely competes with agricultural crops, it impedes the flow of water through irrigation canals and pumps and thereby hinders crop production.

The water hyacinth has spread from its native region—the tropical lowlands of South America—to more than 50 countries on five continents. In 1824 Karl Friedrich Philipp von Martius found the plant in Brazil and formally described it as *Pontederia crassipes;* later it was found to belong to the tropical genus *Eichhornia.* For the next six decades, however, the water hyacinth received little attention from botanists; apparently it was considered nothing more than a well-behaved plant. Its show of good behavior did not last for long.

Although the spread of the water hyacinth is difficult to document with complete accuracy, it appears that popular fascination with the plant began in 1884. In that year water hyacinths imported from the lower Orinoco River in Venezuela were distributed as gifts by a Japanese delegation at a cotton exposition in New Orleans. Water hyacinths have beautiful clusters of violet and yellow flowers perched atop floating rosettes of bulbous green leaves. They proved irresistible to the delegates. The botanical gifts were taken to surrounding districts and cultivated in garden ponds. They multiplied at a prodigious rate.

From the ponds the water hyacinth spread throughout the southern U.S. The plant's growth soon restricted river transportation of commodities such as corn, cotton and lumber, causing those industries to lose millions of dollars. Particularly troublesome was the 1895 invasion of the Saint Johns River in Florida: gale-force winds blew the water hyacinths up and down the river for more than 160 kilometers, creating huge floating mats as much as 40 kilometers long.

REMARKABLE SIMILARITIES between the water hyacinth, *Eichhornia crassipes* (*left*), and its docile relative, *E. azurea* (*right*), have helped botanists to single out the water hyacinth's aggressive traits. Although the roots of *E. azurea* restrict its growth to the edges of waterways, *E. crassipes* floats freely, and so the hyacinth can spread over vast stretches of water surface.

Word of the water hyacinth's beauty apparently traveled faster to Southeast Asia than news of its destructive powers. In 1894 the caretakers of the Bogor Botanical Garden in Java reported that the water hyacinth had become such a nuisance that specimens were routinely discarded into a river flowing through the gardens. Many local infestations soon followed. Today mats of water hyacinth can be found all over Southeast Asia as well as in the warm lowlands of India, Sri Lanka, China and Japan.

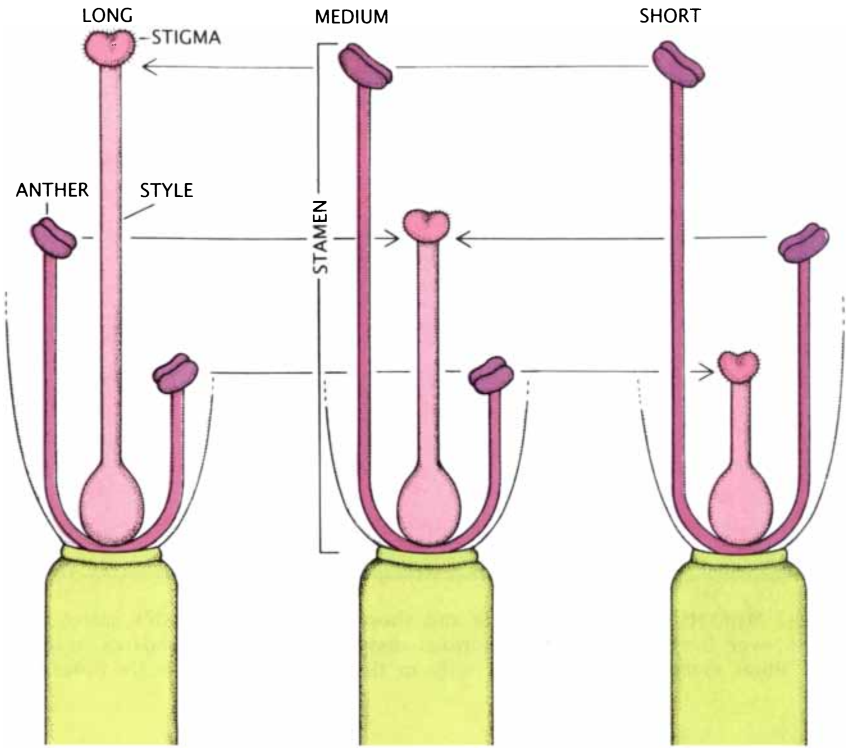
The water hyacinth has spread just as rapidly through the African continent. Boat traffic along the Congo and Nile rivers and their tributaries seems to have promoted the invasion: plants that became attached to paddle-wheel steamers sometimes hitchhiked 1,600 kilometers upstream.

As one might expect, most research on *Eichhornia crassipes* has been aimed at understanding its growth and devising ways to control it. Virtually all of this work has been done in the nonnative range of the species, particularly in the southern U.S., India and Southeast Asia.

More than a decade ago I began an investigation of *Eichhornia* in its native range. I wanted to understand more about the population biology of the remarkable *E. crassipes* and to learn something of the behavior of its little-known relatives. These studies have clarified several misconceptions about the reproductive biology of the water hyacinth and in addition have provided ecological explanations for many of its unusual reproductive characteristics.

The genus *Eichhornia* belongs to the monocotyledonous family Pontederiaceae, which includes the North American pickerelweed, *Pontederia cordata*. *E. crassipes* is one of eight species of freshwater plants in *Eichhornia*. All are native to the tropics of Central and South America, except for the African species, *E. natans*. Most species of *Eichhornia* are distributed widely throughout their native regions and regenerate by cloning. Yet *E. crassipes* is the only member of *Eichhornia* that has shown any tendency to become a noxious, aggressive weed.

That fact is particularly puzzling when one considers that the morphology of *E. crassipes* is quite similar to that of *E. azurea*. Both species form floating mats and produce large, showy flowers. More significant, *E. azurea*, like *E. crassipes*, has been exported from South America to decorate ponds and has occasionally escaped into local aquatic environments. Yet *E. azurea* has never become a serious weed problem. What makes *E. azurea* just another ornamental pondweed and *E. crassipes* the world's most aggressive aquatic weed?



CHARLES DARWIN studied the reproductive organs of morphs, or forms, of *E. crassipes* that have long and medium styles and deduced the existence of a short-style morph. He observed that the anthers of the long stamens of the medium-style morph corresponded to the stigma of the long style and that the anthers of the medium stamens of the long-style morph corresponded to the stigma of the medium style. However, the short stamens of the long- and medium-style morphs did not have sexual partners. Darwin therefore predicted that there must be a third floral morph that had not yet been discovered. The author found the short-style morph in 1974.

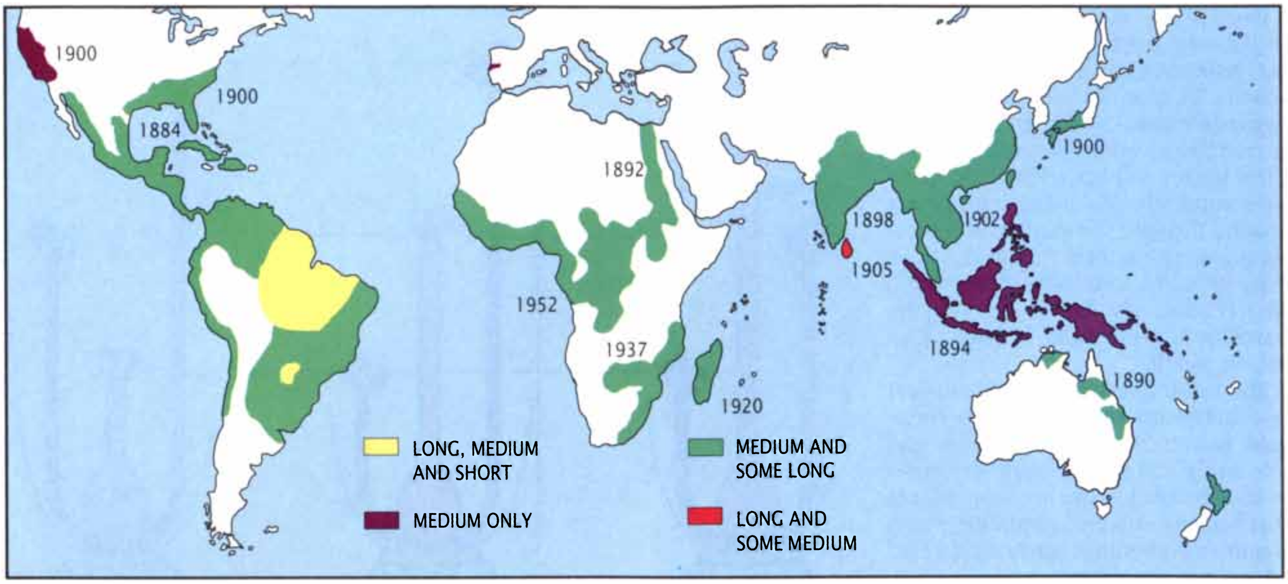
The answer lies primarily in their differing abilities to fragment into pieces that develop into whole individuals. *E. crassipes* breaks apart more readily because its rosettes of floating leaves are held together only by delicate horizontal stems called stolons. *E. azurea*, on the other hand, regenerates more slowly; its colonies cannot grow rapidly unless the plants are rooted firmly to the mud bottom. The roots of *E. azurea* restrict its distribution to shallow ponds and the edges of lakes and rivers. In contrast, because *E. crassipes* floats freely, it can grow and multiply on the surface of deep waters, away from most competitors.

Why has *E. crassipes* evolved the ability to float freely and fragment rapidly? The answer lies in the ecological conditions and habitats that *E. crassipes* occupies in its native region.

The two areas where the water hyacinth is thought to have originated are the Amazon basin and the extensive lakes and marshes of the Pantanal region in western Brazil. The two regions provide a dynamic aquatic habitat. Water levels of local lakes and rivers fluctuate dramatically because of seasonal changes in rainfall. The waters of the Amazon River, for example, rise and fall about 10 meters annually, even as far as 2,000 kilometers upstream from the Atlantic Ocean. Under these conditions the free-floating habit is highly adaptive, whereas rooted plants often perish during periods of submersion in deep, muddy water.

The Amazon basin and the Pantanal also contain many shallow, interconnected, nutrient-rich lakes and pools created by annual floods. These seasonal lakes provide ideal conditions for the explosive growth of water hyacinth. If a small colony of *E. crassipes* finds itself stranded in a lake after a flood, it proliferates by absorbing the abundant resources.

Ecological studies in the native region led me to another interesting finding. *E. crassipes* produces a great



FLORAL MORPHS with long, medium and short styles have spread over five continents in a curious distribution. The three floral morphs grow together only in the water hyacinth's native region, the lowland tropics of South America. The dates on the map show when, according to historical records, the water hyacinth was introduced to a particular region.

number of seeds that can survive dry spells. The seeds help to regenerate populations after the desiccation of colonies. This observation and others laid a common misconception to rest. Many investigators assumed that clones of the water hyacinth were sexually sterile and could not regenerate from seed. This assumption was largely based on two generalizations. First, plants that grow exclusively by vegetative propagation over long periods (such as sweet potato, sugar cane and many ornamental plants) often lose their ability to reproduce sexually: genetic mutations that impair pollen and seed fertility tend to accumulate over time. (Mutations of this type are continuously eliminated from the gene pool of species that reproduce sexually in regular cycles.)

The second generalization that suggested sterility was based on misconceptions about the water hyacinth's breeding system. Flowers of *E. crassipes* can be divided into three sexual types that differ in the length and position of their reproductive organs, the male stamens and the female pistil. These types, or floral morphs, are distinguished by their long, medium or short styles—prolongations of the ovary. Hence, *E. crassipes* is described as tristylous.

Tristylous plants are usually self-incompatible and intramorph-incompatible. In other words, very few seeds are produced as a result of self-pollination and pollination by other plants of the same morph. (Crosses between flowers of different morphs yield many seeds.) The theory that heterostylous flowers are self-incompatible and intramorph-incompatible came from Charles Darwin. He first investigated the floral morphology and breeding relationships of heterostylous plants. In 1877 he published his findings in *Different Forms of Flowers on Plants of the Same Species*. Because of the self-incompatibility of the water hyacinth, many botanists believed it would produce few seeds in regions where one style morph grows, as is the case in most areas.

Yet my own experiments indicate that high levels of seed fertility are achieved in many single-morph colonies of *E. crassipes*. I found that individual clones are self-fertile and intramorph-fertile. In fact, most clones can produce thousands of viable seeds.

The failure to recognize that many seeds are often produced in water hyacinth populations has complicated weed-control efforts. One technique, still employed in various parts of the world, involves the drainage of water from infested canals and reservoirs during certain periods of the year. This practice, known in the U.S. as drawdown, destroys the vegetative parts of aquatic plants through desiccation. Yet drawdowns also provide excellent opportunities for weeds to germinate and establish seedlings: they remove the leaves of floating mats that usually shade seedlings, and they establish moist sediments, in essence mimicking the water-level fluctuations of the Amazonian habitat.

Although the water hyacinth was an exception to Darwin's findings that heterostylous plants are usually self-incompatible, he proved correct about other features of the plant. Darwin received dried flowering specimens of *E. crassipes* from southern Brazil. He identified the flowers as examples of the long- and medium-style morphs. He then deduced that there should also be a short-style morph, because both the long- and medium-style morphs had short stamens. Darwin's deduction later sparked considerable controversy, but no definitive evidence was found to prove that the short-style morph existed.

In the 1950's, in an effort to find the missing form, the geneticist J.B.S. Haldane enlisted the help of schoolchildren in India to survey populations of water hyacinth. This effort and many others failed. By the early 1970's botanists had reached the conclusion that the short-style morph was extinct and that *E. crassipes* had two floral forms, not three.

I was aware of this controversy in 1974, when I began working in the lower Amazon basin of Brazil. There I encountered my first flowering colony of water hyacinth growing in marshes associated with the river Jari. The flowers had short styles! Darwin's deduction was confirmed.

Later I conducted a more extensive geographic survey of the morphs of water hyacinth in North and South America with Wendy Forno of the Commonwealth Scientific and Industrial Research Organization (CSIRO) in Queensland, Australia. We found that the short-style morph has a more restricted distribution than the long- and medium-style morphs have [*see illustration on opposite page*]. The short form grows primarily in the Amazon basin and the Pantanal, and it has been sighted in the Paraguay and Paraná rivers. From a global perspective the medium-style morph predominates throughout the nonnative range; the long-style morph occurs less frequently. What factors account for this curious geographic pattern of style morph distribution?

Although the short-style floral morph grows as rapidly and floats as freely as the other morphs, it is confined to parts of South America probably because of its relationship with a local pollinator. Several different types of bees visit the large flowers of the water hyacinth to feed on pollen and nectar. Most of them contact the stigma (the pollen receptor) at the tip of the long and medium styles and thereby mediate pollination. The narrow flower of the water hyacinth conceals the stigma at the tip of the short style from most pollinators, however. The long-tongued bee *Ancyloscelis gigas* is the only known pollinator that can easily contact the stigma of the short-style morph, and that may account for the restricted range of the short-style morph. The fact that the short-style morph is absent from the Old World range appears to be simply the result of chance. The short-style morph has not been exported to other regions.

The geographic distribution of style morphs reflects a principle introduced by the evolutionary biologist Ernst Mayr of Harvard University: a new population that is started by a few individuals displays less genetic variation than does the parent population. Hence, if a few clones are isolated from their ancestral gene pool and inaugurate a new population, they can give rise to genetically uniform populations. Just such a sequence of events appears to have happened many times during the spread of the water hyacinth.

Most populations in the nonnative range are composed of the medium-style morph. Historical record and geographic distribution seem to imply that many invasions may have originated in Venezuela, where that morph predominates. The fact that the long-style morph appears periodically (albeit infrequently) in the nonnative region does not necessarily mean that it was introduced separately. A population of medium-style morphs can give rise to long-style morphs because of the pattern of inheritance of tristyly.

In the water hyacinth two genes determine the length of the style. One gene controls whether or not the style will be short. If the short trait is not expressed, a second gene controls whether the style is medium or long. The second gene has two alleles, dominant (M) and recessive (m). If a plant is homozygous for m (both of its copies of the gene are recessive), its flowers will have a long style. If a plant has at least one copy of the dominant gene (it is mM or Mm or MM), its flowers will have a medium style.

In heterozygous plants the pollen and ovules, each of which carries a copy of the gene, can have either form, dominant (M) or recessive (m). When the forms combine during sexual reproduction, the gene complement of the offspring has one chance in four of being homozygous for a dominant

VAST MATS of water hyacinth can cover several square kilometers of water surface and can grow as much as two meters thick. The plants deplete nutrients and block sunlight, starving out other plants and animals. They can also slow boat traffic.

medium-style gene (*MM*), two chances of being heterozygous for the dominant gene (*mM* or *Mm*) and one chance of being homozygous for the recessive gene (*mm*)—and of having a long style. In a population of medium-style morphs that are heterozygous, then, about one fourth of the offspring of sexual reproduction will have flowers with long styles. If a few long-style plants appear in populations that mainly have medium-style plants, one can assume that the population probably reproduces sexually, which in turn indicates that its environment allows seeds to germinate and seedlings to grow.

In California, where only medium-style plants occur, seeds were collected and segregated to grow long-style plants in a greenhouse. The absence of the long-style morph in California, then, indicates that the water hyacinth probably does not reproduce sexually under the ecological conditions typically found in the state. It is also possible that populations in California originate from one or a few clones that have propagated vegetatively since the beginning of the century. Many other regions of the introduced range also exhibit this genetic uniformity.

Botanists hope to exploit genetic uniformity as a way to control the water hyacinth. For a time, it was believed that the tropical sea cows called manatees could be introduced into regions to feed on the water hyacinth and thus control its growth. This program was successful in Guyana. In other areas, however, researchers discovered that the manatee does not find the water hyacinth very tasty and often prefers other vegetation to it.

Although the results of programs for eradicating the water hyacinth have been disappointing, success has been achieved in controlling the world's second most troublesome aquatic plant, the kariba weed, *Salvinia molesta*. This curious free-floating fern is so tiny and delicate that most people are surprised to learn that it is an aggressive invader.

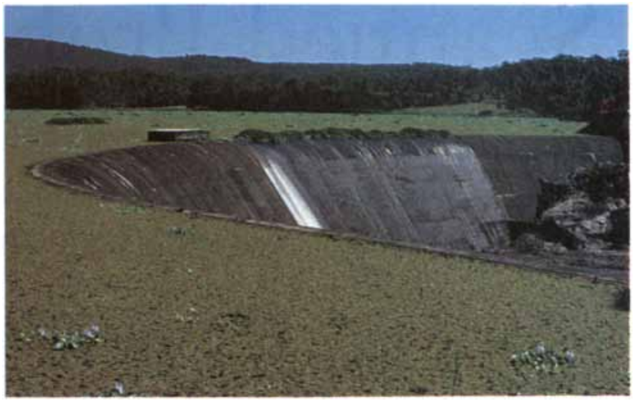
The kariba weed forms mats about one meter thick and spreads out over water surfaces in a way similar to the water hyacinth. Under favorable conditions the kariba weed can double its biomass in as little as 2.2 days—five times faster than the water hyacinth.

In the past 50 years the explosive growth of *S. molesta* has had an adverse socioeconomic effect in parts of Africa, Asia and Australia. A notable invasion took place at Lake Kariba on the Zambezi River in Africa, where at its peak in 1962 the weed covered 1,000 square kilometers, nearly one quarter of the total area of what was then the world's largest reservoir. The infestation earned the plant a common name, the "kariba" weed.

Nowhere has the plant had more of an impact than in the floodplains of the Sepik River in Papua New Guinea, north of Australia. After its introduction in the early 1970's, a colony of kariba weed covered all the lakes in the lower half of the floodplain, a total of 250 square kilometers of water surface. The invasion threatened the lives of 80,000 people who depend on the river for food and transportation.

The kariba weed had been identified as *Salvinia auriculata*, a native of

MANATEE, commonly known as the tropical sea cow, munches away on water hyacinth in Blue Spring State Park in Orange City, Fla. The manatee has been introduced into tropical waterways to help control the rapid spread of water hyacinth.



DELICATE LEAVES of the kariba weed, *Salvinia molesta* (*left*), belong to a single genetic individual that blankets waterways around the globe. The weed is perhaps the most massive entity in the botanical world. A floating mat of it is shown at the right.

South America, until the 1970's. In 1972 David S. Mitchell of CSIRO in New South Wales located a herbarium specimen of *S. auriculata*. The specimen had been discovered in 1941 along with two other related species. All three specimens originated from the botanical garden in Rio de Janeiro. Botanists began to suspect that the kariba weed was a horticultural hybrid, an offspring of the two related species from the garden. The fact that the kariba weed was sterile seemed to confirm the plant's hybrid origin.

Later Mitchell described the kariba weed as a new species and named it *Salvinia molesta*, the epithet signifying its aggressive nature. In 1978 the CSIRO team of Forno and K. L. S. Harley at last discovered the native range of *S. molesta* in southeastern Brazil and cast doubt on the original theory that the plant was an artificial hybrid.

Although it now seems unlikely that *S. molesta* originated in a botanical garden, its characteristic hybrid vigor is undoubtedly one of the secrets of the plant's extraordinary behavior. Hybrid plants often grow rapidly and are usually sterile. This is the case for *S. molesta*. In contrast to the water hyacinth, which can reproduce sexually, *S. molesta* propagates entirely by clonal means.

The fact that *S. molesta* is asexual means that the world's entire population of the plant may be a single genetic individual. Because many millions of tons of the weed are distributed around the world, from a genetic viewpoint *S. molesta* may lay claim to being the largest individual organism on the earth.

The aggressive behavior of the kariba weed has largely been displayed outside its native South American range. This is true in the case of most biological invasions. Plants and animals usually populate their native environments at densities commensurate with their ecological role in a balanced community, but when species are introduced to another part of the world, they leave behind the coevolved competitors and enemies that normally keep their populations in check. The absence of natural enemies in the nonnative range allows alien populations to increase rapidly and leads to "ecological release."

Knowledge of the causes behind the different behaviors of species in their native and introduced ranges has led to novel methods of managing pest and weed outbreaks. These methods, known as biological control, reduce population numbers to acceptable levels through the planned release of host-specific natural enemies.

Biological control of the kariba weed was initiated shortly after botanists located its native range. While exploring the plant's Brazilian homeland, Forno, P. M. Room and P. A. Thomas of CSIRO in Queensland discovered a new beetle species that feeds exclusively on *S. molesta*. (The beetle was later named *Cyrtobagous salviniae*.) Their work is the most successful example so far of biological control of an aquatic weed.

The beetle was brought to Australia and released at Lake Moondarra, where it rapidly destroyed an infestation of kariba weed covering two square kilometers. The beetle's most spectacular success was achieved in Papua New Guinea, where between 1983 and 1985 the weed cover was reduced from about 250 square kilometers to two square kilometers. It was estimated that the beetles consumed two million metric tons of the weed in just two years. Other beetle programs are now under way in India and Namibia and are rapidly achieving control.

Although biological control is not a universal solution to all pest and weed problems, the technique may be applied to many other waterweed invasions. A survey conducted by Jeremy J. Burdon and Don Marshall of CSIRO in Canberra examined 81 attempts to control 45 weed species. They found a correlation between the level of control and the reproductive system. Asexual species were controlled much more effectively than those that rely on sexual reproduction.

Presumably biological control succeeds if genetic variation is limited. If this assumption is true, the water hyacinth and many other aquatic weeds that regenerate primarily by clonal means should be excellent targets for biological control. Clonal propagation, which allows water hyacinth and kariba weed to dominate the world's waterways, may yet provide a way to sink these noxious invaders.

FURTHER READING

TRISTYLY IN *EICHHORNIA CRASSIPES* (MART.) SOLMS (WATER HYACINTH). Spencer C. H. Barrett in *Biotropica*, Vol. 9, No. 4, pages 230-238; December, 1977.

THE WORLD'S WORST WEEDS: DISTRIBUTION AND BIOLOGY. L. G. Holm et al. University Press of Hawaii, 1977.

SEXUAL REPRODUCTION IN *EICHHORNIA CRASSIPES* (WATER HYACINTH). S. C. H. Barrett in *Journal of Applied Ecology*, Vol. 17, pages 101-124; 1980.

BIOLOGICAL CONTROL OF WEEDS WITH PLANT PATHOGENS. Edited by Raghavan Charudattan and H. Lynn Walker. John Wiley & Sons, Inc., 1982.

TAXONOMY AND CONTROL OF *SALVINIA MOLESTA*. P. A. Thomas and P. M. Room in *Nature*, Vol. 320, No. 6063, pages 581-584; April 17, 1986.

Scanned-Probe Microscopes

By examining a surface at very close range with a probe that may be just a single atom across, they can resolve features and properties on a scale that eludes other microscopes

by H. Kumar Wickramasinghe

Objects smaller than the wavelengths of visible light are a staple of contemporary science and technology. Biologists study single molecules of protein or DNA; materials scientists examine atomic-scale flaws in crystals; microelectronics engineers lay out circuit patterns only a few tens of atoms thick. Until recently this minute world could be seen only by cumbersome, often destructive methods such as electron microscopy and X-ray diffraction. It lay beyond the reach of any instrument as simple and direct as the familiar light microscope.

A family of new microscopes opens this realm to direct observation. The devices can map atomic and molecular shapes, electrical, magnetic and mechanical properties and even temperature variations at a higher resolution than ever before, without the need to modify the specimen or expose it to damaging, high-energy radiation. The achievement seems implausible. More than 100 years ago, after all, the German physicist and lensmaker Ernst Abbe described a fundamental limitation of any microscope that relies on lenses to focus light or other radiation: diffraction obscures details smaller than about one half the wavelength of the radiation.

H. KUMAR WICKRAMASINGHE is Manager of Physical Measurements at the IBM Thomas J. Watson Research Center in Yorktown Heights, N.Y., where he specializes in scanned-probe microscopes and their applications in microelectronics manufacturing. After earning his Ph.D. in electrical engineering from University College, London, in 1974, he moved to Stanford University, where he did some of the fundamental work in acoustic and photoacoustic microscopy. He joined the faculty at University College in 1978 and went to IBM in 1984. In his spare time Wickramasinghe enjoys hiking with his family.

The new microscopes—typified by the scanning tunneling microscope, for which Gerd Binnig and Heinrich Rohrer of the IBM Zurich Research Laboratory received a Nobel prize in 1986—overcome this Abbe barrier with ease. The principle by which they do so was first described in 1956. In that year J. A. O'Keefe, then of the U.S. Army Mapping Service, proposed a microscope in which light would shine through a tiny hole in an opaque screen, illuminating an object directly in front of the screen. Light transmitted through the specimen or reflected back through the hole would be recorded as the sample was scanned back and forth. O'Keefe pointed out that the resolution of such a "scanning near-field microscope" would be limited only by the size of the hole and not by the wavelength of the light. In principle the device could make superresolving images—images showing details smaller than half a wavelength.

O'Keefe acknowledged that technology capable of positioning and moving an object with the needed precision did not exist. By resorting to long-wavelength radiation, however, Eric Ash of University College, London, adopted the O'Keefe strategy in 1972 to circumvent the Abbe barrier. He passed microwave radiation at a wavelength of three centimeters through a pinhole-size aperture and scanned an object in front of it to record an image with a resolution of 150 microns—one two-hundredth of a wavelength.

By that time, means of controlling sample position and movement with the precision needed to surpass the resolution of a conventional light microscope were becoming available. In the same year as Ash's demonstration, Russell D. Young of the National Bureau of Standards succeeded in manipulating objects in three dimensions with a precision of about a nanometer (a billionth of a meter). He relied on piezoelectrics—ceramic materials that change size ever so slightly when an electrical potential across the material is changed. Piezoelectric controls opened the way to the development, in 1981, of the supreme example of a scanning near-field microscope, the scanning tunneling microscope, or STM [see "The Scanning Tunneling Microscope," by Gerd Binnig and Heinrich Rohrer; SCIENTIFIC AMERICAN, August, 1985].

In the STM the "aperture" is a tiny tungsten probe, its tip ground so fine that it may consist of only a single atom and measure just .2 nanometer in width. Piezoelectric controls maneuver the tip to within a nanometer or two of the surface of a conducting specimen—so close that the electron clouds of the atom at the probe tip and of the nearest atom of the specimen overlap. When a small voltage is applied to the tip, electrons "tunnel" across the gap, generating a minuscule tunneling current. The strength of the current is exquisitely sensitive to the width of the gap; typically it decreases by a factor of 10 each time the gap is widened by .1 nanometer—half the diameter of an atom.

X and *y* piezoelectric controls (which govern motion in the two dimensions of a plane) move the probe back and forth across the specimen surface in a raster pattern, its parallel tracks separated by perhaps a fraction of a nanometer. If the probe maintained a steady height, the tunneling current would fluctuate dramatically, increasing as the tip passed over bumps such as surface atoms and falling to nothing as it crossed gaps between atoms. Instead the probe moves up and down in concert with the topography. A feedback mechanism senses the variations in tunneling current and varies the voltage applied to a third, *z*, control. The *z* piezoelectric moves the probe vertically to stabilize the current and maintain a constant gap between the microscope's tip and the surface.

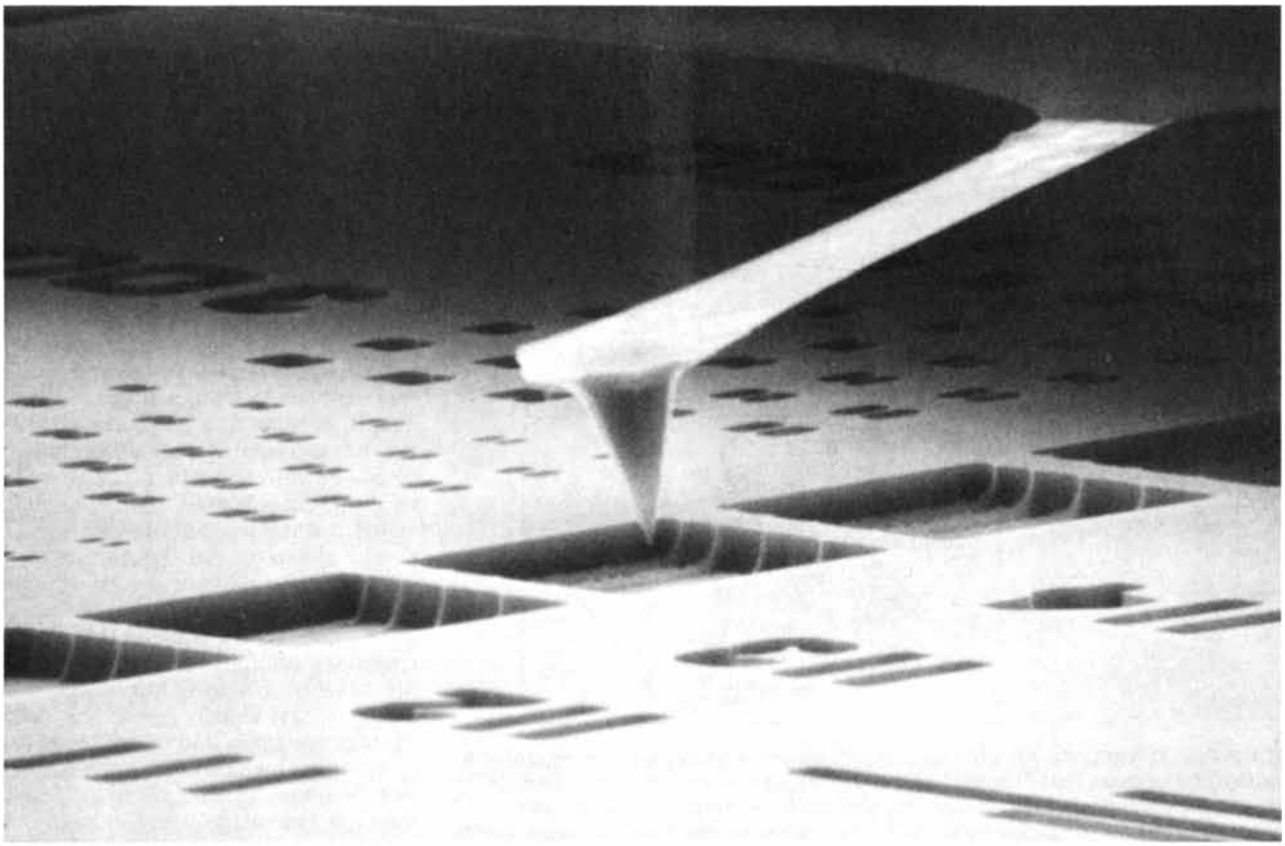
The variations in the voltage applied to the *z* piezoelectric are electronically translated into an image of surface relief. If the sharpness of the probe, the precision of the controls and the fineness of the raster scan are all sufficient, STM images can reveal individual atoms, as small as .2 nanometer in diameter. The images are superresolving: the quantum-mechanical wavelength of the tunneling electrons in the probe—the "radiation" that gives rise to the image—is approximately one nanometer.

What such images map is not topography in the usual sense but a surface of constant tunneling probability. The tunneling probability is affected by topography, but it is also affected by variations in the abundance and energies of surface electrons. When the specimen is composed of just a single element, tunneling probability closely follows topography, but "topography" can also reveal atom-by-atom variations in composition. A contaminant atom on an otherwise uniform surface, for example, may appear as an anomalous pit or bump, depending on its electronic properties.

The STM's success gave workers confidence that a probe can be scanned over a sample with atomic precision by piezoelectric controls. Since then, the microscope has imaged surface structure on a host of substances and has even been harnessed as a nanometer-scale tool: the tip can apply a precisely targeted voltage that can dissect molecules or probe their electronic properties. What is more, the STM gave birth directly to a whole family of scanned-probe microscopes based on similar technology. The first spinoff addressed one of the progenitor's major limitations: the STM is largely restricted to imaging electrical conductors. Even conducting or semiconducting materials, such as silicon, are often covered with an insulating layer of oxide. Biological materials, too, are in general nonconducting, although biological samples placed on a conducting surface and immersed in an electrolyte have been profiled with the STM.

In 1985 Binnig, together with Calvin F. Quate of Stanford University and Christoph Gerber of IBM Zurich, introduced the atomic force microscope (AFM), a scanned-probe device that does not need a conducting specimen. Like the STM, the device moves a minute tip—in this case, an atomically sharp shard of diamond mounted on a strip of metal foil—over the specimen in a raster pattern. In place of tunneling current, the AFM records contours of force—the repulsion generated by the overlap of the electron cloud at the tip with the electron clouds of surface atoms. In effect the tip, like the stylus of a phonograph, "reads" the surface. The foil acts as a spring, keeping the tip pressed against the surface as it is jostled up and down by the atomic topography.

In the original AFM design, a tunneling current flowing between the foil and an STM tip mounted just above it measured the foil's deflection. A feedback mechanism responded to variations in the tunneling current by ad-

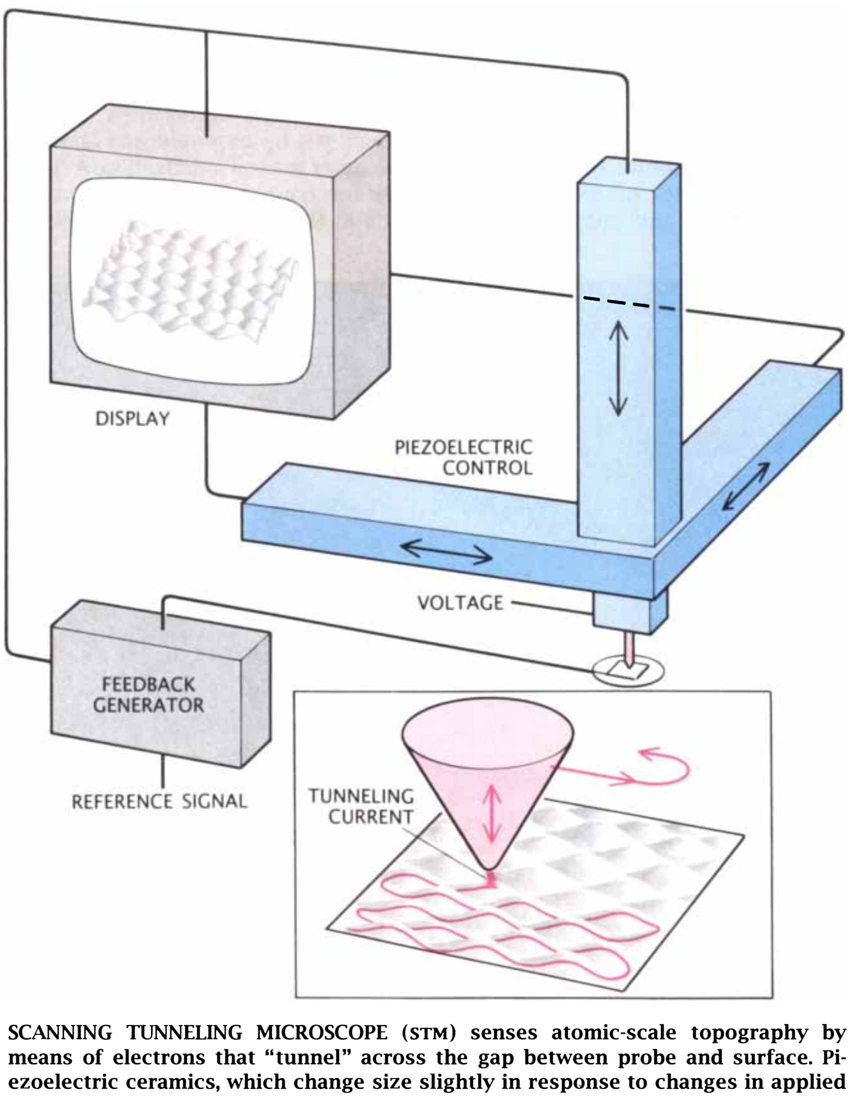


SILICON PROBE of a laser force microscope, tapered to a point just atoms across, hovers over a surface. The microscope, developed in the author's laboratory, images surface relief by scanning a tiny vibrating probe a few nanometers (billionths of a meter) above the sample, where it "feels" weak attractive forces from the surface. The scanning electron micrograph, at a magnification of about 1,300, is by Olaf Wolter of the IBM German Manufacturing Technology Center in Sindelfingen.

justing the voltage on a *z* piezoelectric control, which moved the sample up or down. The deflection, and hence the repulsive force, was thereby held constant; the variations in the voltage on the *z* piezoelectric mimicked the sample's topography and served as the basis for the image. In this way the microscope could reveal individual atoms. Like the STM, the AFM was superresolving, its resolution limited only by the fineness of the diamond tip and not by any wavelength.

Even though the first AFM could in principle image any nonconducting or conducting substance, the pressure of the diamond stylus (about a millionth of a gram) was enough to distort or move many biological molecules. Recently Paul K. Hansma of the University of California at Santa Barbara and his colleagues have reduced the pressure by a factor of 10. One factor that increases the pressure of the tip is the thin film of water and contaminants that inevitably collects on both the tip and the sample. As the tip nears the surface and the contaminant films come into contact, adhesive forces draw the tip and sample together, thereby increasing the tracking pressure. Hansma's group has eliminated this effect by submerging the tip and sample in a drop of water.

The workers have also replaced the STM mechanism with a new means of detecting tip deflection, developed at the IBM Thomas J. Watson Research Center in Yorktown Heights, N.Y.—a laser beam that is reflected off the foil. Movement of the foil shifts the path of the reflected beam, so that a photocell placed in the beam path some distance away can detect tiny movements of the foil. The signal from the photocell activates the *z* piezoelectric to maintain a steady tip displacement. The optical sensor provides a more reliable measure of tip deflection than the tunneling sensor does, and so it renders the AFM's touch more consistent and gentle. As a result of these improvements, Hansma and his colleagues have been able to image biological substances in nearly atomic detail. They have even taken snapshots of a molecular process as it unfolds—the polymerization of the protein fibrin, a major component of blood clots.

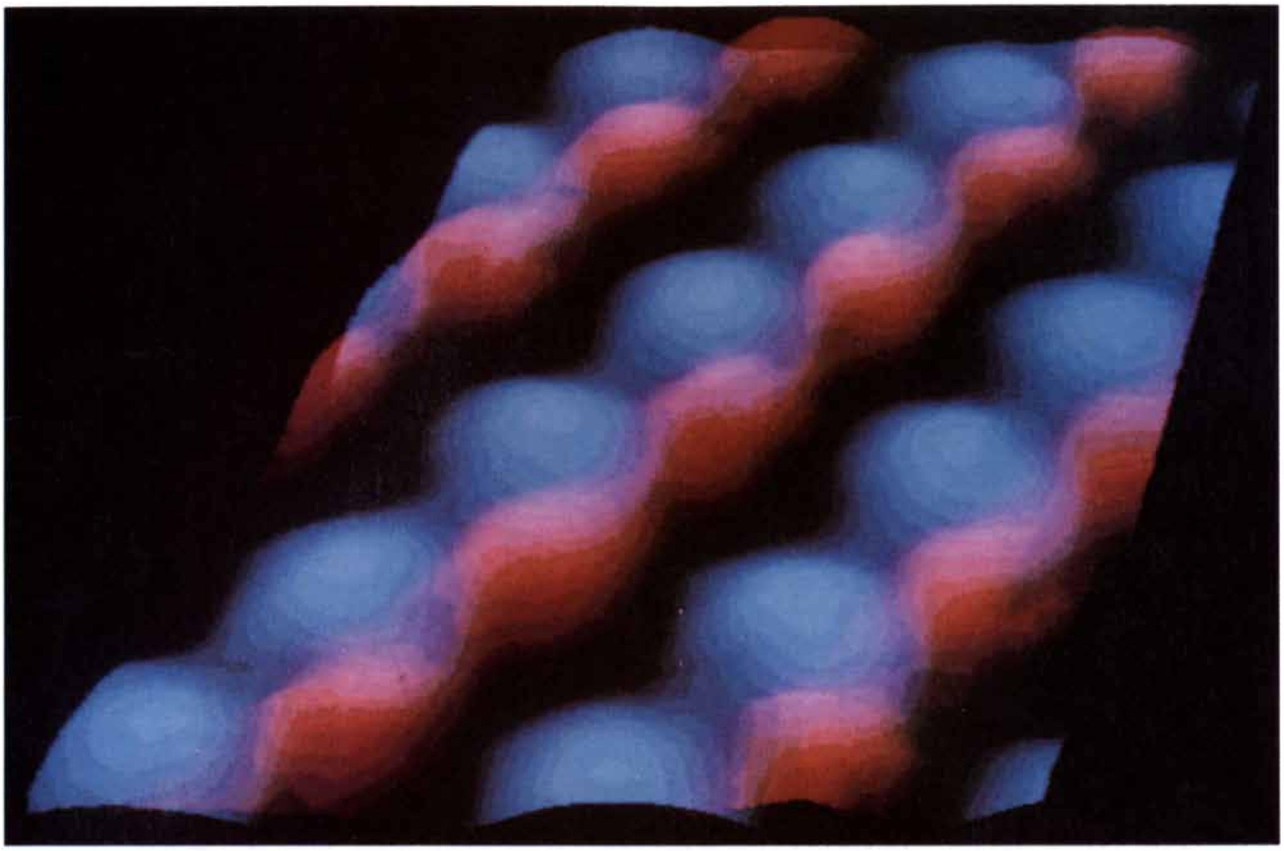


SCANNING TUNNELING MICROSCOPE (STM) senses atomic-scale topography by means of electrons that "tunnel" across the gap between probe and surface. Piezoelectric ceramics, which change size slightly in response to changes in applied voltage, maneuver the tungsten probe in three dimensions. A voltage is applied to the tip, and it is moved toward the surface (which must be conducting or semiconducting) until a tunneling current starts to flow. The tip is then scanned back and forth in a raster pattern. The tunneling current tends to vary with the topography; a feedback mechanism responds by moving the tip up and down, following the surface relief. The tip's movements are translated into an image of the surface.

At the time the AFM was being developed, my group in the manufacturing research department of IBM in Yorktown Heights was searching for new ways to monitor production quality in microelectronics manufacturing. In an effort to build faster and more powerful computer hardware, engineers are steadily shrinking circuit elements and data-storage devices. Inspecting circuits for flaws demands a resolution at least 10 times finer than the smallest element, and the dimensions of circuit elements themselves are already approaching the theoretical resolution limit of conventional optical microscopes, about 250 nanometers.

For some time the scanning electron microscope (SEM) has been a standard tool in microelectronics: it can resolve details as small as a few nanometers. Yet the SEM requires that the specimen be coated with metal and imaged in a vacuum, and it has poor three-dimensional resolution. Moreover, its high-energy electrons can damage or destroy a semiconductor device, which limits the SEM's value for monitoring production quality. Scanned-probe microscopes—simple, direct and powerful—seemed to answer our needs.

Yet the STM usually requires a conducting specimen, and microelectronic devices include semiconductors and insulators as well as conductors. The AFM can image a wider range of materials, but even the improved, gentler AFM makes contact with enough force to contaminate or damage delicate circuit elements. My colleagues and I have therefore developed a new



ATOMS are aligned like beads in an STM image of a sample of the semiconductor gallium arsenide. The image is a composite: because atoms of gallium (*blue*) and of arsenic (*red*) have different electrical properties, they were imaged separately, under different conditions. For the gallium atoms, the tunneling current flowed to the sample from a negatively charged tip; for the arsenic atoms, the STM's tip was positively charged and the current flow was reversed. The image was provided by Randall M. Feenstra of the IBM Thomas J. Watson Research Center in Yorktown Heights, N.Y.

family of scanned-probe microscopes that can profile electronic devices regardless of composition, without touching the surface.

Chief among them is the laser force microscope (LFM), invented in collaboration with Yves Martin and Clayton C. Williams. The "force" of the LFM is the small attractive force that develops between a surface and a probe from two to 20 nanometers away—a much larger gap than in the STM and AFM. On semiconductors and insulators the force results largely from the surface tension of water that condenses between the tip and the sample, but van der Waals interactions (weak, transient electrostatic attractions between atoms or molecules) also contribute.

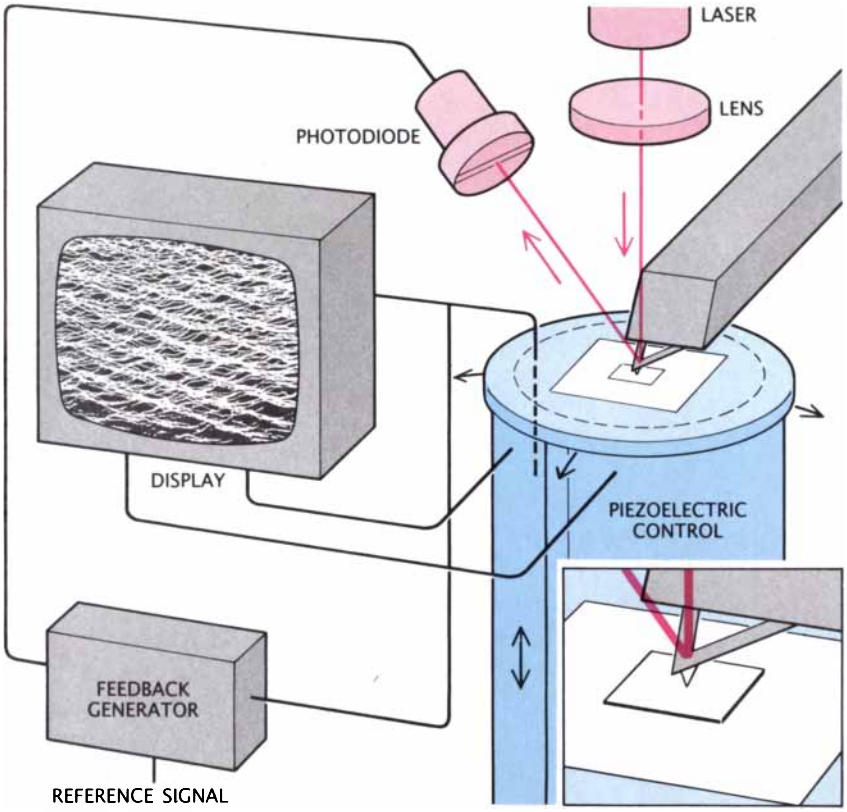
The attractive force is minute—1,000 times smaller than the interatomic repulsions recorded by the AFM. The LFM detects the force by its effect on the dynamics of a vibrating probe—a tapered tungsten wire half a millimeter long with a downturned tip that is etched to a diameter of 50 nanometers or less. (Recently silicon probes ending in an atomically fine tip have been introduced.) A piezoelectric transducer at the base of the wire converts alternating current into vibration. The driving frequency of the transducer lies just above the wire's lowest mechanical resonance (typically about 50 kilohertz).

Because the wire is vibrating at close to its resonance frequency, like the reed in a musical instrument, it amplifies the driving signal. The tip may oscillate by half a nanometer even though the transducer at the other end is moving only a hundredth as far. When the vibrating tip approaches the sample, however, the weak attractive forces it feels in effect "soften" the wire: they lower its resonance frequency. The driving frequency now lies farther from the resonance frequency, and so the vibration amplitude diminishes.

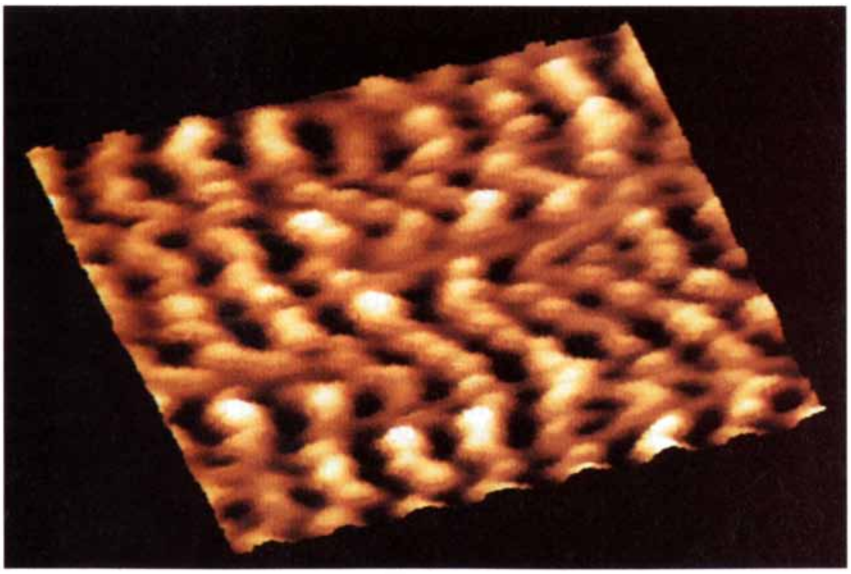
A laser sensor detects the amplitude change. It does so by interferometry, a technique for precise distance measurement much applied in astronomy and geophysics, for example. A laser beam is split into a reference beam, reflected by a stationary mirror or prism, and a probe beam, reflected by the back of the tip. The reflected beams are recombined, and they interfere with each other to yield a beam whose phase is sensitively dependent on the path length of the probe beam. The phase shifts back and forth with each excursion of the tip; the extent of the shift reveals the amplitude of the vibration. In this way the interferometer can detect amplitude changes as small as 10^{-5} nanometer.

The amplitude tends to fall over topographic highs, where the attractive forces are strongest, and to increase over depressions. As in the other scanned-probe microscopes, a feedback mechanism responds to changes in the tip's behavior by varying the voltage across a *z* piezoelectric to stabilize the vibration amplitude and hence the probe-to-surface gap. The fluctuations in piezoelectric voltage are converted into a surface profile.

In this way the LFM can detect surface relief of as little as five nano-



ATOMIC FORCE MICROSCOPE (AFM) scans a sample with a shard of diamond mounted on a thin metal arm. The electron cloud of the diamond tip (which may end in a single atom) presses against the clouds of individual atoms in the sample, generating a repulsive force that varies with the surface relief. The force deflects the tip, whose movements are monitored by a laser beam reflected from the top of the arm to a photodiode sensor. A feedback mechanism responds to the changes in the beam's path by activating a piezoelectric control, which adjusts the sample's height so that the deflection of the arm remains constant. The sample's movements are translated into a surface profile. Unlike the STM, the AFM can readily image electrical insulators.



POLYMER CHAINS of the amino acid alanine corrugate the surface of a microscope slide, where a solution of the polymer was allowed to dry. In this AFM image, by Paul K. Hansma of the University of California at Santa Barbara and his colleagues, each bump is an individual amino acid. The image shows an area three nanometers wide.

meters (some 25 atomic thicknesses), and because it senses topography from a distance, it can inspect features inside deep, narrow clefts. It promises to be valuable not only for examining finished microcircuits but also for monitoring the quality of the silicon surfaces on which they are built. As circuit features shrink to a few hundred nanometers in width and tens of nanometers in thickness, a substrate that departs from perfect smoothness by more than a few atomic thicknesses can be unacceptable. By the same token, the increasing density with which information is stored on magnetic disks means that the data must be written and read on a finer scale, by smaller heads flying closer to the disks. To avoid colliding, both the disks and the heads must be smooth on a scale of atoms.

A variant of the LFM, the magnetic-force microscope (MFM), enables us to look at the actual performance of such heads—the definition, uniformity and strength of the magnetic field they produce. In place of the tungsten or silicon needle, the MFM has a magnetized nickel or iron probe. When the vibrating probe is brought near a magnetic sample, the tip feels a force that changes its resonance frequency and hence its vibration amplitude. The MFM can trace the magnetic-field pattern emanating from data-recording heads at a resolution of better than 25 nanometers. The device can also study the structure of the magnetic bits that store data on disks and other media, giving insight into both head performance and the quality of the storage medium.

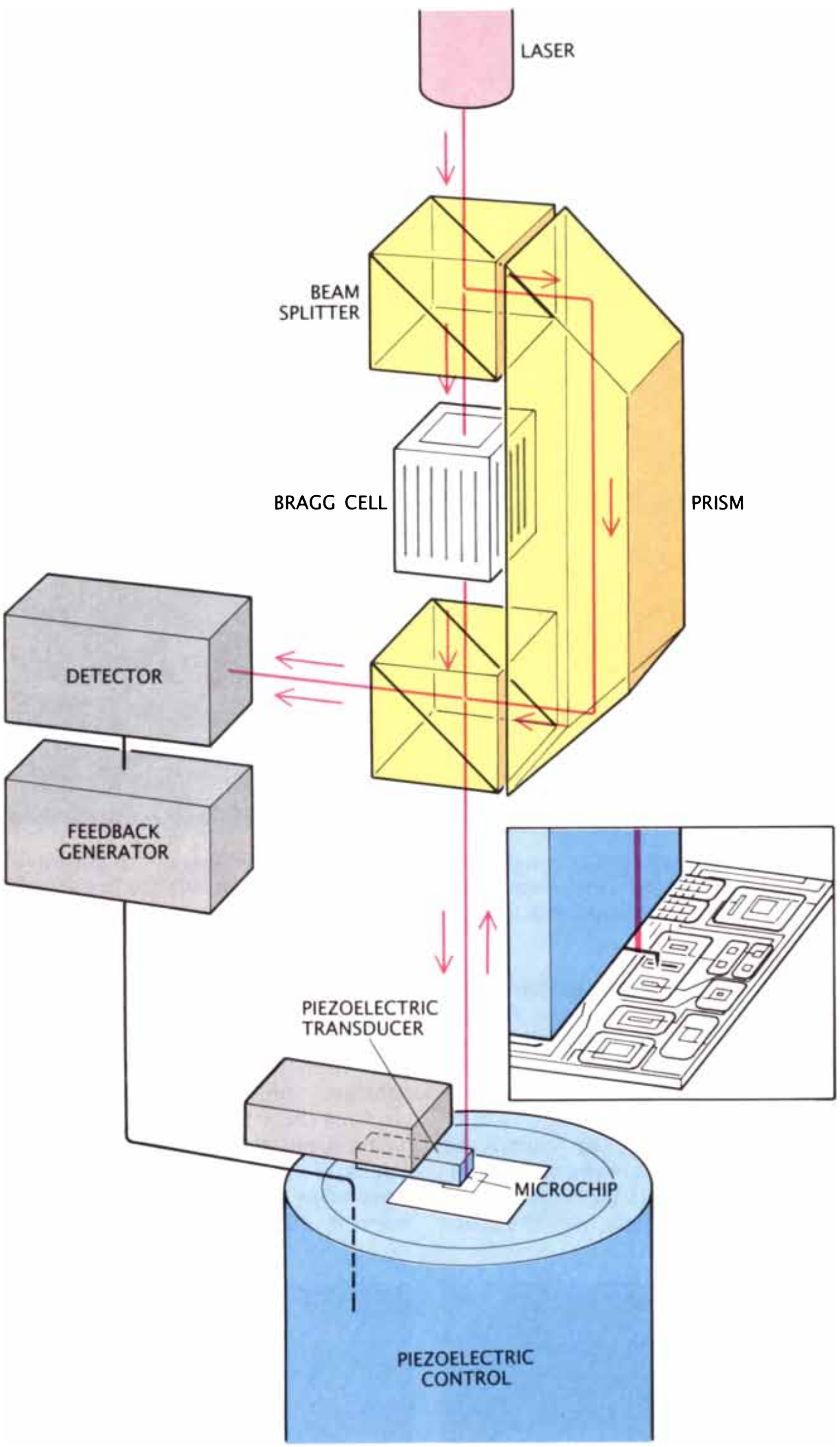
Yves Martin, David W. Abraham and I have developed another specialized offspring of the LFM, the electrostatic-force microscope, which promises to fill yet another role in microelectronics design and manufacturing. Here the vibrating probe bears an electric charge, and its vibration amplitude is affected by the electrostatic forces resulting from charges in the sample. With this microscope we can map the electrical properties of microcircuits on a very fine scale. For example, small numbers of impurity atoms, called dopants, are added to silicon to modify its properties; either the dopants donate electrons, which are then free to wander through the silicon, or they capture electrons, leaving positively charged "holes" that also can migrate through the crystal lattice.

The distribution and concentration of dopant atoms plays a critical role in chip performance. One way to map these dopants is to apply a voltage across the gap between the probe of an electrostatic-force microscope and the surface. The voltage mobilizes the conduction electrons or holes beneath the probe, leaving a charged region that exerts an electrostatic force on the tip. The consequent movements of the tip provide a precise, fine-scale measure of the charge and hence of the number of mobilized electrons or holes and the concentration of dopant atoms.

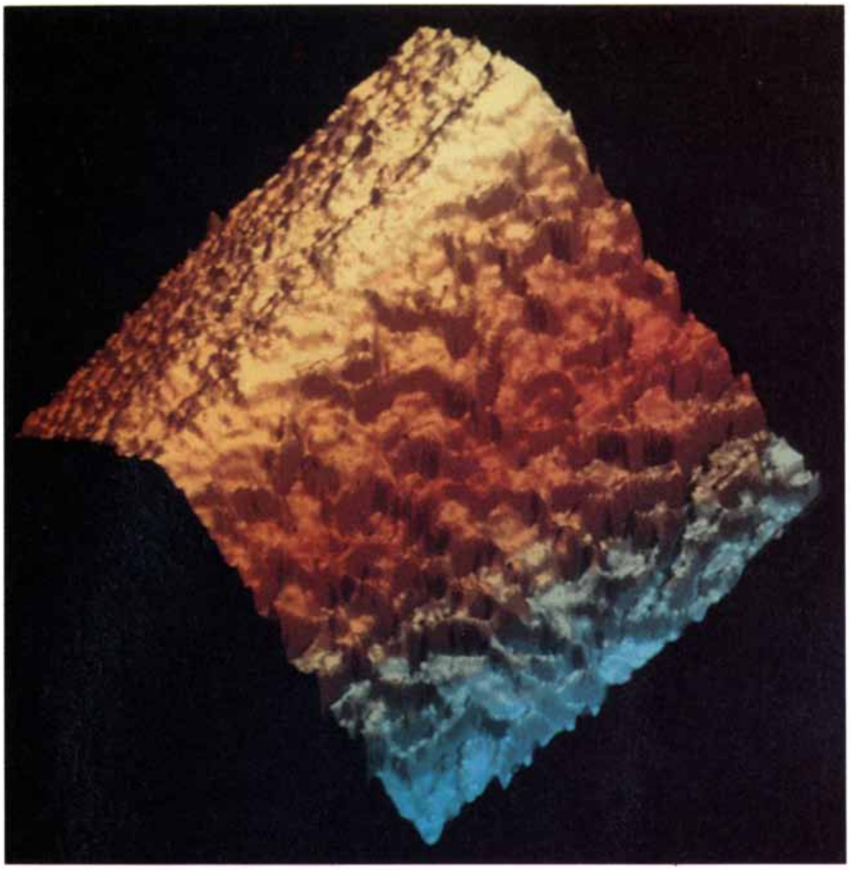
The magnetic- and electrostatic-force microscopes are just two of the scanned-probe systems that have been developed to map surface properties other than topography. Another is the scanning thermal microscope, which I developed in collaboration with Williams. The microscope's probe may be the world's tiniest thermometer: it can measure surface-temperature variations of a ten-thousandth of a degree on a scale of tens of nanometers. The probe consists of a tungsten wire ending in a point perhaps 30 nanometers across. The wire is coated with a second metal, nickel, which is separated from the tungsten by a layer of insulator everywhere except at the tip of the probe. The tungsten-nickel junction acts as a thermocouple, generating a voltage proportional to its temperature.

We developed the scanning thermal microscope early in our search for a way to profile microelectronic devices, before the introduction of either the AFM or the LFM. In our initial experiments, in 1985, we found that it can indeed serve as a rapid surface profiler. First a direct current is passed through the tip to heat the probe; when the energy lost to the air as heat equals the energy supplied as current, the temperature of the tip stabilizes, typically at a few degrees above its surroundings.

When the heated tip approaches a sample, which—being a solid—is a much better heat conductor than air, its rate of heat loss increases. It cools, and the resulting drop in the voltage across the thermocouple junction provides a measure of the tip's proximity to the sample. (To reduce the system's sensitivity to temperature fluctuations, what we actually do is to vibrate the tip by less than a nanometer, at a rate of about a kilohertz. The voltage falls every time the tip approaches the sample and increases as it swings away; the amplitude of the voltage fluctuation increases as the average gap between the vibrating tip and the sample is reduced.) Thus, heat



LASER FORCE MICROSCOPE senses topography by moving a tungsten or silicon probe across a sample (typically a microelectronic component) at a height of a few nanometers. The tip is set vibrating at close to its resonance frequency. Attractive forces—so-called van der Waals attractions and the surface tension of water that condenses between tip and sample—pull on the tip, changing its resonance frequency and so reducing its vibration amplitude. The variation in the forces, and hence in tip travel, reveals surface relief. A laser probe tracks the tip by means of a beam that is split into two. One of the beams is reflected through a stationary prism; the other passes through a Bragg cell—a device that shifts the beam's frequency—and is reflected from the back of the probe. The beams are recombined, and their interference produces a signal (at the Bragg-cell frequency) that measures the tip vibration.



NANOMETER-SCALE RIDGES roughen one side of a micron-wide groove etched in a silicon surface. Pierre Levy, then of IBM in Yorktown Heights, made the image with the laser force microscope, which can profile both coarse and fine-scale topography.

loss reveals surface topography as the probe is scanned across the sample, in the same way as tunneling current, interatomic repulsion or van der Waals attractions do in other scanned-probe microscopes.

The thermocouple tip cannot be made much finer than about 30 nanometers, which limits the resolution of surface profiles made with the scanning thermal microscope. The advent of the LFM, with its much higher resolution, freed the thermal probe for more specialized uses. It can map temperature variations in living cells, yielding clues to patterns of metabolism; it can measure the flow rate of a minute stream of liquid or gas by monitoring heat loss when the tip is warmed and placed in the stream. It also brings an analytical technique called photothermal absorption spectroscopy to an unprecedented level of resolution.

Photothermal absorption spectroscopy takes advantage of the fact that specific elements absorb light most efficiently at distinctive wavelengths. In this method a tunable laser is trained on a sample. The temperature of the sample is monitored as the laser wavelength is varied; a sharp increase in temperature at a specific wavelength indicates efficient absorption. A complete "fingerprint" of absorption wavelengths can reveal the composition of the sample. The thermal probe opens the way to doing such spectroscopy on a tiny scale, which will make it possible to map small-scale variations in surface composition.

My group has recently extended the resolution of the technique even further—to the level of individual atoms. We dispensed with the thermocouple probe and returned to the bare, atomically sharp tungsten tip of the STM. The tip acts as one of the two metals in a thermocouple junction; either the sample itself (if it is a metal) or a metal electrode on which it rests acts as the other. At first the instrument is operated as a tunneling microscope. A voltage difference is established between the tip and the metal, and the tip is moved toward the surface until a tunneling current begins to flow.

Then the current is switched off, and the surface is heated by the tunable laser. The tiny thermocouple formed by the heated surface and the probe (which is warmed by its proximity to the surface) generates a voltage proportional to its temperature. The temperature, in turn, indicates how much energy was absorbed by the atom over which the probe is poised. The absorption can be monitored

MAGNETIC BITS a micron or two across appear as craters in images made with a magnetic-force microscope—a laser force microscope equipped with a magnetized iron or nickel probe, which makes the microscope sensitive to gradients of magnetic force. The bits, which store information on magneto-optical disks, were formed by exposing a disk to a magnetic field while a tightly focused laser heated the surface, thereby allowing the magnetic domains in the heated region to realign themselves. The images, made by Yves Martin of IBM in Yorktown Heights, compare the magnetic structure of a normal bit (*left*) with bits that were written under reduced laser power (*middle*) and an unusually weak magnetic field (*right*).

atom by atom at each laser wavelength to tease apart a sample's composition at the finest level.

Those recent developments in my laboratory alone give some idea of the extraordinary versatility of scanned-probe technology. Still other variants have been developed elsewhere. Gary M. McClelland and his colleagues at the IBM Almaden Research Center in San Jose, Calif., have modified the AFM to measure friction on an atomic scale. The stylus of an AFM is swept sideways across a surface while a laser interferometer measures the deflection of the foil arm. By this means the workers have, for example, gauged the resistance offered by the atomic corrugations of a graphite surface. They hope to determine how such atomic-scale friction affects the bulk properties of materials.

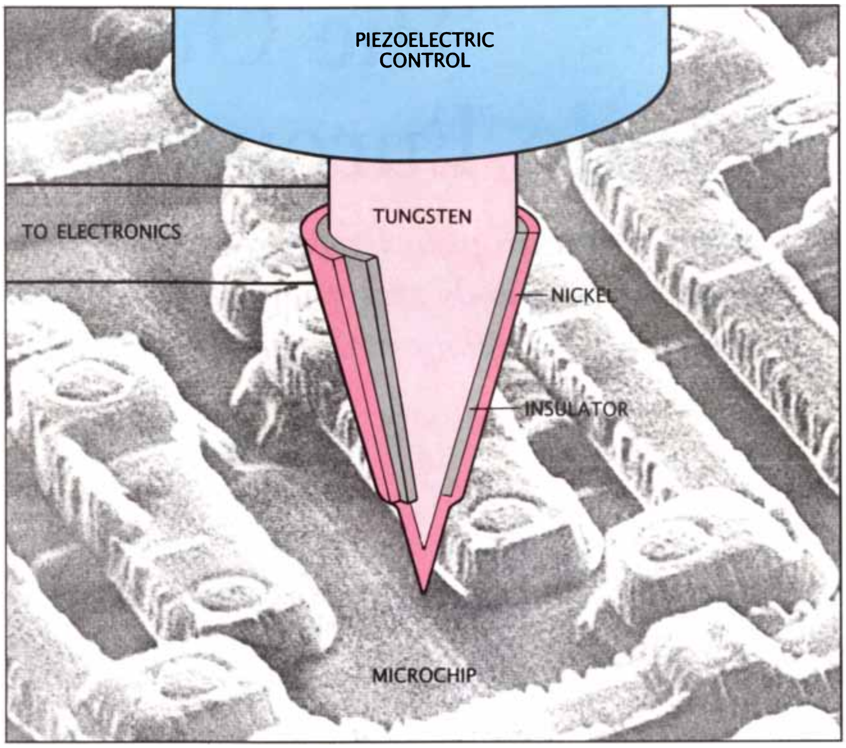
Hansma's group, meanwhile, has developed the scanning ion-conductance microscope (SICM), which scans a sample not with a metal or diamond tip but with a glass micropipette containing a tiny electrode. The workers immerse the sample and the pipette in a bath of electrolyte (such as a salt solution) and establish a voltage difference between the electrode within the pipette and another electrode in the bath. A current of ions tends to flow between the electrodes by way of the pipette opening, but when the pipette approaches the surface of the sample, the current weakens. It disappears altogether when the scanned pipette touches the sample, which blocks the opening.

Equipped with a feedback mechanism that maintains a constant ion current through the pipette opening, the SICM can profile a sample's topography. The fineness of the pipette does limit the SICM's resolution to about .2 micron (10 nanometers may eventually be feasible). The device is therefore best suited to more specialized functions—studying a living cell's electrical behavior, for example. When a cell is stimulated, so-called ion channels in its membrane open and close, allowing tiny currents of ions to pass. Hansma hopes to detect individual ion channels and study their behavior once the SICM's resolution has been improved.

Another scanned-micropipette device represents what one might call a return to the roots of scanning near-field microscopy. Three decades ago O'Keefe conceived of the strategy as a way of overcoming the Abbe barrier and enhancing the resolution of optical microscopes, but at that time the technology for scanning a sample with the precision of a fraction of a wavelength did not exist. The nanometer-precise controls pioneered by the STM have opened the way to the superresolving optical microscope envisioned by O'Keefe.

Two groups, one led by Dieter Pohl of IBM Zurich and the other by Michael S. Isaacson of Cornell University, have demonstrated scanning near-field optical microscopes. One side of a very thin sample is illuminated, and a micropipette, its walls coated with an opaque aluminum film, is scanned across the other side. A photodiode at the top of the pipette measures the light collected by the tip. In this way, transmitted-light micrographs have been recorded that have a resolution (limited by the pipette diameter) of about 50 nanometers, or about a tenth of a wavelength. Still finer pipettes may eventually yield a resolution of 10 nanometers—a 25-fold improvement over the best conventional optical microscopes.

Far from superseding optical microscopy, then, scanned-probe technology promises to extend its capabilities. Microscopists are now in a position to transfer the refinements of the past 100 years—techniques for enhancing contrast and highlighting specific features, such as polarization contrast, immunofluorescence and phase contrast—to scanning near-field optical microscopy. With existing scanned-probe microscopes we can "see" the nanometer realm of molecules and microcircuits. As the same technology is applied to optical microscopy, we may eventually come to see that world in the familiar terms of light, shadow and color.



THERMAL PROBE can map temperature variations as small as a ten-thousandth of a degree on a scale of tens of nanometers. The probe has a tungsten core, jacketed in nickel but insulated from it everywhere except at the 30-nanometer-wide tip. There the two metals join to form a thermocouple, which generates a voltage proportional to its temperature. The device can profile surface relief by sensing variations in the heat loss from a probe that is warmed and then scanned across the sample.

FURTHER READING

VACUUM TUNNELING: A NEW TECHNIQUE FOR MICROSCOPY. Calvin F. Quate in *Physics Today,* Vol. 39, No. 8, pages 26–33; August, 1986.

TIP TECHNIQUES FOR MICROCHARACTERIZATION OF MATERIALS. Y. Martin, C. C. Williams and H. K. Wickramasinghe in *Scanning Microscopy,* Vol. 2, No. 1, pages 3–8; March, 1988.

SCANNING TUNNELING MICROSCOPY AND ATOMIC FORCE MICROSCOPY: APPLICATION TO BIOLOGY AND TECHNOLOGY. P. K. Hansma, V. B. Elings, O. Marti and C. E. Bracker in *Science,* Vol. 242, No. 4876, pages 209–216; October 14, 1988.

The Origins of Indo-European Languages

Almost all European languages are members of a single family. The author contends that they spread not by conquest, as has been thought, but along with the peaceful diffusion of agriculture

by Colin Renfrew

One of the most contentious problems in the entire field of archaeology and prehistory is the explanation of the remarkable relations that link nearly all the European languages, many of the languages spoken in India and Pakistan and some of those in the lands between. For more than two centuries it has been known that the "Indo-European" languages are related. But what prehistoric process underlies that relationship? How did related languages come to be spoken over such a wide area? What implications does their distribution hold for European prehistory and history? (Keep in mind as well that as a result of the colonial expansion of the 16th to 19th centuries, more people speak Indo-European languages than speak those of any other language group.)

The traditional view of the spread of the Indo-European languages holds that an Ur-language, ancestor to all the others, was spoken by nomadic horsemen who lived in what is now western Russia north of the Black Sea near the beginning of the Bronze Age. As these mounted warriors roamed over greater and greater expanses, they conquered the indigenous peoples and imposed their own proto-Indo-European language, which in the course of succeeding centuries evolved in local areas into the European languages we know today.

In recent years, however, many scholars, particularly archaeologists, have become dissatisfied with the traditional explanation. I have analyzed the arguments for the accepted view and found them unconvincing. In this article I offer a different view, based on new insights into how cultural change comes about. According to this view, the spread of the Indo-European languages did not require conquest. On the contrary, it was likely to have been a peaceful diffusion linked to the spread of agriculture from its origins in Anatolia and the Near East. This proposed solution, which differs markedly from the accepted one, has profound implications for European prehistory and for linguistic studies of Indo-European languages.

COLIN RENFREW is Disney Professor of Archaeology at the University of Cambridge and Master of Jesus College there. He has conducted archaeological excavations in Greece and in the British Isles. His previous articles for *Scientific American* were on radiocarbon dating, the obsidian trade and the megalithic monuments of Europe. The arguments made here are presented in greater detail in Renfrew's book *Archaeology and Language: The Puzzle of Indo-European Origins,* published by Cambridge University Press in 1988.

The starting point of the problem of the origins of Indo-European is not archaeological but linguistic. When linguists look at the languages of Europe, they quickly perceive that these languages are related. The connections can be seen in vocabulary, grammar and phonology (rules for pronunciation). To illustrate the relatedness in vocabulary, it is sufficient to compare the words for the numbers from one to 10 in several Indo-European languages [*see illustration on page 109*]. Such a comparison makes it clear that there are significant similarities among many European languages and also Sanskrit, the language of the earliest literary texts of India, but that languages such as Chinese or Japanese are not members of the same family.

More detailed comparisons enable linguists to subdivide the languages of Europe further, into families. The earliest family to be recognized, that of the Romance languages—all known to be descended from Latin—includes French, Italian, Spanish, Portuguese and Romanian. The Slavonic languages include Russian, Polish, Czech, Slovak,

ÇATAL HÜYÜK is an early agricultural site on the central Anatolian plain in

Serbo-Croat and Bulgarian. The Germanic languages include German, Norwegian, Danish and Swedish. These families, which link the most closely related tongues, can themselves be regrouped to form the larger Indo-European language family. Indeed, only a few European languages (Hungarian, Finnish and Basque, for example) are not members of this family.

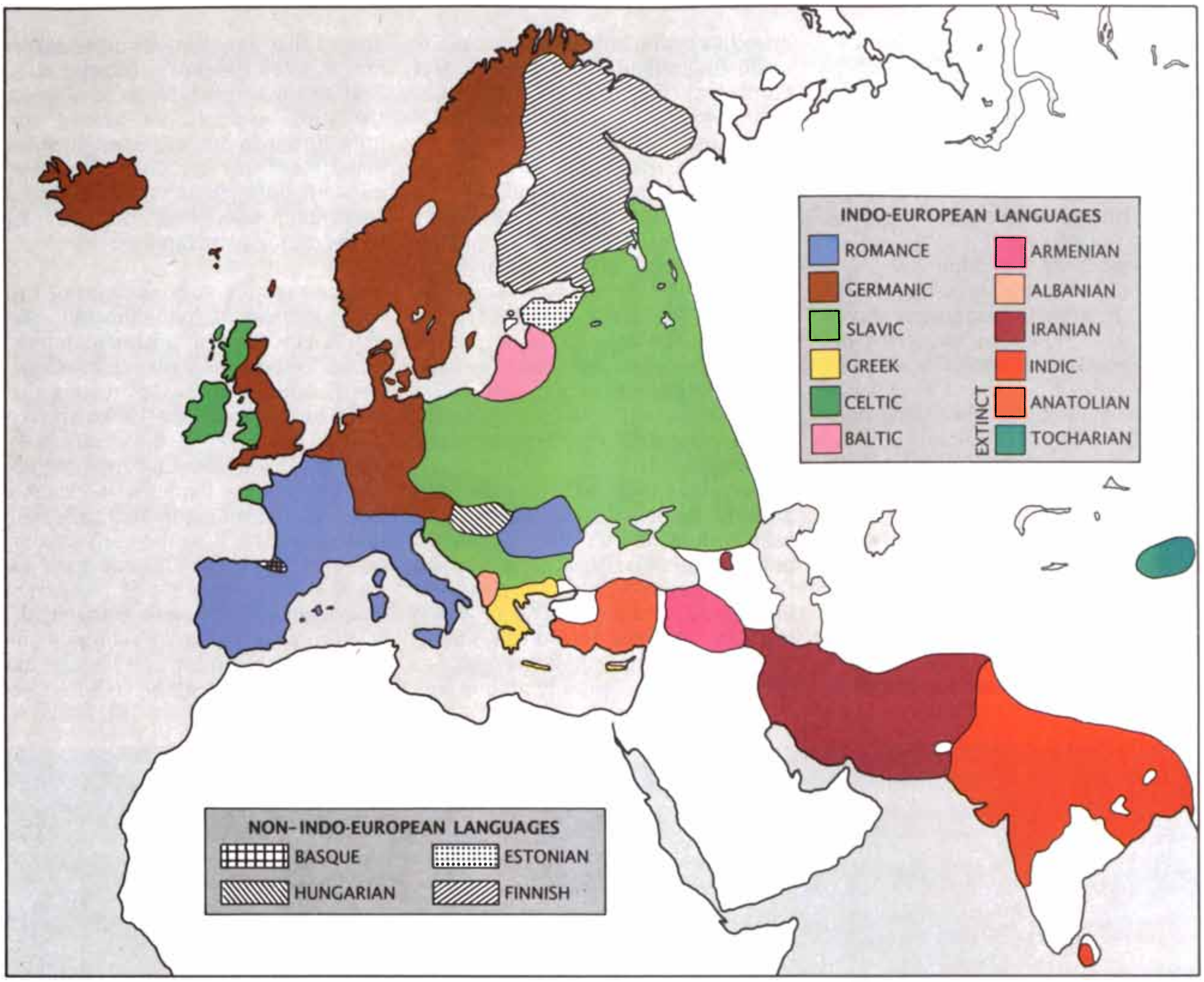
How did this complex arrangement arise? The Romance languages served as the first model for answering the question. Even to someone with no knowledge of Latin, the profound similarities among Romance languages would have made it natural to suggest that they were derived from a common ancestor. On the assumption that the shared characteristics of these languages came from the common progenitor (whereas the divergences arose later, as the languages diverged), it would have been possible to reconstruct many of the characteristics of the original proto-language. In much the same way it became clear that the branches of the Indo-European family could be studied and a hypothetical family tree constructed, reaching back to a common ancestor: proto-Indo-European.

This is the tree approach, which was pioneered in the early 1860's by the German philologist August Schleicher; it provides the framework for the way most historical linguists still think about the development of language families. The basic process represented by the tree model is one of divergence: when languages become isolated from one other, they differ increasingly, and dialects gradually differentiate until they become separate languages.

Divergence is by no means the only possible tendency in language evolution. Only a decade after Schleicher proposed his tree hypothesis, another German linguist, Johannes Schmidt, introduced a "wave" model in which linguistic changes spread like waves, leading ultimately to convergence—that is, growing similarity among languages that were initially quite different. In 1939 the Soviet linguist N. S. Trubetskoy went so far as to suggest that the resemblances among the Indo-European languages might have come about this way. Today, however, most linguists dismiss Trubetskoy's suggestion and think primarily in terms of linguistic family trees.

Although the starting point of the problem of Indo-European origins may be linguistic, its solution certainly entails archaeology, which offers a means of testing the linguistic hypotheses. The archaeologists began tackling the problem in the early 1900's. At that time the archaeology of the Romans, Greeks and Celts was becoming increasingly well understood back to the early historical period, perhaps as far back as the first millennium B.C. Moreover, archaeological finds were being made that promised to extend scholarly understanding deep into prehistory, indeed back to Paleolithic (Old Stone

modern Turkey. Farming arose here and at nearby sites by about 7000 B.C.; not long afterward it began to spread north into Europe. The author contends that the prototypical form of the Indo-European language spread along with agriculture.



INDO-EUROPEAN LANGUAGES are distributed from Ireland to India. Almost all the languages of Europe fall into this family. Among the exceptions are Finnish and Hungarian, in the Finno-Ugric group, and Basque, which exists in splendid isolation.

Age) times in most regions of Europe. It began to seem plausible that such material evidence might make it possible to trace the cultures of the speakers of European languages back to their origins.

In those early days of archaeology it was assumed that most significant cultural changes came about as the result of migrations of entire peoples, or tribes. Migration routes could be traced, it was thought, by means of characteristic weapons, tools and pottery left along the way. It was also assumed that a specific assemblage of artifacts—what archaeologists call a "culture"—could document a specific tribe of people with their own language. In this way the movements of tribes, as tracked by the archaeological record, would account for the dispersal of the early Indo-European languages. The question then became one of finding the original "homeland" of the Indo-Europeans and tracing their dispersals, using the record of archaeological cultures.

The search for this homeland proved controversial, however, and the discussion has not always remained purely academic. For most of this century German scholars have generally preferred an Indo-European homeland in northern Europe. Some of their work was taken up by the Nazis in an effort to prove that the original Indo-European language had been spoken by a master race of "Aryans" in Germany. The Semitic languages, which form a different group, were associated with a race thought by the Nazis to be inferior. From such garbled misuse of linguistics and anthropology sprang part of the sordid rationale for the Third Reich. Not surprisingly, in recent years scholars have trod carefully (if at all) in the field.

Yet the concept of a northern European homeland for Indo-European was not the most influential in scholarly circles. In 1926 V. Gordon Childe of the Royal Anthropological Institute in London published a book called *The Aryans* in which he argued for a homeland in the steppe areas north of the Black Sea, in what is now Russia, sometime between the late Neolithic period and the beginning of the Bronze Age, which was well established in some parts of Europe by 3000 B.C.

Childe's text comprised both archaeological and linguistic arguments, the latter being particularly ingenious. They drew on the success of linguists in establishing a "core" vocabulary that was common to many Indo-European languages. This core, it was inferred, had survived from the proto-Indo-European language spoken in the homeland. "Core" words for plants and animals were employed to obtain a picture of the environment in which early speakers of the language lived. Other words provided a means of dating the formation of the proto-language. There was no core word for iron or bronze, but there were words for the horse and the wheel. Therefore, it seemed that the initial dispersal of the Indo-Europeans must have taken place before the beginning of the Bronze Age but after the time when horses were domesticated and wagons introduced.

Childe went on to link these linguistic ideas with the archaeological evidence. In particular, he focused on Corded Ware (a type of pottery decorated by pressing cords into the wet clay), which is widely found in sites dating to the beginning of the Bronze Age. In northern and eastern Europe this pottery is often discovered—accompanied by stone battle-axes—in earthen mounds, called kurgans in Russian, that served as tombs for prominent men. Childe proposed that these artifacts were the material remains of groups of nomadic pastoralists, armed and on horseback, who migrated from their steppe homeland north of the Black Sea at the inception of the Bronze Age. They were, in short, the Indo-Europeans.

In recent years this argument has been set forth in impressive detail by Marija Gimbutas of the University of California at Los Angeles. Drawing on the evidence described by Childe and buttressing it with recent findings, Gimbutas reconstructed a series of "Kurgan invasions" flowing west from the lands north of the Black Sea. This view has now been widely accepted by historical linguists. Many archaeologists have also come to accept it, and other archaeological arguments are now frequently tailored to make them conform to the Kurgan-invasion hypothesis. Yet to my mind it is not a satisfactory story.

My reasoning is severalfold. In the first place, the archaeology is not convincing. Today many archaeologists see the Corded Ware burials as essentially local phenomena in which prestige goods were buried with members of emerging local aristocracies. Nor is the argument from the core words a strong one. Some of the so-called core words for plants and animals may have changed in meaning over time; in any case they are not necessarily specific to a particular geographic area. Words that provide the basis for datings are likewise suspect. Robert Coleman of the University of Cambridge has questioned the notion that the words for the wheel and the horse were actually part of a "protolexicon" prior to a generalized dispersal.

Perhaps the strongest objection is simply the lack of conviction behind the whole story. Why on earth should hordes of mounted warriors have moved west at the end of the Neolithic, subjugating the inhabitants of Europe and imposing the proto-Indo-European language on them? What enormous upsurge of population on the steppes could have been responsible? Although its construction is elegant, the story does not ring true for this listener.

The underlying problem, I believe, is that insufficient attention has been paid to the question of how change in language might actually be reflected in the archaeological record. Many of the traditional arguments, as I mentioned above, tend to equate a given assemblage of artifacts with a supposedly well-defined group, such as a tribe. Archaeologists now realize, however, that it is they who recognize and define archaeological "cultures" and that the equation with supposed tribes is problematic. What is more, the further equation between a people, so defined, and a particular language or language group is much less than a straightforward proposition.

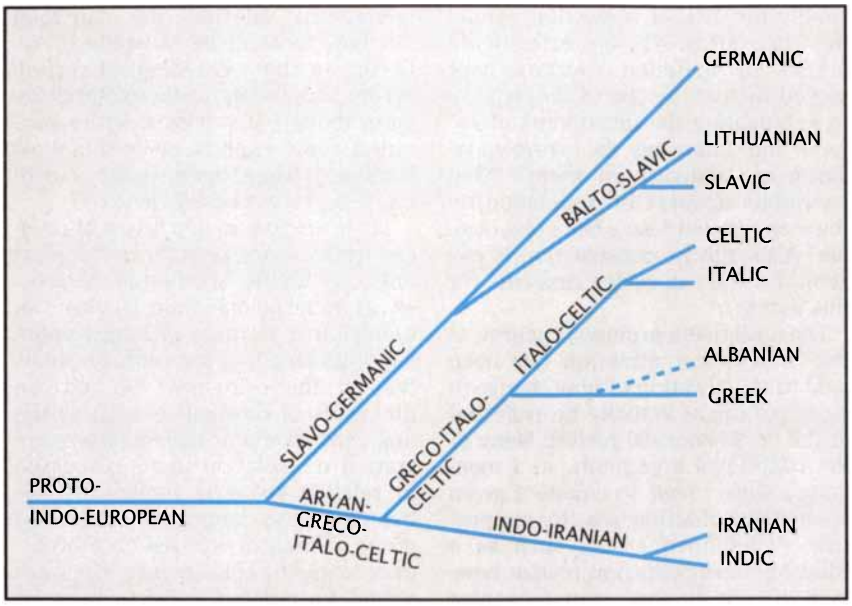
In my view one should avoid equating a particular style of pottery such as Corded Ware with a people or with a language. Instead the analysis should focus on processes of cultural change. What social, economic and demographic processes, it should be asked, might be correlated with changes in language? Having answered that question, one can then ask how those changes would be reflected in the archaeological record. Before proceeding to the material record, though, it is necessary to construct some explicit models of how language change might occur according to a process-based view.

There are four main classes of models, which space permits me to treat only very briefly. The first is the process of initial colonization, by which an uninhabited territory becomes populated; its language naturally becomes that of the colonizers. Second are processes of divergence, such as the linguistic divergence arising from separation or isolation that I discussed in relation to early models of the Indo-European languages. The third group of models is based on processes of linguistic convergence. The wave model, formulated by Schmidt in the 1870's, is an example, but as I said, convergence models have not generally found favor among linguists.

Now, the slow and rather static operation of these processes is complicated by another factor: linguistic replacement. That factor provides the basis for a fourth class of models. In many areas of the world the languages initially spoken by the indigenous people have come to be replaced, fully or partially, by languages spoken by people coming from outside. Were it not for this large complicating factor, the world's linguistic history could be faithfully described by the initial distribution of *Homo sapiens sapiens*, fol-

ENGLISH	OLD GERMAN	LATIN	GREEK	SANSKRIT	JAPANESE
ONE	AINS	UNUS	HEIS	EKAS	HIITOTSU
TWO	TWAI	DUO	DUO	DVA	FUTATSU
THREE	THRIJA	TRES	TREIS	TRYAS	MITTSU
FOUR	FIDWOR	QUATTUOR	TETTARES	CATVARAS	YOTTSU
FIVE	FIMF	QUINQUE	PENTE	PANCA	ITSUTSU
SIX	SAIHS	SEX	HEKS	SAT	MUTTSU
SEVEN	SIBUM	SEPTEM	HEPTA	SAPTA	NANATSU
EIGHT	AHTAU	OCTO	OKTO	ASTA	YATTSU
NINE	NIUN	NOVEM	ENNEA	NAVA	KOKONOTSU
TEN	TAIHUM	DECEM	DEKA	DASA	TO

WORDS FOR NUMBERS from one to 10 show the relations among Indo-European languages and the anomalous character of Japanese, which is not part of that family. Such similarities stimulated interest in the origins of Indo-European languages.



TREE MODEL of the origins of the Indo-European languages is based on divergence from a common root: the proto-Indo-European language. This diagram was devised in the 1860's by the German linguist August Schleicher, pioneer of the tree approach.

lowed by the gradual, long-term workings of divergence and convergence.

It seems to me that linguistic replacement has a key role to play in explaining the origins of the Indo-European languages. The archaeological record indicates that Europe has been populated continuously since far back in the Old Stone Age. Hence initial colonization is unlikely to supply much of the answer. It seems unlikely that simple divergence could explain the complex pattern of relations seen among the European languages. The suggestion of unity through convergence, proposed by Trubetskoy, has been widely rejected. Almost by default it seems that a model based on linguistic replacement is necessary. The Kurgan-invasion model does fall into that category, but, as I said, it is not wholly satisfactory. What are the alternatives?

There are several ways one language might replace another in a specific region. The first primarily entails demographic and economic processes. The existing population of the region will generally have a well-established subsistence economy. Whether the economy is based on hunting-and-gathering or on farming, it will already have begun to approach its appropriate "carrying capacity." If a group of newcomers is to establish itself by peaceful means, it must have a technology that will enable it to exploit a different ecological niche or compete successfully in the same one. Only in such cases will the incoming population expand enough so that their language begins to predominate.

Other forms of replacement are possible. When the incoming group is well organized and possesses superior military technology, it may be able to take over the existing social system and dominate by force of arms. In such cases the newly dominant elite may impose their own language. Such elite dominance, however, requires the fulfillment of several preconditions. One is the possession of a superior military technology. Another is that both social orders—of the inhabitants and of the occupiers—must have a high degree of organization. The incoming group must be organized in order to exert dominance; the indigenous society must be highly organized if it is to be taken over at all.

The Kurgan invasions would be a good example of the elite-dominance model if it could be shown that these criteria had been fulfilled. But that is not likely. The supposed military advantage of the Kurgan warriors (the fact that they were mounted) is hypothetical, since it is not clear that there were mounted warriors at the time. Furthermore, it has yet to be shown that either the incoming invaders or the inhabitants of Europe were sufficiently highly organized before the beginning of the Bronze Age to have undergone such a process. Indeed, it is likely that pronounced social stratification emerged in Europe only with the Bronze Age; in the preceding Neolithic period most societies are likely to have been largely egalitarian.

Two other forms of replacement ought to be mentioned, at least in passing. When a highly centralized society collapses, peoples formerly kept under control beyond the frontier can take advantage of the power vacuum and move in, as they did at the end of the Roman Empire. In the cases of such system collapse, the language of the invading "barbarians" may supplant that of the imperial center.

Alternatively, when long-distance trade builds up in an egalitarian society, a trading language, or lingua franca, often develops. A pidgin language (a simplified version of a language originally spoken outside the territory in question) is an example. When a pidgin language begins to be spoken as a mother tongue by some inhabitants, it is termed a creole, and creolization, which is a type of replacement, is now regarded as an important aspect of linguistic development.

Applying these forms of linguistic replacement—demographic change, elite dominance, system collapse and lingua franca—to European history and prehistory brings us considerably closer to our goal. Elite dominance and system collapse both require a degree of social organization that was probably lacking prior to the Bronze Age. It is unlikely that any European trading system before the Bronze Age was sufficiently intensive to foster the development of a lingua franca. That leaves the demographic and subsistence models. If one surveys European prehistory there is an event wide-ranging and radical enough in its effect to be a candidate, and that event does indeed fall squarely into the subsistence category: the coming of farming.

In the seventh millennium B.C. a novel agricultural economy began to spread across Europe, based on the cultivation of wheat and barley and the herding of sheep and goats. These species did not grow wild in Europe; they were imported. If one traces their ancestry back across Europe to the nearest region where their prototypes then existed in the wild state, one comes to central Anatolia, which today is part of Turkey. Actually, the domestication of these species seems to have taken place at about the same time in several adjoining regions of the Near East, but Anatolia is most relevant here because it was from there that the new domesticates reached Europe.

What was this spread like in demographic terms? Albert J. Ammerman and Luca L. Cavalli-Sforza at Stanford University have offered an elegant answer in the form of a model they call the "wave of advance." Their model presupposes that the agricultural economy was carried by local movements of farmers and their offspring.

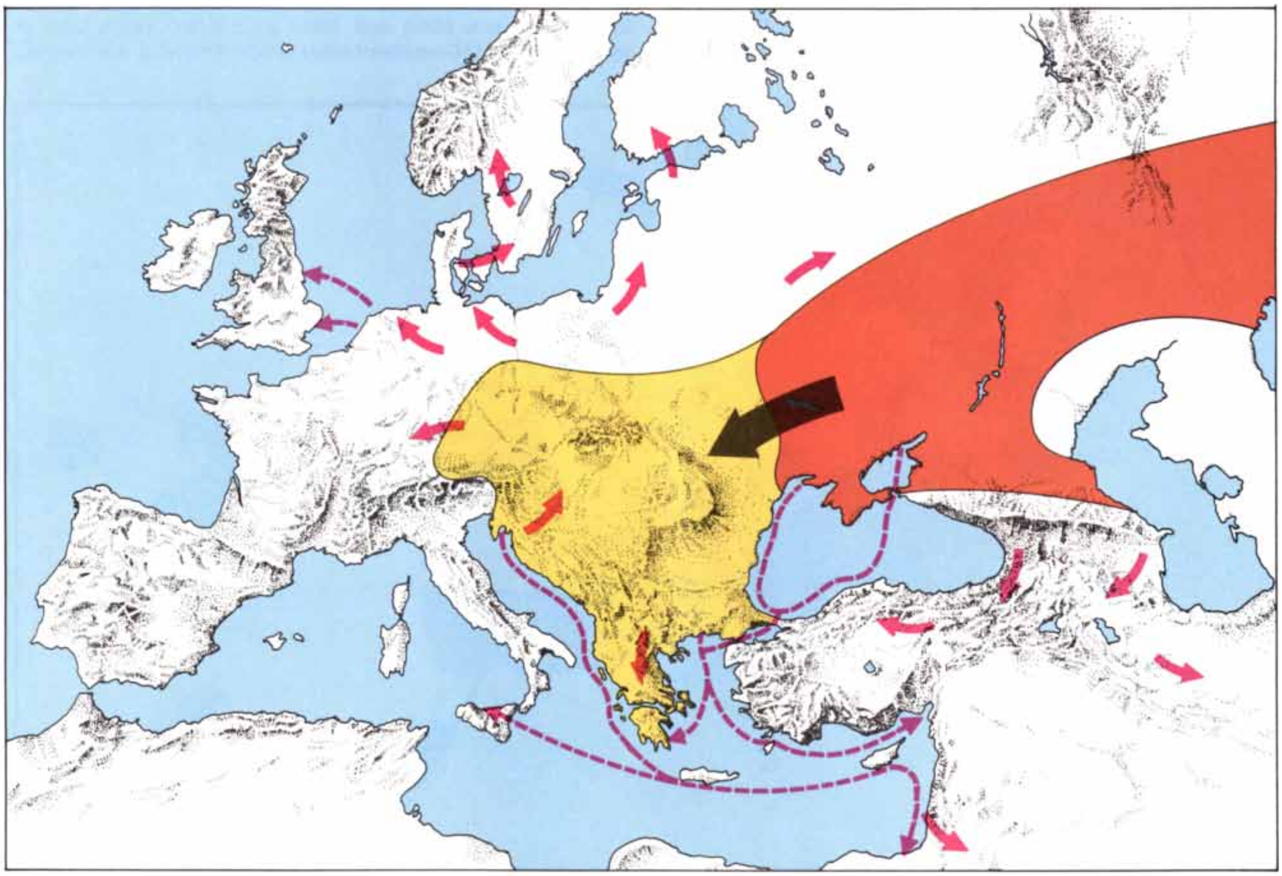
Once agriculture reached any given area, the population density there would have increased rapidly. Indeed, Ammerman and Cavalli-Sforza note that farming could have increased the population density 50-fold over the one person per 10 square kilometers that was probably typical of early hunter-gatherer economies. The increase in population density has potent consequences in the wave-of-advance model.

Ammerman and Cavalli-Sforza assume an interval between generations of 25 years. They further assume that on coming of age each farmer moves 18 kilometers (in a random direction) from the parental homestead to establish his own farm. With those assumptions providing the basis for calculations, Ammerman and Cavalli-Sforza show that agriculture would have spread across Europe as a wave progressing at an average velocity of one kilometer per year. At that pace it would have taken about 1,500 years for the farming economy to reach northern Europe from Anatolia, which accords quite well with the available archaeological evidence.

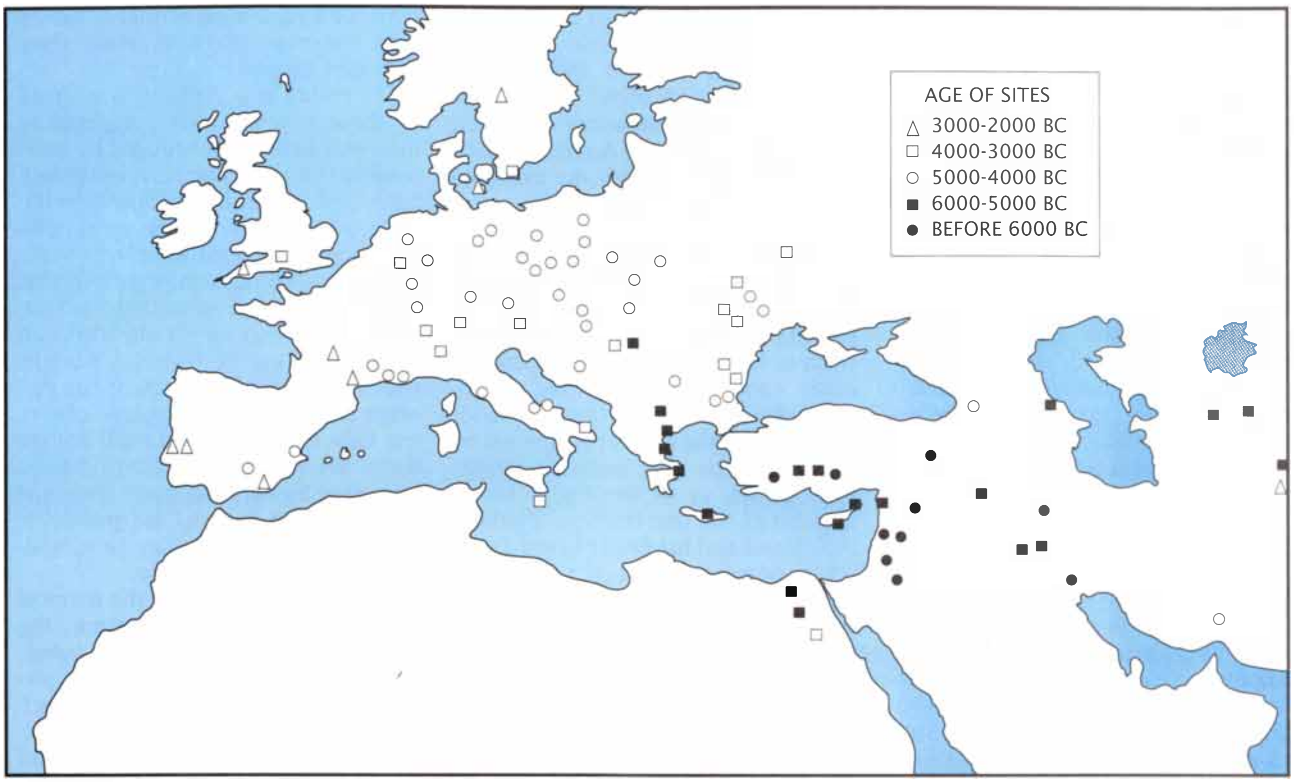
Of course, no single model could adequately describe a social process as complex as the coming of farming to Europe. Variations in terrain and in climate, among other conditions, imply that reality will differ significantly from the model. Moreover, the wave of advance is by no means the only possible applicable model, as archaeologist Marek Zvelebil of the University of Sheffield in England and his father Kamil Zvelebil, a linguist of Czech origin now based in the Netherlands, have pointed out. If the local hunter-gatherer population adopted farming from their neighbors, farming might have spread somewhat more slowly and without linguistic replacement, because the farmers would be the natives with their new economy rather than the newcomers speaking their own, new tongue.

The reality was probably a mixture of these two processes. Agriculture may well have been brought by newcomers to Greece and then to the Balkans, central Europe and southern Italy. In other regions, however, agriculture may have been adopted by the indigenous population, a pattern that would explain the anomalous persistence of several non-Indo-European languages. One of these is Basque, which survives to the present day. Another is the Etruscan language of central Italy, which survived until Roman times. Several other shadowy tongues, including Iberian, the early language of Spain, and Pictish, the pre-Celtic language of Scotland, may be amenable to similar explanations.

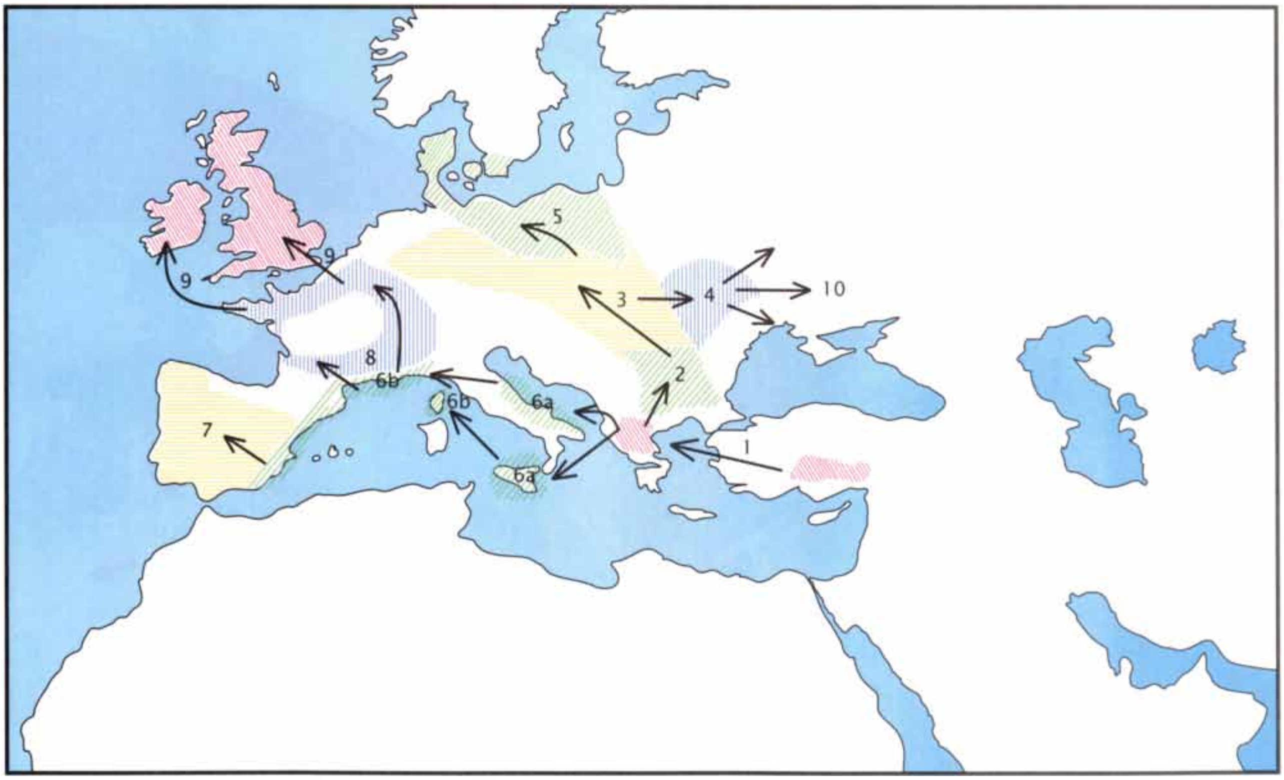
Whatever the details of the entry of agriculture into specific regions, the process as a whole provides a coherent alternative to the conventional picture of how the Indo-European language came to Europe. The new picture differs markedly from the old



"KURGAN INVASION" hypothesis pictured the original Indo-Europeans as mounted warriors ranging out from a homeland north of the Black Sea (*orange*) beginning in about 4000 B.C. This map is based on one drawn by Marija Gimbutas of the University of California at Los Angeles. The first wave of invasions, according to the model, brought the warriors to Greece by about 3500 B.C. Thence they spread north and south; colored arrows show their movements after about 2500 B.C.

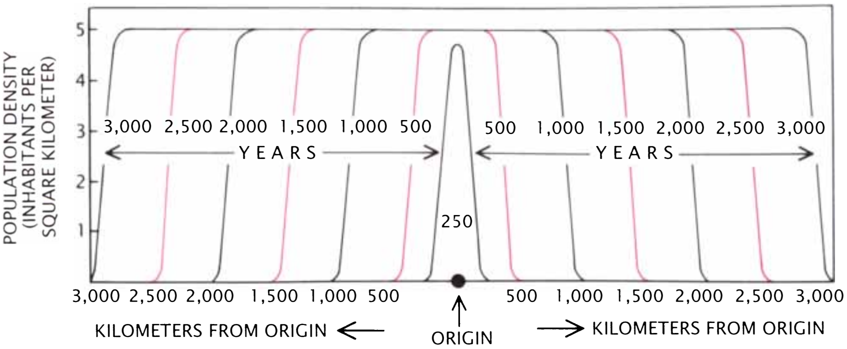


DIFFUSION OF AGRICULTURE over Europe from its zone of origin in the Near East took a little more than 2,000 years. The map depicts sites where remains of crops typical of early farming have been found. The early-crop assemblage reached Greece between 6000 and 5000 B.C.; 1,000 years later it was distributed throughout what is now Germany and Poland.



SEQUENCE of linguistic transformations has been proposed by the author as a parallel to the diffusion of agriculture; here each transformation is designated by a number. The initial transformation (*1*) was from the early agricultural culture of Anatolia, home of the proto-Indo-European language, to that of central Greece, where the language ancestral to Greek later developed. Each transformation after that led in turn to the subsequent formation of a new language or language group.

one. Its immigrants come from Anatolia rather than from the steppes and at a date (6500 B.C. or so) several thousand years earlier than has generally been suggested. My hypothesis also implies that the first Indo-European speakers were not invading warriors with a centrally organized society but peasant farmers whose societies were basically egalitarian and who in the course of an entire lifetime moved perhaps only a few kilometers.



WAVE-OF-ADVANCE MODEL, formulated by Albert J. Ammerman and Luca L. Cavalli-Sforza at Stanford University, simulates the effects of farming on population density. Agriculture can support densities many times greater than those of hunter-gatherer economies. As farming takes hold in new regions, even small movements of farmers away from family homesteads to set up their own new farms will cause agriculture to spill over into new territories in a "wave of advance." These curves measure population density in relation to the origin of agriculture as simulated by the model.

This hypothesis has some significant corollaries for the prehistory of the steppe region of Russia and for European prehistory in general. Indeed, my model would reverse the direction of influence between the steppes and western Europe as outlined by Childe and Gimbutas. In the new hypothesis one would expect farming and early Indo-European speech to reach the steppe lands of Russia from the west, rather than the other way around. And there is evidence of early farming villages in the Ukraine whose wheat and barley was almost certainly acquired from the west: from the Balkans, where cereal farming had arrived from Anatolia by way of Greece. Hence the first speakers of proto-Indo-European in the steppe region probably communicated in a tongue of Anatolian origin that had already passed through Greece and the Balkans by the time it reached the lands north of the Black Sea.

In a more general sense, if the arrival of Indo-European in Europe is pushed back to 6500 B.C., then there is considerably greater continuity in European prehistory than has previously been believed. There was no sudden discontinuity at the beginning of the Bronze Age, as represented by the "Coming of the Indo-Europeans" described in many textbooks of prehistory. Nor was there a discontinuity in the Iron Age as has often thought to be represented by the arrival of the Celts in northern Europe. The Celtic language would have evolved in western Europe from Indo-European roots. Rather than an alien group obliterated by the Indo-Europeans, the people who built Stonehenge and the other great megalithic monuments of Europe were Indo-Europeans who spoke a form of Indo-European ancestral to the Celtic languages of today.

In this light the whole early history of Europe appears as a series of transformations and evolutionary adaptations on a common proto-Indo-European base augmented by a few non-Indo-European survivals. The story is not predicated on a series of migrations from without but on a series of complex interactions within a Europe that was already fundamentally agricultural in economy and Indo-European in language.

So far I have concentrated on Europe, but the hypothesis that the spread of language is linked to the dissemination of agriculture has implications well outside that continent. The archaeological evidence makes it clear that Anatolia was not the only region in which early domestication took place. The zone of origination of agriculture had at least two other lobes, more or less self-contained regions within the larger zone. These were the Levant, a strip some 50 to 100 kilometers wide on the Mediterranean coast of what is now Jordan, Lebanon, Syria and Israel, and the Zagros region of Iraq and Iran [*see illustration on next page*].

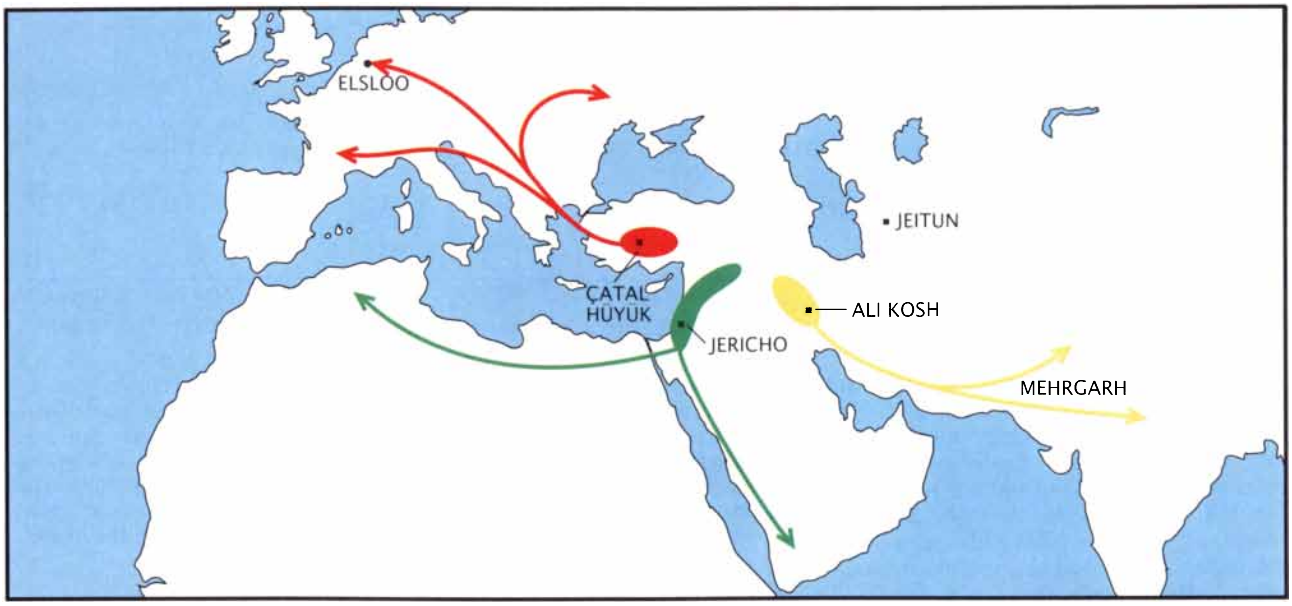
Given that the wave-of-advance model of demic diffusion is based chiefly on the capacity of farming to increase population density, it should be expected that wherever agriculture originates, a wave comparable to the European one should radiate outward. Where the Levant is concerned, the terrain dictates that such a wave would move south into the Arabian peninsula and west into northern Africa. In the case of the Zagros region, the wave would likely move to the southeast and the east, into Asia.

Now, there is accumulating evidence that farming did come to Africa north of the Sahara not long after its arrival in Europe. I would like to propose that it arrived there by a process of demic diffusion that mirrors the one in Europe. But what of the linguistic aspect of the process? In much of northern Africa the dominant linguistic group is the Afro-Asiatic, which includes Ancient Egyptian and the Berber languages as well as the Semitic group, which are sometimes thought to have originated in Arabia. It is possible, however, that all these languages can be traced back to a proto-Afro-Asiatic root in the Levantine lobe of agriculture's zone of origin.

Turning to the third lobe, the one based in the Zagros, one would predict the propagation of the farming economy east across southern Iran and as far as Pakistan. In this connection it is interesting that the linguist David McAlpin of the University of London has recently shown that Elamite, a language known to have been spoken in the ancient kingdom of Elam (now part of Khuzistan in southwestern Iran) is related to the Dravidian languages of India. It may be that the southeastern wave of advance carried the ancestor of the Elamite and Dravidian languages across to India and Pakistan. Later the proto-Dravidian tongue would have been displaced by the Indo-European languages that are now spoken in India.

This somewhat expanded version of the wave-of-advance model has the effect of situating the ancestral languages of the Indo-European, Afro-Asiatic and Dravidian groups quite close together in the Near East about 10,000 years ago. Although still hypothetical, this picture finds remarkable support from recent work in linguistics and genetics.

More than 20 years ago the Soviet linguists Vladislav M. Illich-Svitych



ZONE OF ORIGINATION of agriculture had three "lobes," each of which may have given rise to a great family of languages by diffusion. The Anatolian lobe, containing Çatal Hüyük, may have been the cradle of the Indo-European languages. A second lobe, containing Jericho, may have been the homeland of the languages of Egypt and northern Africa. A third lobe, containing Ali Kosh, may have been the source of a group of languages in India and Pakistan that were later replaced by languages of the Indo-European group. Elsloo, Jeitun and Mehrgarh are early farming sites in these three great diffusional pathways. The processes that are depicted on the map are hypothetical but are supported by recent findings in linguistics and genetics.

and Aron Dolgopolsky proposed that a number of Eurasian language families, including, among others, the Indo-European, the Afro-Asiatic and the Dravidian, are related in a "superfamily" they called the Nostratic. The recognition of superfamilies, which may represent a breakthrough in linguistics, is still regarded as controversial. Indeed, the work of the two Soviet scholars is only now becoming known in the West. It is notable, however, that they have also suggested Anatolia as the homeland of the proto-Indo-European language. Since I was unaware of their views when I made my proposal, the convergence of views is striking.

That convergence is buttressed by some new genetic findings by the research group of Cavalli-Sforza and that of Allan C. Wilson of the University of California at Berkeley. Both teams employed statistical methods to analyze blood groupings of living populations and infer their genetic affinities. They conclude that there is a close genetic relationship among the speakers of the Afro-Asiatic, Indo-European and Dravidian languages, among others. Their findings thus harmonize with the Nostratic hypothesis and perhaps with the view that the coming of farming is intimately linked to the formation and distribution of present-day languages.

Taking the final step back and adopting the most global of perspectives, the proposal that one can make meaningful statements about proto-languages and language groupings as far back as 10,000 B.C. may ultimately open the way to a better understanding of the whole phenomenon of human linguistic diversity. Most (although by no means all) scholars today believe that the comprehensive linguistic ability seen in human populations emerged with *Homo sapiens sapiens,* the anatomically modern form of our species. New findings from Israel and southern Africa suggest that the transition to *Homo sapiens sapiens* took place about 100,000 years ago. Not long after that date modern humans were probably spreading out of Africa and populating large regions of the globe. This biological evolution and dispersal provides the framework within which human language and linguistic diversity must be explained.

It would be wrong, however, to assume that the last word has been spoken here. Actually, although I have ended with global considerations, I began with a relatively limited end in view: criticizing the accepted explanation of the origin of the Indo-European languages. My provisional proposal of an early Anatolian origin does find some validation in recent linguistic and genetic research. The final picture will no doubt be more complex than the one I have sketched, involving several historical episodes and a variety of theoretical models. Yet I predict that when a more complete understanding is achieved, the spread of farming from Anatolia into Europe will prove to be a significant part of the story.

FURTHER READING

THE ARYANS: A STUDY OF INDO-EUROPEAN ORIGINS. V. Gordon Childe. Alfred A. Knopf, Inc., 1926.

INDO-EUROPEAN AND INDO-EUROPEANS. Edited by George Cardona, Henry M. Hoenigswald and Alfred Senn. University of Pennsylvania Press, 1970.

THE NEOLITHIC TRANSITION AND THE GENETICS OF POPULATIONS IN EUROPE. Albert J. Ammerman and L. L. Cavalli-Sforza. Princeton University Press, 1984.

ARCHAEOLOGY AND LANGUAGE: THE PUZZLE OF INDO-EUROPEAN ORIGINS. A. Colin Renfrew. Cambridge University Press, 1988.

NOSTRATIC. Mark Kaiser and V. Shevoroshkin in *Annual Review of Anthropology,* Vol. 17, pages 309-329; 1988.

RECONSTRUCTING LANGUAGES AND CULTURES: ABSTRACTS AND MATERIALS FROM THE FIRST INTERNATIONAL INTERDISCIPLINARY SYMPOSIUM ON LANGUAGE AND PREHISTORY, ANN ARBOR, NOVEMBER 8-12, 1988. Edited by Vitaly Shevoroshkin. Studienverlag Dr. Norbert Brockmeier, 1989.

IN SEARCH OF THE INDO-EUROPEANS: LANGUAGE, ARCHAEOLOGY AND MYTH. J. P. Mallory. Thames and Hudson, 1989.

THE AMATEUR SCIENTIST

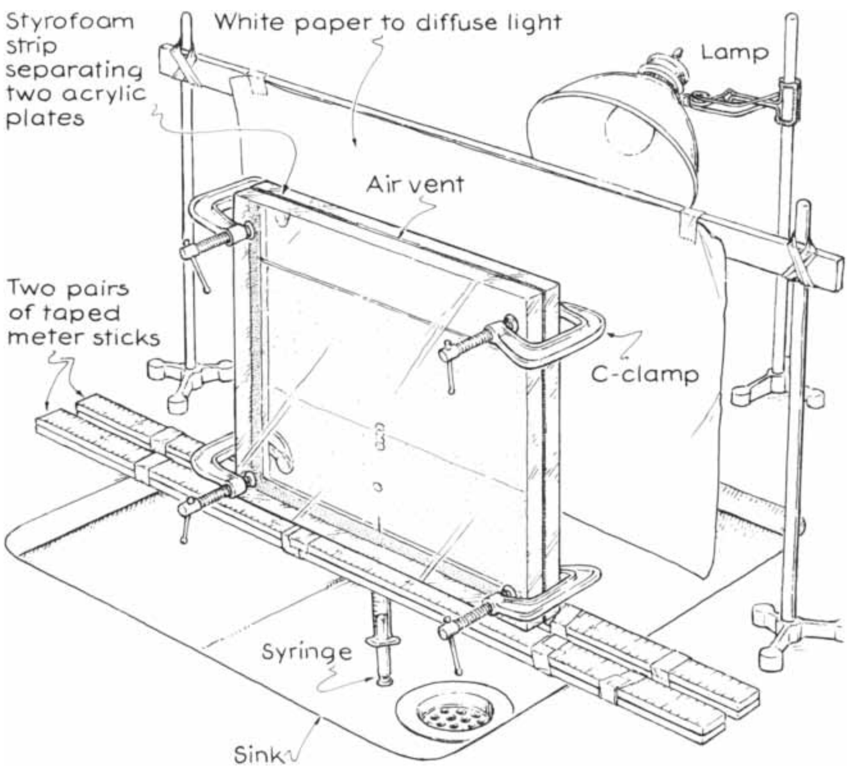
How to build a Hele-Shaw cell and watch bubbles playing tag in a viscous fluid

by Jearl Walker

Air bubbles rising through a viscous fluid sometimes take on curious shapes, combining and interacting in perplexing ways. Such behavior often goes unnoticed because it is hard to observe the bubbles in an ordinary container, particularly when the fluid is opaque. One way to enhance the visibility and at the same time somewhat reduce the complexity of the various goings-on is to place the fluid in a Hele-Shaw cell, a device named for Henry S. Hele-Shaw, the English engineer who devised it around the turn of the century.

The cell consists primarily of two transparent plates separated by a narrow gap. A thin spacer runs along the internal edges of the plates to maintain their separation and keep the fluid from leaking out. Air bubbles are introduced into the cell through a port along one of the edges. The fluid can be pushed or pulled through the cell by a pump connected to other ports. Alternatively, the cell can simply be propped up at a slant or mounted vertically so that gravity and buoyancy move the fluid and the bubbles.

The most obvious advantage of a Hele-Shaw cell is that the bubbles are always visible. Even if the fluid in bulk is opaque, a thin layer of it is transparent because it hardly absorbs any light. Moreover, the flow is effectively two-dimensional and so is easier to analyze than when it is three-dimensional in a wider container.



The experimental setup with the Hele-Shaw cell

Hele-Shaw cells have been used extensively for investigating various features of bubble formation and fluid flow, but it was not until recently that some rather curious features of bubbles were noticed. In 1986 Tony Maxworthy of the University of Southern California reported that when a series of bubbles were released individually in a viscous silicone oil, they tended to queue and then move up through the fluid like a wagon train. Often a "bubble stack," as Maxworthy called it, became unstable, the lead bubble splitting down its center. As the stack continued to ascend, the instability and splitting progressed down through the length of the stack, bubble by bubble, until there were two separate stacks.

In additional research Maxworthy and several other investigators studied the shapes air bubbles assume when they move through glycerin and other viscous fluids in a Hele-Shaw cell. You might think that surface tension would keep a bubble always circular or slightly oval, but actually a number of odd shapes happen to be stable.

I wondered if the bubble stacks and odd bubble shapes might be attainable in a homemade Hele-Shaw cell. From a local store specializing in plastic supplies I purchased two square acrylic plates that measured 15 inches on a side and were 3/4 inch thick. To separate the plates I used narrow, thin strips of Styrofoam that are sold in hardware stores for weatherproofing windows. The strips come with a sticky side that adheres well to the acrylic. I applied strips along three edges on the face of one plate. Along the fourth edge I applied two shorter strips so that a wide central hole was left between them to serve as an air vent.

I next laid the plate on paper towels and poured or squirted a fluid onto the surface inside the strips to cover about half of the area. Then I placed the second plate on the first one, taking care to align their edges. I squeezed the plates by tightening four C-clamps along their left and right sides, one near each corner. The plates were then separated by less than a millimeter, but the width varied across the cell because of the pressure of the fluid. The fluid filled most of the cell.

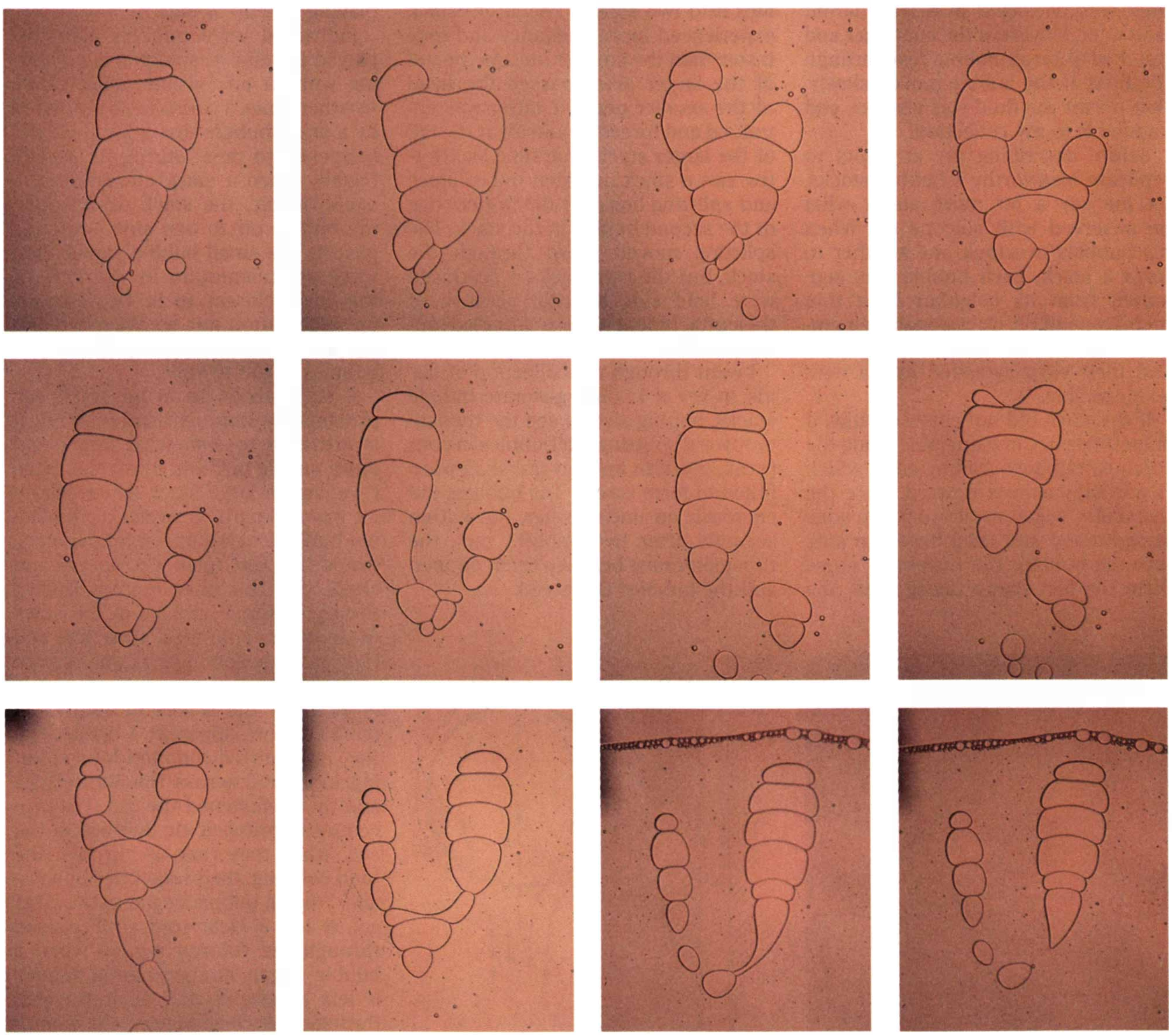
I planned to let gravity drive air

bubbles through the fluid, and so I needed to mount the cell upright with the air vent at the top. Because some of the fluids I employed leaked slowly through the bottom Styrofoam strip, I worked over a sink. To support the cell I fashioned a bridge from four metersticks. They were taped in pairs with wide plastic tape, and the pairs were laid across the sink with a small separation between them. I balanced the cell on top of the two-beam bridge, with the cell's bottom Styrofoam strip centered on the gap between the beams [*see illustration on opposite page*]. On the other side of the sink I put a flood lamp clamped to a laboratory stand. To diffuse its light I suspended a wide sheet of white paper between the lamp and the cell.

When I placed the cell on the bridge, often some air bubbles were already in the fluid, having been trapped when I clamped the plates. Each bubble was outlined with a distinct dark line, whereas the regions within the bubble and elsewhere in the cell were uniformly lit. The dark line was caused by refraction of the light as it passed through the shallow, curved surface of the bubble that straddled the gap between the plates. The refraction sent the light rays off to one side of my field of view, which left the border darker than the other regions. The bubbles were so stark that they resembled a penciled sketch.

To make additional bubbles I passed the needle of a syringe through the gap between the bridge beams and then carefully forced it up through the bottom Styrofoam strip in the cell. The syringe was filled with air, and pushing on the plunger injected a small amount of the air into the fluid, where it formed a bubble. The bubble gradually rose away from the syringe and traveled up through the cell. Injecting a large bubble was easy, but producing small ones was more difficult: the resistance of the fluid required a strong push on the plunger before the fluid yielded—with a sudden "give" that often produced a bubble larger than I wanted. I found that if I pulled back on the plunger before the bubble left the tip of the syringe, I could remove some of the air and thereby shrink the bubble.

To photograph the bubbles I magnified them by mounting on my 35-millimeter camera three lenses from a close-up kit. During the photography I switched off the room lights so that only the diffused light from the



Sequence of photographs **(left to right, top to bottom)** ***of a bubble stack splitting and recombining***

flood lamp illuminated the cell. The viscosity of the fluid was usually high enough, and the confines of the cell narrow enough, so that the bubbles moved sedately, allowing me to bring them into focus and take a photograph before they shifted appreciably. (The slow migration of the bubbles also meant that each test of a fluid in the cell required at least an hour of observation and sometimes much more.) I had some film that was color-balanced for the flood lamp, but often I used film that was easier to find but was not balanced; it yielded photographs with an unnatural tint, as may be evident in some of the accompanying illustrations.

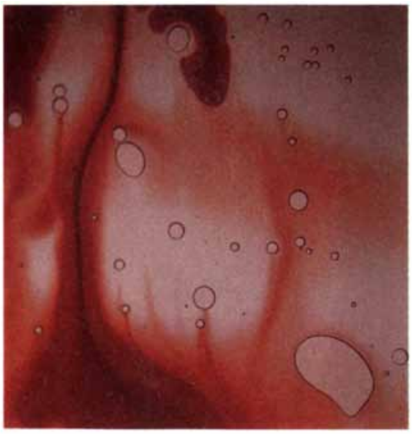
For the tests I collected a variety of fluids. I found glycerin in a drugstore and then explored the shelves there and in several food stores for other appropriate materials. When a fluid was packaged in a transparent container, I inverted the container and watched as an air bubble rose through the fluid. If the bubble moved slowly, that meant the fluid was viscous, and I added it to my collection.

Before describing my attempts to replicate Maxworthy's bubble stacks, let me say a bit more about what he observed with silicone oil. When his bubbles overtook one another to form a stack, each bubble was separated from its neighbor by a thin membrane of oil. Occasionally a membrane failed: two bubbles coalesced, and their sides wrestled into a new, wider bubble.

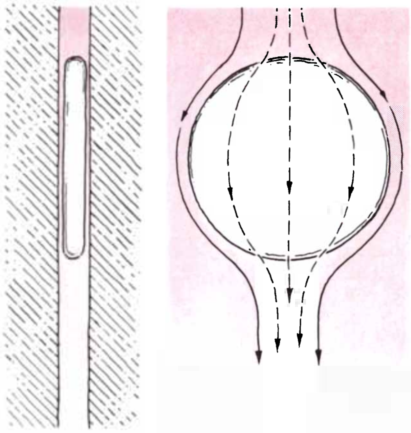
If the stack did not coalesce fully, it sometimes became unstable along the topmost bubble's leading edge, which is normally convex upward. Once the instability began to develop, the edge straightened and then began to sink into the bubble. The instability arose from the fact that a dense fluid (the oil) overlay a less dense fluid (the air) along the edge. If surface tension along the edge kept the edge upwardly convex for a time, it kept the oil from sinking into the bubble. As the bubble advanced through the oil, however, tiny waves played along the leading edge. Sometimes one of the waves grew large enough to upset the stable curvature of the edge, and then the edge began to collapse. A collapse could also be initiated if the edge happened on a small bubble that was almost stationary. As the edge approached the small bubble, the fluid pressure between them increased and could become large enough to initiate a collapse.

If the collapse extended to the trailing edge of the top bubble, the bubble split into two parts. The collapse and splitting then marched through the rest of the stack, separating the bubbles into two stacks. The larger stack experienced more buoyancy and rose faster than the smaller one. As the tail of the larger stack passed the head of the smaller one, the latter was entrained and forced to join up at the tail of the larger stack. One time Maxworthy saw a stack in which the collapse and splitting began at the upper edge of the second bubble in the stack. The splitting moved down through the stack, but the two stacks it produced were held side by side because of their attachment to the unaffected topmost bubble.

I went through my collection of fluids to see if I could generate bubble stacks, hoping also to see the tendency toward splitting. Air bubbles in corn syrup, glycerin and several shampoos failed to form stacks. The bubbles did encroach on one another, but within seconds after two bubbles met, the thin membrane between them popped and the bubbles coalesced.



The trails left by bubbles



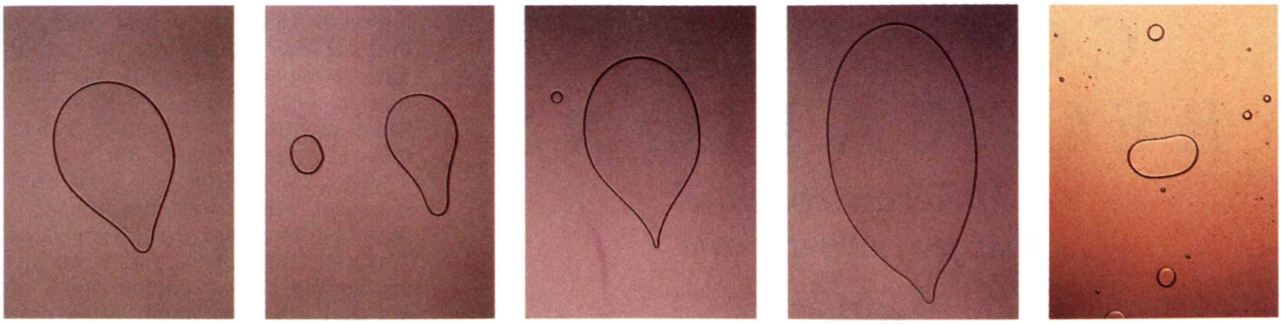
The formation of a trail

I had better luck with some other shampoos. The best was a product called Shower Gel, which consists of propylene glycol, glycerin and soluble collagen, along with several other ingredients. The series of photographs on the preceding page shows how one stack behaved. The top bubble in the complex stack was initially unstable but then restabilized on the left side of the stack. Meanwhile, the instability passed to the second bubble, which collapsed and split. The bifurcation of the bubbles then traveled down through the stack until finally a stack of two bubbles remained on the right side. The longer stack on the left side then began to split at the top. As it rose, the short stack was pulled into its tail and finally underwent collapse and splitting. As I watched this play of bubbles in the partial darkness of the room, I thought of a band of trilobites playing a game of tag.

Individual large bubbles also displayed instability, sometimes collapsing without any visible provocation. At other times a collapse was initiated by a small bubble that a large bubble happened to meet during its ascent. Usually when a stack encountered a small bubble, the stack would push the bubble off to one side and slide past it. The small bubble would then force an indentation in the side of the stack closest to it. Occasionally the indentation was so extensive that the side of the stack collapsed, and its bubbles began to split.

A stack tends to avoid small, approximately stationary bubbles, but it is attracted to somewhat larger and more mobile bubbles. In several cases I watched a large stack deviate from its upward path to overtake individual bubbles off to one side; the stack would actually pick up speed and chase after the other bubbles. During the chase some bubbles in the stack extended and thinned, their left and right sides becoming straight or even sinking inward.

To make the wakes of individual bubbles more apparent, I opened up the cell, squirted common food coloring in a band across the Shower Gel and then refastened the cell. The procedure introduced air bubbles in the cell. When they rose up through the food coloring, their wakes became visible. When a bubble began in the clear gel, it left a clear trail as it passed through the colored region; when a bubble began in the colored region, it left a colored trail as it traveled through the clear region; the photograph at the left shows this effect. Usually the length of one of these



Some shapes of bubbles in glycerin and* (right) *in a shampoo

trails was several times the diameter of the bubble that created it.

I believe the trails are left by fluid that is slowly discharged from thin layers next to the bubble. Consider a vertical section that extends from plate to plate and through a bubble [*see drawing on opposite page*]. The air in the bubble is separated from each plate by an intermediate layer of fluid. The layers are so thin that during the bubble's ascent the fluid can pass through them only slowly. As fresh fluid enters the top part of each layer, the fluid in the bottom part of the layer is discharged below the bubble. Now consider a vertical section that is parallel to the plates and extends through the bubble. As the fluid passes the left and right sides of the ascending bubble, it converges just below the bubble. The convergence compresses the fluid that is discharged from the thin layers.

Suppose that the bubble travels from the clear gel into the colored gel. The fluid coming around the left and right sides is then colored, but the gradual discharge from the thin layers remains clear for a while. The flow of colored gel into the region below the bubble compresses the discharge into a thin clear trail that lies along the path taken by the bubble. A similar story accounts for the colored trail a bubble leaves when it travels from colored gel into clear gel.

Several investigators have found that an air bubble traveling through a viscous fluid can take on slightly different shapes depending on circumstances. The tug-of-war between surface tension, buoyancy and fluid pressure tends to keep a bubble circular or elliptical, but it can instead be flattened on both the leading and trailing edges or be shaped like a pear, with the leading edge noticeably wider than the trailing edge. The flat and pear shapes are typically seen when the cell is horizontal and the fluid is forced through it.

Last year Anne R. Kopf-Sill and George M. Homsy of Stanford University reviewed the theory behind these bubble shapes and added observations of new shapes that were even curiouser. They found bubbles that had tails—some of them short and rounded and others much longer than the diameter of the main part of the bubble, with a small circle at the tip. Kopf-Sill and Homsy also were the first to sight a shape, which had been predicted from theory, in which the trailing edge is flat and the leading edge is slightly sunken.

I tested the fluids in my collection, looking for the various shapes that had been reported. Very small bubbles in glycerin, pancake syrup, corn oil, Shower Gel and two other shampoos called Agree and Prell were circular. Somewhat larger bubbles were ellipses whose long axis was vertical.

In one case, working with Prell, I spotted what might have been an example of a bubble with a sunken leading edge, but my attention had been focused on other bubbles at the time, and so I cannot be sure that the bubble had not simply been stretched to one side and distorted by a larger bubble that had passed it. Large bubbles in syrup, corn oil and glycerin often developed short tails once they got going. The tails were persistent but wiggled slightly. Strung across them was a faint border that seemed to complete the generally oval boundary of the bubble.

I concentrated on the bubbles in glycerin. When one bubble overtook another, the membrane between them lasted only seconds before the air in the bubbles broke through. What remained in the breached region were shallow, curved ridges of glycerin that clung to both plates. As the newly formed bubble ascended, the ridges descended and their left and right ends distorted the bubble noticeably, rendering it pear-shaped.

Sometimes a tiny spot of glycerin appeared inside a bubble. It may have been caught up on an imperfection in one of the plates, or it could have been clinging to some contamination in the cell. When the spot reached the lower edge of the bubble, it shoved the edge outward. As the tail of the bubble neared the height of the spot, the tail extended over to it, as if it were being wagged.

Although I usually worked with the cell standing upright, I sometimes wanted to slow the ascent of the bubbles by tilting the cell. To do so, I laid a large sheet of white paper on a tabletop, placed the lower end of the cell on the paper and propped the upper end up on two boxes. I angled the lamp so that its light scattered from the paper up through the cell. The lower end of the cell was at the very edge of the table, and so I could still inject it with air with a syringe.

If you build your own Hele-Shaw cell, you might test other viscous fluids such as honey, an ungelled gelatin dessert or motor oils. (A warning about the particularly viscous motor oils: they not only make a mess but also are extremely difficult to clean out of the cell when you have finished with them.) If you cannot locate a syringe for your work, you can entrap air bubbles in a fluid simply by lifting up the top plate and then clamping it down again before the fluid has a chance to level out.

FURTHER READING

BUBBLE FORMATION, MOTION AND INTERACTION IN A HELE-SHAW CELL. T. Maxworthy in *Journal of Fluid Mechanics*, Vol. 173, pages 95-114; December, 1986.

STABILITY OF BUBBLES IN A HELE-SHAW CELL. S. Tanveer and P. G. Saffman in *Physics of Fluids*, Vol. 30, No. 9, pages 2624-2635; September, 1987.

THE AMATEUR SCIENTIST. Jearl Walker in *Scientific American*, Vol. 257, No. 5, pages 134-138; November, 1987.

BUBBLE MOTION IN A HELE-SHAW CELL. Anne R. Kopf-Sill and G. M. Homsy in *Physics of Fluids*, Vol. 31, No. 1, pages 18-26; January, 1988.

COMPUTER RECREATIONS

A Tinkertoy computer that plays tic-tac-toe

by A. K. Dewdney

"I first had that experience [universality of computation] before I went to school. There weren't any [computers] yet, but we had toy construction sets. One was called TinkerToy.... What's strange is that those spools and sticks are enough to make anything."

—MARVIN MINSKY,
in preface to *LogoWorks*

How many of us remember Tinkertoys, those down-home kits of colored wooden sticks and spools with holes in them? Amid our childhood constructions of towers or cranes, how many of us pondered the outer limits of the Tinkertoy world? Did we conceive of contraptions that reached the ceiling? Perhaps, but we lacked the kits or the time to make it happen. Such a Tinkertoy fantasy took place several years ago when a student group from the Massachusetts Institute of Technology constructed a computer entirely (well, almost entirely) out of Tinkertoys!

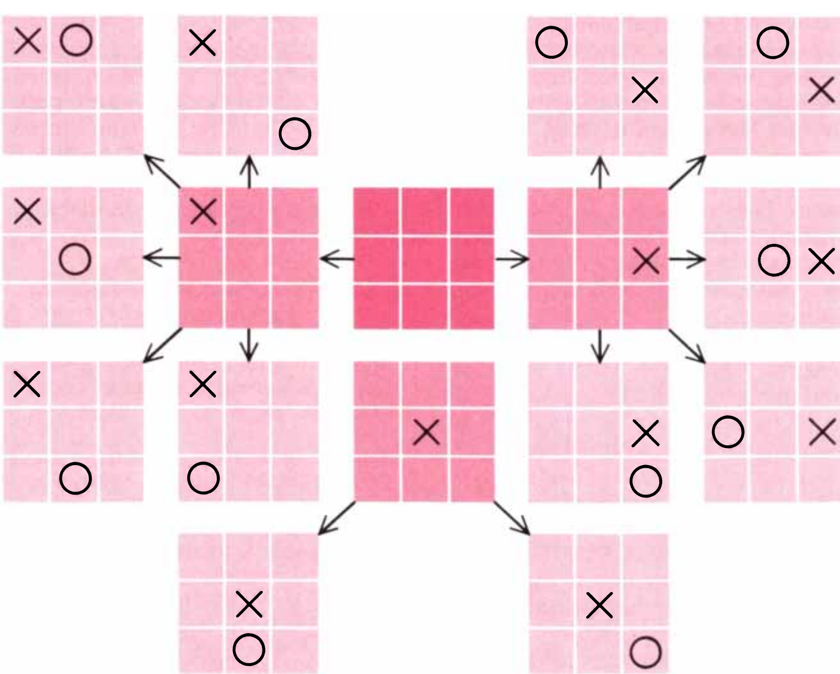
From a distance the Tinkertoy computer resembles a childhood fantasy gone wild or, as one of the group members remarked, a spool-and-stick version of the "space slab" from the movie *2001: A Space Odyssey.* Unlike the alien monolith, the computer plays a mean game of tic-tac-toe. A Tinkertoy framework called the read head clicks and clacks its way down the front of the monolith. At some point the clicking mysteriously stops; a "core piece" within the framework spins and then with a satisfying "kathunk" indirectly kicks an "output duck," a bird-shaped construction. The output duck swings down from its perch so that its beak points at a number—which identifies the computer's next move in a game of tic-tac-toe.

What precisely does the read head scan as it feels its way down the monolith? Nothing less than 48 rows of Tinkertoy "memory spindles" encoding all the critical combinations of *X*'s and *O*'s that might arise during a game [*see illustration on opposite page*]. Each spindle is a sequence of smooth spools connected axially by sticks and arranged in nine groups of three each, one group for each square of the tic-tac-toe board. The presence or absence of spools from a group indicates that a corresponding square of the tic-tac-toe board is vacant or is occupied by an *X* or *O*.

The Tinkertoy computer is not fully automatic: a human operator must crank the read head up and down and must manage its input. After the computer's opponent makes a move, the operator walks to the front of the machine to adjust the core piece inside the read head, registering the contestant's move. The operator then pulls on a string to cock the core piece for its impending whirl of recognition. When it discovers a memory that matches the current state of the game, the core piece spins, and the computer indicates its move.

The best way to understand how the machine works in detail is to recount the story of its creation at the hands of the M.I.T. students: Erlyne Gee, Edward Hardebeck, Daniel Hillis, Margaret Minsky and brothers Barry and Brian Silverman. Most of the group has long since graduated and entered various computer professions. Perhaps the best-known team member is Hillis. He was the moving force behind Thinking Machines, Inc., which produces the well-known parallel supercomputer called the Connection Machine. (Perhaps Tinkertoys have something to teach us.)

In 1975, when Hillis and Brian Silverman were in their sophomore year, they participated in a class project to build something digital from Tinkertoys. The students sat down to play. One made an inverter—a logic device that converts a binary 1 signal to a 0 signal and conversely. Another made an OR gate; if either of the device's two input signals happened to be a 1, then its output would also be a 1. It quickly became clear to the students that Tinkertoys were "computation universal," the theoretical term for a set of components from which a fully program-



The first three levels of the tic-tac-toe game tree

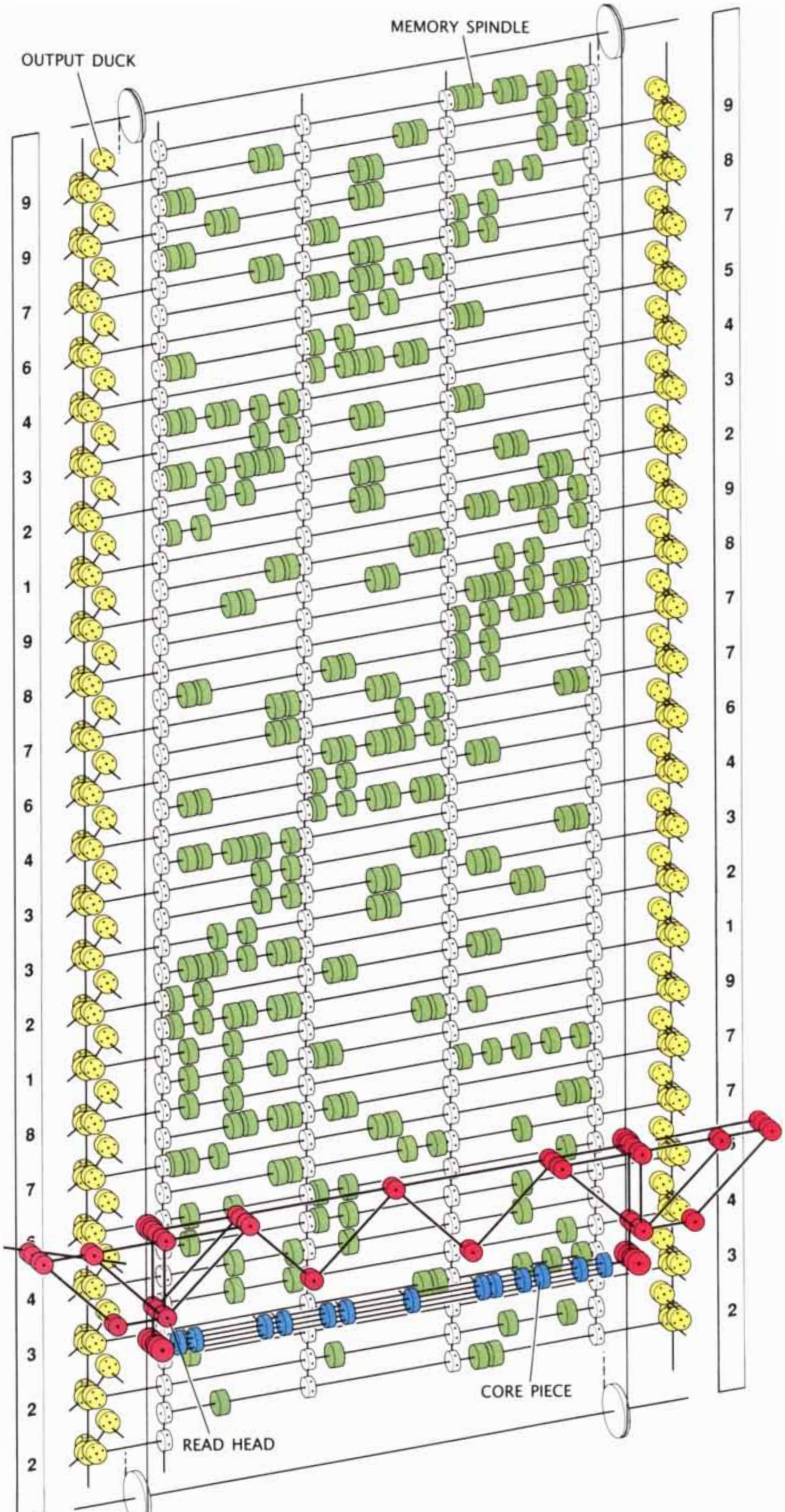
mable computer can be constructed. Theoretical possibility was one thing, the practical demands of money and time another.

The demands were met in a rather roundabout manner through Hillis's interest in robots. From time to time he had mused openly about building a robot. Word of his idea somehow reached the ear of Harry Loucks, then director of the Mid-America Center in Hot Springs, Ark. Would the students like to construct a robot as a display in the center's museum? The students agreed in principle, but the project seemed too complicated. Just then the old Tinkertoy dream resurfaced. Would the center like a computer made out of Tinkertoys instead?

Hillis and company set out to assemble the first Tinkertoy computer in a laboratory at M.I.T. The first model, unlike its successor, was a bulky cube with sides about one meter long. It was impressively complicated. Packed with logic devices made entirely of wooden sticks and spools, the machine signaled its moves by waving nine flags from the top of the framework. The prototype Tinkertoy computer had to be taken apart for the trip to Hot Springs, and once it was reassembled on site, the machine never quite worked properly again. On the other hand, it did make an intriguing exhibit. (It is currently on display at the Computer Museum in Boston.)

In 1979 Loucks contacted the group again. Could they make a new Tinkertoy computer, one that worked? By this time Silverman was in Ottawa and Hillis in Boston, each pursuing a new career. Over the telephone Hillis and Silverman worked out an improved design. It was to be reliable, and that meant simple. They decided to lay out all the possible tic-tac-toe boards in a row and devise some kind of reading mechanism that would move from row to row until it found a pattern matching the current board. The very act of recognition could trigger a preset response.

While Hillis contemplated ways to represent tic-tac-toe boards with digital Tinkertoy components, Silverman analyzed the game. To appreciate the complexities involved even in this childhood pastime, readers might consult the game tree shown on the opposite page. In the middle of the tree sits the initial board, a three-by-three grid empty of *X*'s and *O*'s. From this initial board nine new ones can arise, depending on which of the nine squares *X* plays. The figure shows just three possibilities; the remaining six are rotated versions. Each of the three



The Tinkertoy computer: ready for a game of tic-tac-toe boards at the second level gives rise to other cases. For example, the board in which *X* plays the center square and then another square results in two different boards. The other two boards at the second level each generate five new boards at the third level.

I pruned many branches from the tic-tac-toe tree by appealing to a symmetry argument: the excluded boards are merely rotations or reflections of the included ones. Symmetry seems simple to humans, but a computer must be programmed or wired to recognize it. In a world of Tinkertoy engineering, symmetry operations would require elaborate structures.

Silverman was dealing with a tree, therefore, that was many times larger than the fragment shown in the illustration. But perseverance paid off, especially when Silverman employed a computer program that analyzed the game of tic-tac-toe and discovered that a great many boards could be collapsed into one by a forced move. Suppose, for example, that two squares in a row contain *O*'s and the third is blank. The contents of the remaining two rows are irrelevant since an opponent must fill the third square with an *X* or lose the game.

Silverman was delighted when he tallied up the final total of relevant boards: only 48. For each of them he noted the appropriate move by the machine. The surprisingly short list of possible board positions heartened Hillis. The group converged on Hot Springs, Silverman says, "with the list of 48 patterns and only a vague idea of how to interpret them mechanically."

(Readers who have a fanatical bent—or are stranded in airline terminals—may enjoy working out the game tree on a few sheets of paper. How long does it take, after all, to draw 48 tic-tac-toe patterns? Four symbols should help sort things out: *X, O,* blank and a dash for "don't care.")

Once settled in Hot Springs, the team assembled the raw material for their spool-and-stick odyssey: 30 boxes of Tinkertoys, each containing 250 pieces. Some team members put together the supporting framework that would hold all 48 memory spindles. To explain precisely how the spindles were made, I must digress for a moment and describe the conventions employed by the team to encode tic-tac-toe positions.

First, the squares of a tic-tac-toe board were numbered as follows:

1 2 3
4 5 6
7 8 9

Then a memory spindle was divided conceptually into nine consecutive lengths in which information about the status of each tic-tac-toe square was stored from left to right.

Each length was further subdivided into three equal sections, one for each possible item one might find in a square: an *X*, an *O* or a blank. Each possibility was encoded by the lack of a spool. For example, if an *X* happened to occupy a certain square, the memory spindle would have no spool in the first position, one spool in the second and one spool in the third. Similarly, a spool missing in the second position denoted an unplayed square, and one missing in the third position symbolized an *O*. Finally, if all three spools were missing, it meant that what occupied the square was irrelevant.

One can hardly mention the subject of memory spindles without bringing up the core piece, a thing of digital beauty. Here the Latin *digitus* came into its own, the construction resembling a kind of rotating claw with nine fingers. The core piece and a sample memory spindle are shown in the illustration below.

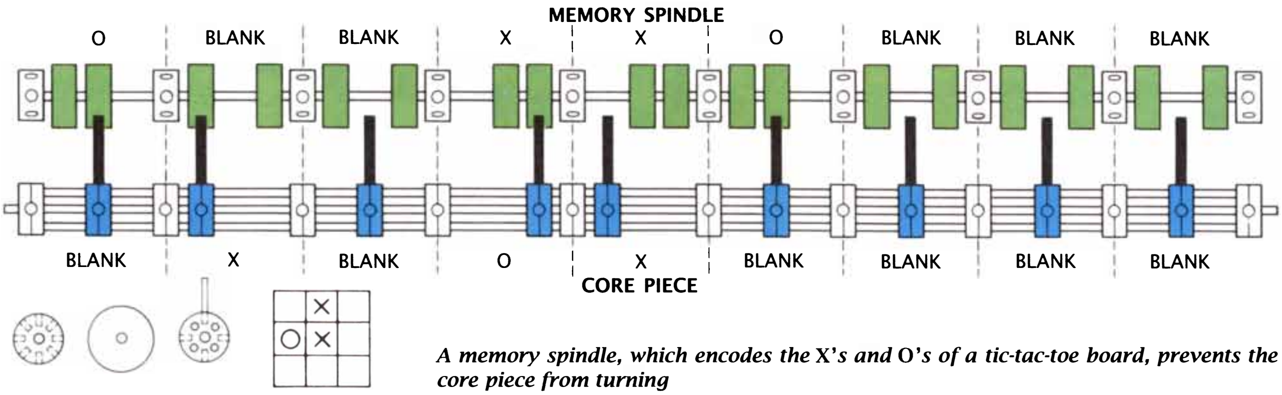
The core piece consisted of nine equal sections. Each had its own finger, a short stick protruding from the rim of a sliding spool. Within each section the finger could be moved along the axis of the core piece into any of three possible positions: one for *X*, one for *O* and one for blank. The core piece could therefore store any possible tic-tac-toe board by virtue of the positions of its nine fingers as moved by the operator for each play by human or machine. In the illustration below, fingers in the consecutive positions 2, 1, 2, 3, 1, 2, 2, 2, 2 would represent the board shown.

If the current situation of play is stored in the core piece, does the Tinkertoy computer require any other memory? Could spool-and-stick logic devices be strung together to cogitate on the position and ultimately to signal a move? Well, yes—but such a Tinkertoy computer would be complicated and immense. The memory spindles eliminated the need for most of the computer's cogitation. All the Tinkertoy computer had to do was to look up the current board in the memory spindles. The only purpose of the search, naturally, was to decide what move to make.

A glance at the illustration on the preceding page makes it clear that each memory spindle was accompanied by a number written on a paper strip hanging next to its output duck. These numbers were the machine's responses. As the read head clicks down the rows of spindles, the core piece wants to turn but cannot as long as at least one memory-spindle spool blocks one of the core piece's nine fingers. Only when the read head falls adjacent to the spindle that matches the current board do all nine fingers miss. Then the core piece whirls.

By a mechanism that would do Rube Goldberg proud, a stick protruding from the end of the core piece engages another stick connected to the output duck. The spinning core piece thus kicks the duck off its perch to peck at a number writ large on the paper strip.

Computer purists will ask whether the Tinkertoy contraption really deserves the title "computer." It is not, to



A memory spindle, which encodes the* X*'s and* O*'s of a tic-tac-toe board, prevents the core piece from turning

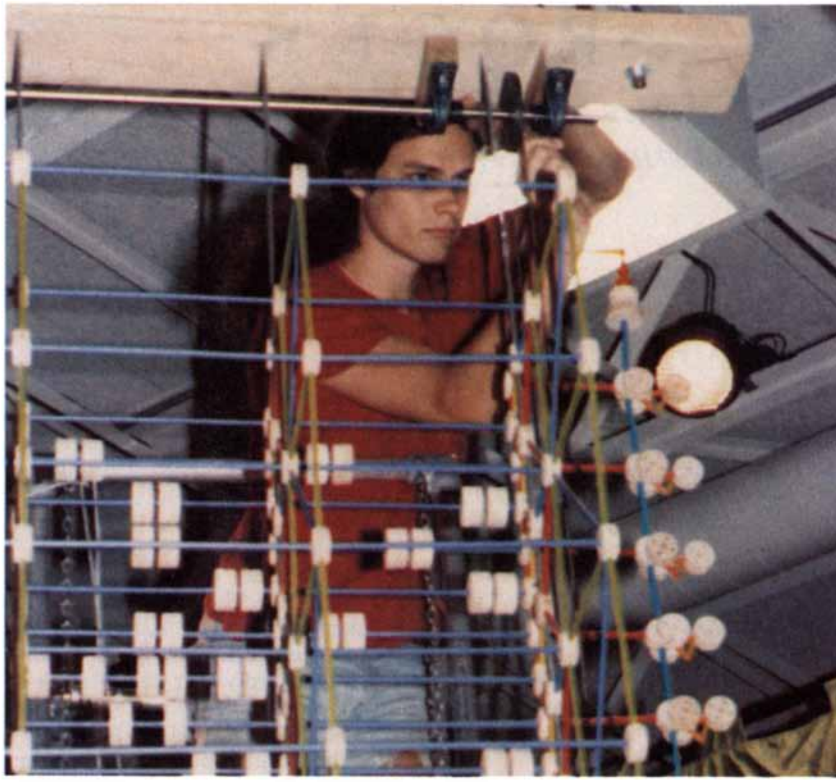
be sure, programmable in the usual sense: one cannot sit at a keyboard and type in a program for it to follow. On the other hand, one could certainly change the memory spindles, albeit with some difficulty, and thus reprogram the computer for other games. Imagine a Tinkertoy device that plays *go-moku narabe* (a game played on an 11-by-11 board in which one player tries to place five black stones in a row while preventing an opponent from creating a row of five white stones). A Tinkertoy computer programmed for *go-moku narabe,* however, would probably tower into the stratosphere.

The real lesson the Tinkertoy computer can teach us resides in a rather amazing feature of digital computation: at the very root of a computation lies merely an essential flow of information. The computer hardware itself can take on many forms and designs. One could build perfectly accurate computers not only of Tinkertoys but also of bamboo poles, ropes and pulleys [see "Computer Recreations," SCIENTIFIC AMERICAN, April, 1988], plastic tubes and water—even, strange to think, electrical components. The last-named are preferred, of course, because of their speed. It would be shortsighted indeed to sneer at a computer made of Tinkertoys merely because it is not electronic. After all, even electrons and wires may not be the best materials for quick computer processing. Photons and fibers are gaining on them fast.

Actually, Tinkertoys are well suited to digital computing. For example, the memory spindles use a binary principle: the presence or absence of spools denotes the status of a particular square on a tic-tac-toe board. The core piece exhibits digital logic: it can turn only if all its fingers miss corresponding spools on a memory spindle. Such an operation is called "and." One can trace the logic for the core piece in the illustration on the opposite page: if the first spool is absent from the first section of the memory spindle *and* the second spool is absent from the second section *and* the third spool is absent from the third section *and* so on—only if all nine conditions are met will the core piece turn. The beauty of the Tinkertoy computer is not just its clever mechanics but its subtle logic.

Tinkertoy purists will be happy to know that the M.I.T. students stuck to the original wooden sticks and spools with only a few exceptions. An occasional aluminum rod runs through the framework to strengthen it. Two wire cables, an axle and a crank transmit motive power to the awesome machine for its next move. Finally, the very joints of sticks and spools were made firm by glue and escutcheon pins—pieces of hardware that commonly hold commemorative plaques in place. The team inserted the pins in holes drilled through the rim of the spool down to the original, central hole and through its stick—a task they had to repeat more than 1,000 times. (When Hillis walked into a hardware store to obtain several thousand escutcheon pins, the manager looked bewildered. "We have," Hillis said with a straight face, "a lot of escutcheons.")

The Tinkertoy tic-tac-toe computer suffered the fate of most museum exhibits. It was taken apart and crated. It sits in storage at the Mid-America Center, waiting to reemerge, perhaps, into the limelight. It may yet click its way to victory after victory, a monument to the Tinkertoy dreams of childhood.



Edward Hardebeck helps to assemble the Tinkertoy computer

Well into my sixth year of "Computer Recreations," I am as painfully aware as ever that there are many things the department cannot do. It cannot, for example, teach readers how to program, nor can it mention the hundreds of fascinating programs and the many computer stories and ideas that readers send in, given the limitations of space and time. It took six years to discover a remedy to these and other needs: a newsletter. Its name is *Algorithm: The Personal Programming Newsletter,* and the first issue is now available.

The newsletter will appear bimonthly. It seeks to pack a lot of information between its covers. In particular it will have two columns for people who like to program. One will be for beginners and the other for more experienced practitioners. A "bulletin board" at the back of the newsletter will make some of the world's underground programs public for the first time. Letters, state-of-the-art-icles and speculative pieces will aim to lead the mind into unexplored territory. I shall be delighted to send a free sample of the first issue to anyone who writes to me in care of *Scientific American.*

FURTHER READING

CHARLES BABBAGE: ON THE PRINCIPLES AND DEVELOPMENT OF THE CALCULATOR AND OTHER SEMINAL WRITINGS. Charles Babbage et al. Edited by Philip Morrison and Emily Morrison. Dover Publications, 1961.

OPTICAL COMPUTING. Special issue edited by Sing H. Lee and Ravindra A. Athale in *Optical Engineering,* Vol. 28, No. 4; April, 1989.

BOOKS

Aeries at the edge, noisy signals, domestic science, fleshing out the bones

by Philip Morrison

OSPREYS: A NATURAL AND UNNATURAL HISTORY, by Alan F. Poole. Cambridge University Press, 1989 ($27.95).

It is a "highblue summer day" in the salt marsh. The white-breasted hawk perched above the big pile of sticks that is her nest whistles again and again that slow guard call. Soon the smaller male flies in, lugging the half-pound fish he has just snatched from the shallows nearby. The two ospreys, mated for life, are fondly rearing this year's three downy young.

These hawks, only slightly modified for fishing the surface waters, are perhaps the best-known birds of prey. "Live fish have comprised over 99 percent of the diet" of every osprey population that has so far been studied. Ospreys therefore rarely nest more than a couple of miles from the water's edge. They hunt on the wing, usually from a slow, circling, scanning flight that ends in a precipitate plunge, razor talons first—a near free-fall subtly guided by wing and tail. Divebombers, they cannot follow finny prey into the depths; they hold sway over only the uppermost yard of water. The birds thrive best in a shallow-water habitat, whether freshwater lagoons, salty bays, estuaries or lonely islets of lake and sea. Theirs is clearly the ecology of the edge.

Indeed, ospreys dwell on two margins at once. One edge is the physical frontier between water and land; the other is the changing boundary between wildlands and the territory under immediate human dominion. In Victorian Scotland the most famous place to see ospreys (and to rob their nests!) was archetypical: an aerie atop an abandoned castle ruin beside the dark and silent Loch an Eilein. For today's ospreys, photographs document a range of environments. One nest is atop a light tower in a busy urban parking lot along the Connecticut coast; four other nests are on successive tall steel pylons of one power highline that crosses the lake district of northern Germany. This independent fish hawk, although never tamed to falconer's hood or hand, is superbly adaptable.

"Few birds are more cosmopolitan." They are present on every continent save Antarctica. They nest from Swedish lakes to the coral islands of the Great Barrier Reef. They winter in the tropics but hardly ever nest there (except for Mexico, Belize and the northern coasts of Australia). Dr. Poole estimates the world population by region: a total of 30,000 pairs. Two thirds of these birds breed in North America, most of them along the myriad inland waters that in summer spangle the forests stretching from western Alaska to the Atlantic coast of Canada.

Fish cannot be caught easily in frozen waters, and every winter northern populations of ospreys migrate to the tropics. Decades of bird banding and recovery (the "easy" banding of these sinewy, fierce birds is described here—something of a tour de force) eventually disclosed their routes down the meridians to the wet tropics. From Scandinavia they migrate to the mangroves and rain forests of West Africa, aloft for one 50-hour leg above the fishless Sahara. From the eastern United States they overfly the Caribbean, with one island stopover, to the rain-forest rivers of northern Colombia and Venezuela.

Osprey life is easier under the tropical sun, according to energy budgets measured for breeding birds and for wintering birds. During a Massachusetts spring and summer the average male bird fed himself, his mate and three young. He caught six or eight fish a day. He spent about 60 percent of his share of the total energy intake on hunting and flying—more than three hours of hunting time per day. In Senegal a carefree bird, wintering alone, caught a couple of fish a day. Half an hour on the wing met daily caloric requirements. Yet ospreys do not breed anywhere in those hospitable wintering grounds. Apparently the numerous migrant ospreys somehow exclude any resident breeders; we do not know how.

But we know a good deal about ospreys, a knowledge gained from crisis. Two case histories are presented in detail. Like all the British birds of prey, the ospreys of Scotland were in steady decline over centuries from hunting and nest-site destruction, even though—being fish hawks—they threatened no farmer's flocks; it was guilt by association. By the mid-19th century the very fact that the striking birds were so rare had, in itself, begun to accelerate their demise. Keen naturalists collected osprey skins and eggs for sale as specimens; after the fall, one of those men wrote ruefully, "'I am afraid that Mr St John, yourself, and your humble servant, have finally done for the Ospreys.'"

The last Scottish nest was found in 1916. The Scandinavians increased the protection of their birds, and in the 1950's a few nesting pairs, probably emigrants from the Swedish population, were once again observed in Scotland. This time the contrite naturalists took a new stance. In 1981 the millionth osprey-watcher was ceremonially made welcome to the observation blind in the preserve that had been established around the nest of the first returnees; ospreys have become a Highlands tourist attraction, and the osprey population is growing fast.

The second case history ran its silent course along the coastal waters between New York and Boston. In the 1930's about 1,000 nesting pairs were counted there. We lack such data for a century earlier, but a plausible figure might be twice that; the larger American numbers absorbed the losses to hunters that extinguished the British birds. Around the salt marshes near the mouth of the Connecticut River, a kind of protected concentration grew up. It was noticed that nested ospreys drove away the hawks that took barnyard chickens, and so farmers often set up an old cartwheel on a pole as a nesting platform to attract the protectors.

Then in 1957 a Yale graduate student, Peter Ames, began to survey the mouth of the Connecticut again. He soon saw a catastrophic decline in osprey number, at the rate of 30 percent a year. At first predation was blamed, for in the marshes the birds nested on the flat ground. Protected nests safe from raccoons were built for the birds, but the decline continued. A series of studies, involving chemical analysis and control transfers of nestlings from less contaminated sites, showed that it was DDT that had reduced reproductive success for the osprey. The salt marshes had been heavily sprayed against the mosquito; the fish were badly contaminated, the birds still more.

British research first drew the connection between the pesticidal organochlorines and raptor eggshell thickness. A Swedish comparison of old museum eggshells with those collected by the early 1970's makes the statistical case eloquently. As shells thin by some 15 to 20 percent, osprey populations decline sharply: the fragile eggs break during incubation. (By now even the enzyme whose loss depleted the shell wall is known.) Sudden population drops occur because the shell thickness is most sensitive to the first five or 10 parts per million of the metabolic product of DDT; heavier doses make less difference. Now DDT use is banned in the U.S., and the ospreys seem to face little threat from dieldrin, PCB's, mercury or acid rain.

The ospreys nesting along the New England coastline now double in number every six or seven years; they are all but dooryard birds, nurtured by plentiful fish and by the provision of safe nesting platforms—many on buoys, pilings and poles and many more on specially built yard-wide platforms, the "ultimate birdhouse." Osprey nesting poles are in high demand among those humans who, like the osprey, enjoy summer residence on the New England shore but winter by choice in warmer places. On Martha's Vineyard there are now 45 active nests within sight of house, road or busy harbor—more osprey nests on that small island than in the wild lochs and streams of all of Scotland.

"It is this adaptability that defines the species." The ospreys of New England were sensitive indicators of pesticide contamination, although not of the intensity of economic development. Their tolerance to human settlement put them at grave risk, but the visibility that came from that very proximity saved them. The links of human concern and scientific study were essential: in saving the osprey, we had helped to secure our own tenure. An even wider adaptability—reason—defines *Homo sapiens;* given a chance, it will save us as well. Ospreys worldwide rightly share.

The author, 15 years a student of ospreys, tells his definitive story in engaging language, with the help of line drawings by Margaret LaFarge. He takes particular care to cite and clarify the evidence for just about every statement—a compelling style, the model of scientific exposition and absent in too many natural histories.

IMAGES OF THE ICE AGE, by Paul G. Bahn and Jean Vertut. Facts on File, Inc., 1988 ($35.00).

At once an enigma and a joy, the Franco-Cantabrian art of the Paleolithic, large and small, has gripped the imaginations of scholars and visitors alike over a full century. This striking new volume brings a reader into direct and critical contact with the work and thought of four men who have pondered over that expressive past.

The book opens with a brief foreword that introduces as persons the two authors of the work; it is by Count Robert Bégouën. Under his family lands lie decorated caves sealed by nature until they were visited in 1912 for the first time since the Paleolithic. The caves (one is linked underground with the huge cavern system of the Volp) have been lovingly and rigorously preserved ever since. The whole context there is safe: not only are the unmatched sculptures in place, but so too are the footprints of children, the discarded flint tools, the cave-bear jaws picked up long ago to remove the prized canines and thrown down again.

The book closes with an appendix by Alexander Marschack, the originator of microscopic techniques for the study of Paleolithic markings, who surveys the methods and approach of his friend the brilliant engineer Jean Vertut, long the best-known photographer of cave art. It was Vertut who made most of the new photographs for the definitive survey published by André Leroi-Gourhan in 1971. Vertut's untimely death in 1985 kept from us his choices and his comment on a hundred of his remarkable pictures in color presented here.

Paul Bahn, a tireless young English archaeologist, has written the eight chapters in between. They are a history and an up-to-date summary: flowing, logically acute, candid and unfailingly close to the evidence. They will set any reader into the midst of what we know (a lot) and what we do not know (even more) about the 275 decorated European sites and the thousands of portable objects that so beguilingly encode "our most direct contact with the beliefs and preoccupations of our ancestors."

How old is the art? It is hard to date wall markings. The slow growth of encrustations of calcite is not uniform, even within a single cave; such a translucent covering (of which the frontispiece shows a superb example) guarantees against historic fraud but provides dates only qualitatively. We have some carbon dates for mobile objects but not for the older finds, most of them. Style has been used to construct chronologies, as has the tally of extinct species of animals depicted; Bahn musters examples to indicate just how uncertain and subjective such schemes can be. In Altamira similar engraved deer heads are found on the walls and on carbon-dated bones; they fix a date of 13,550 B.C.

We can assign the art overall to the long stretch of human prehistory between about 35,000 and about 10,000 B.C. Famed Lascaux is dated to some 15,000 B.C. by charcoal found in the floor layers—rather like "dating a church by analysing the residue from its candles!" In any case that celebrated cave is not the "dawn of art" but lies plainly at least halfway along the time line of art (up either to the Sphinx or to Andy Warhol—which suggests the dating uncertainties).

Is this art unique to the region on either side of the Pyrenees? As a matter of fact it is worldwide. Even in the New World there are claims of dated and decorated artifacts in Mexico and in Brazil, some 10 to 20 millenniums old. Better authentication goes with certain colored drawings of figures of a rhinoceros and zebras from Namibia. Now Australia has begun to fill up the Paleolithic map: rich concentrations of cave engravings have been newly found (in South Australian pits with surface openings, their locations not made public for reasons of protection). These are not figurative but are finger markings, with circles and lozenges in profusion. Radiometric dating is in progress on the carbonate layers that bear these engravings; it looks as though they cannot be younger than 15,000 or 20,000 years. There is no puzzle here; all the decorations claimed are the work of our own people, our single sapient species.

If time and place hold such doubts, can we read the messages themselves? The record is gloomy. The scholars are Robinsons seeing their own footprints every Friday. Hunting magic? The clay bear of Montespan was seen as "killed" by spear holes stuck into it while the hunters danced around the anticipated prey. But the ceiling is low, the wall a meter away and the holes only the natural texture of the clay. Fertility magic and sex? Pregnant animals are identified from profile views, even though veterinarians cannot do that today. Certain little ivory carvings have long been held to be female figurines with exaggerated breasts or buttocks. Feminist scholars now see them as male genitalia: noisy signals!

Very few of the famous "Venuses" have any known context; we have almost 150 of them, only a few usually reproduced. Most may be much older than the big paintings. They hold no clue to race, nor are they in the main of unusual proportions. A "macho preoccupation with hunting and girls" now appears mere scholarly fashion. Marks and tallies are seen as part of a complex system, not idle doodlings, just as the placement of animal figures is held to be more systematic than random. But elaborated informational interpretations would seem to owe more to the computer age than to the Paleolithic. Shamanistic trance, initiation of the young, mnemonics and records of seasons, cosmology, group boundaries: proposals abound, and many an ingenious association certainly holds some truth.

There are messages to be read, complex and varied, sent over a wide span of purpose, time and place—but we do not know how to read them. Without informants, what can we expect? Modesty is the best stance for interpreters, and we are learning much. We can give a few exciting answers, by no means to everything. Look at the two small clay bison lying on their sides in Tuc d'Audoubert. We know the artist did not live in the dark where the work was done; only three or four people came and went with lamps to the dark chamber. The marks of hand and the simple tools of the sculptor are plain, and the fingerprinted clay "sausages" around them are the sculptor's discarded tests of the damp clay brought from the next room. "Opposite the bison we found the only engraving... like a signature," said Vertut. Our empathy for so manifest a continuity of human skills and aspiration transcends hypothesis.

ON FOOD AND COOKING: THE SCIENCE AND LORE OF THE KITCHEN, by Harold McGee. Collier Books, Macmillan Publishing Company, 1984 (paperbound, $16.95).

Many of us cook, and all of us eat. Diners who seek a subtler nourishment from the rich and surprising order that lies within cooking will find reward on just about every page of this thick, hugely informed, lighthearted and sensible book. It treats the nature and technology of all our principal foodstuffs, arranged in 10 chapters, from milk and cheese to chocolate and distilled spirits. It closes with the elements of digestive physiology and a summary of the physicochemical principles of cookery, organized around the key topic of heat transfer through air, water or oil to the food. Along the discursive way you embark on the history of ideas, peruse economic botany and the sociology of plantation crops, scan a run of outlandish and instructive recipes cited from the deep past ("Said Cato, 'The cabbage surpasses all other vegetables'") and move on to the worries of food additives and fads.

Novel illustrations display the microstructure of those domestic composites bread, butter and ice cream down to the level of the electron microscope. A vast but specialized literature supports the subject of food; after all, it is a major part of the world's economic activity. The specialized columns and journals retail food lore widely, but Harold McGee opens it up at depth, something of the whole system in one satisfying volume, the complexity accessible because it is close to what we see and spoon up and converse about daily. He ends with a few pages on the invisible—atoms, molecules and energy—to aid readers whose recollections of those concepts have faded; the primer is very well done (save for a few awry lines on proton-electron interactions).

There are many tables of properties of foods and simple flowcharts—not recipes, but explications—that present the main steps in the arts of making cheese, soy sauce, bean curd, chocolate, beer and wine. All of these are in context, each made part of an illuminating historical survey. Chocolate to eat—the name is Nahuatl, and today's crop is largely West African—is good for soldiers, wrote a 17th-century author, for it "makes Men to wake all night-long." Chocolate became a sweetmeat only with the advanced processing and the abundant sugar that came in the 19th century.

Soybean, now the largest American cash crop, used worldwide as feed and fodder, has long provided the cheese of China, soybean curd. For dairy cheeses, the proteins of bacteria-soured milk and some of its butterfat are coagulated with the enzyme rennin, extracted from the true stomach of a calf. But the soluble proteins of the filtered milk of the soybean are precipitated out by calcium sulfate, so that "all the Mass is as white as the very Snow.... Alone it is insipid, but very good dress'd as I say," to cite a traveling friar of the 17th century.

Only occasionally does the domestic art respond to the assertions of science. Consider the rise of the idea that the searing of meat before roasting it is necessary to retain juices and nutrition. It was the eminent chemist Justus von Liebig who made the practice common, at least in England and America, within 10 years after his research on muscle chemistry appeared in about 1850. He was a genuinely important chemist, but here he was all wrong. The juices of meat are in fact not lost in roasting, and the crust that forms on searing is not waterproof; even a long-boiled cut still holds most of its protein. But the Victorian cookbooks had found a modern rationale, true or false, and for most of a century, from decades before Fannie Farmer on past Escoffier, the books adhered to Liebig's plausible principle.

A few old cooks grumbled at the new way, but it was not until the 1930's that measurements showed that unseared meat lost less fluid than seared samples. For a generation the new (and also old) low-temperature method of roasting was certified as the method of choice. It did not last. By the 1970's initial searing was back in the popular cookbooks. Now the rationale had changed. "Searing does not seal, but it does brown." Browning unmistakably adds flavor (witness the hybrid microwave oven), and chefs now reasonably divide on a more delicate issue: whether to brown the roast first or last. Liebig's once-famous Extract—prepared from the waters to which long-boiled meat had, they said, given up its proteins—might make a fair soup, but its nutritive strength was long exaggerated. Liebig's untested model—a visible sealing layer able to retain decisive juices—was persuasive to the mind but ineffective to the roast, like many another received theory far from the kitchen.

One bitter taste from this feast of understanding: many plants contain an unstable sugar-cyanide complex. All the seeds of the rose family—apple, pear and citrus, as well as peach and the other stone fruits—share the substance. A single seed is no problem (avoid them as a rule), although people have died from roasting and eating a few dozen apple seeds. The news here is that the homely lima bean too can have plenty of cyanide complex! Only the safest varieties are grown in the U.S., but in Java and in Burma lima beans demand skillful preparation.

DYNAMICS OF DINOSAURS AND OTHER EXTINCT GIANTS, by R. McNeill Alexander. Columbia University Press, 1989 ($30.00).

Can those dry bones live? Professor Alexander, a biomechanicist at the University of Leeds, strings them together so that they lumber or cavort away before your mind's eye—they even sing. About all he needs is a little algebra, a few plastic models, a time recalled when once he chased antelope and giraffe across the grassy plains of Kenya and a finely tuned and contagious sense for apt physical approximation. He has put these simple but far from naive arguments into a wonderful little book that is "meant for everyone... scientists and nonscientists, schoolchildren and professors."

His broad method is the comparative engineering of large organisms, seen in the light of the physical necessities that link structure to function and of the style of design by evolution. The reader's pleasure and instruction spring from the directness of the work. Want to weigh dinosaurs? Buy from the great museum in London their plastic models made at a scale of 1 : 40. It is prudent to check dimensions; those commercial models indeed fit the best-known fossil skeletons species by species. Dinosaur volume is found by weighing the models in and out of water.

Just how dense were the dinosaurs in life? A British zoologist measured the density of nine dead Nile crocodiles; those crocs are about as large as modern reptiles come, and they are close relatives to the old giants. The dead beasts turned out to be 8 percent denser than water. Living crocodiles are seen both to float with all but the nostrils submerged and to lie quite still on the river bottom. Plainly, they can adjust their overall density by filling or deflating their lungs. It seems right, then, to assign to live dinosaurs the density of water. A generation ago another researcher did much the same work using his own plaster models. A table compares the results for 10 species: chunky *Triceratops* weighed in at six tons for Alexander and at nine tons for Edwin H. Colbert; their estimates for the long-necked *Brachiosaurus*—largest of the dinosaurs known from a nearly complete skeleton—were 87 and 47 tons, respectively.

This method depends on the uncontrolled judgment of the modeler who had to flesh out the bare bones. Another estimate has been tried recently by an international trio of zoologists, who measured the circumferences of the femur shafts of many modern quadrupedal mammals, from mice to elephants, with body weights recorded in life. (In fact, they used the main bones of all four legs.) The correlation is pretty tight: a fine power law, or a good straight line on a log-log plot. (You can use the right button on your calculator to take the 2.73 power needed, the author reminds us.) And yet the fit is not perfect: testing it for hippopotamus and elephant treated as unknowns, we miss either way by a factor of about 1.5 or 1.6. The estimate for *Brachiosaurus* is only 32 tons. Here there are no modeler's errors; however, the comparisons—made to creatures that are not very like the dinos—are based on only a few major bones for each. We have a consistent result, but only a rough one, for this simplest descriptor.

By just such means we arrive at the gait and speed of dinosaurs from their footprints, at their strengths of limb and back, at the posture and uses of their long necks and tails (the running bipeds had stiff, balanced tails that were never dragged), at their use of horns for ritual combat and of bony passages on the crests of the skull as resonant chambers employed in the calls and songs of courtship. More physics is needed to analyze motions, of course, but it remains close to everyday experience; the evolutionary arguments drawn on are equally straightforward.

These insights fall short of answering the question of whether the big old reptiles maintained their body temperature by metabolic heat or only by basking in the sun. The physics of heat transfer is now salient, and Professor Alexander shows us a cooling curve he made for a baked potato fresh from the oven. But the answer evades him still; so far physics is inconclusive and the predator-prey populations are too uncertain to make clear whether the carnivores were as few as a hot-blooded state would have required. "One thought that I find startling": if dinosaur energies were reptilelike and not mammalian, and if plants grew as well and rapidly then as now, dinosaurs could have been as numerous, weight for weight, as mammals of one fifth their mass are in modern populations. Think of the East African plains. In such a landscape imagine "for every 3-tonne elephant a 15-tonne *Diplodocus*.... The world may have seemed very full of dinosaurs."

It is pretty full of books about them now, but this one is irresistible for its simple and compelling demonstration of the unity of the natural sciences. If you would like to know about aquatic and flying reptiles or about the fat Madagascar bird with elephant's feet and massive eggs, you will have to read the book.

SCIENTIFIC AMERICAN

CORRESPONDENCE

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

Photocopying rights are hereby granted by Scientific American, Inc., to libraries and others registered with the Copyright Clearance Center (CCC) to photocopy articles in this issue of SCIENTIFIC AMERICAN for the flat fee of $1.50 per copy of each article or any part thereof. Such clearance does not extend to the photocopying of articles for promotion or other commercial purposes. Correspondence and payment should be addressed to Copyright Clearance Center, Inc., 21 Congress Street, Salem, Mass. 01970. Specify CCC Reference Number ISSN 0036-8733/89. $1.50 + 0.00.

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

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

Address subscription correspondence to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 3187, Harlan, IA. 51593. Telephone inquiries: 1-800-333-1199, U.S. only; other 515-247-7631/32. The date of the last issue on subscriptions appears on each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (if convenient, 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 ______________________

ESSAY

The sorry state of science education

by P. Roy Vagelos

All of us who remember Sputnik also remember being jolted out of complacency. We learned that the U.S. was technologically unprepared for the space age. Consequently, the Soviet space coup set in motion a major reappraisal of American education and imparted a new urgency to training for careers in science and engineering.

Now, three decades later, America finds itself lulled again. In spite of having universities and research institutions that remain the envy of the world, the U.S. is clearly failing to make its young people literate in science and mathematics; the educational system is not even awakening their interest in these fields. It is time to ask whether the system can respond to the challenge of maintaining this country's technological leadership.

As a physician and biochemist, heading a company that has been able to discover and introduce a series of important new medicines, I am concerned about the effect the decline of U.S. scientific education must have on the health of Americans and indeed of people throughout the world. Of the medicines introduced to the American market between 1940 and 1987, 62 percent originated in this country. If U.S. innovation flags, the remarkable progress against disease that has continued for half a century is likely to slow as well.

The effect of educational failure will be felt in other high-technology industries as well. In the global competition of the future they too will need workers equipped with a basic scientific and mathematical education. Scientifically literate managers will be correspondingly important.

In a democratic society the resolution of policy questions also demands scientific knowledge. And so it is disturbing that the men and women who will be the country's leaders in the 21st century are not being equipped to think intelligently about the environment, energy, space, defense and biotechnology. Equally disturbing, the vast majority of our population is failing to receive the kind of education that would ensure wider public understanding of, and support for, technological progress. Without such support, in a democracy even the best leadership will be powerless.

Although the list of troubles is already distressingly long, another must be added: the career scientists and engineers needed by universities and industry are not being trained in sufficient numbers. The past decade has witnessed a decline in the proportion of our students majoring in the sciences and engineering or receiving advanced degrees in those fields. Only seven out of every 1,000 U.S. students earn engineering degrees; in Japan the figure is 40. More than half of the new U.S. doctoral degrees in engineering, mathematics and physics go to foreign nationals. In the life sciences the number of Ph.D.'s conferred remains about the same as a decade ago; master's degrees, however, have declined by nearly 25 percent and undergraduate degrees by nearly 30 percent, thus reducing the base for future scientific training and leadership.

Secondary schools do far too little to encourage interest in science. A survey by the National Science Teachers Association of U.S. high schools found that one third of them offered no physics course, one fifth no chemistry and one tenth no biology; almost three fourths offered no earth or space science. Another survey found that in a field of 13 countries, U.S. high school seniors in advanced physics ranked ninth, seniors in advanced chemistry ranked eleventh and those in advanced biology were last.

The shortfall in our capacity to teach science, carry out research projects and find applications for new knowledge can have only one outcome: a reduction in the quantity and quality of the industrial innovations on which the American standard of living depends.

It seems to me that the dismal showing of U.S. high schools can be improved if we can attract young Americans to science at an earlier age—starting such instruction in at least the middle grades. One way to attract them is through the life sciences, which show how attributes that young people respond to—physical vigor and well-being—can be promoted by biomedical research.

For example, I know from letters I have received that children as young as eight or nine are fascinated to learn about onchocerciasis (river blindness) and the sight-saving effects of a Merck drug discovery that we are donating for use in the impoverished tropical countries where this scourge is prevalent. Such a story might well motivate some youngsters to choose a career in science.

In the meantime corporations are continuing efforts to improve educational performance at both university and precollege levels. Merck alone gave more than $8 million to scientific education last year, and we were by no means the largest contributor. Industry funded approximately 8 percent of all university research in 1987; that figure is expected soon to increase to between 10 and 12 percent.

But the crisis cannot be solved by corporate support or university fund-raising alone. Government at all levels must take steps to strengthen the science and mathematics curriculums of schools, colleges and universities. The nation must summon the resources to prepare teachers adequately and to modernize the equipment with which they learn, teach and do research. Finally, increased funding of scholarships and other support programs is needed to attract more of our brightest students to science careers, including research and teaching.

Unless we take such action, other countries will reap the human, economic and cultural benefits of tomorrow's breakthrough products. In my own field, better treatments for cancer, mental illness, osteoporosis, Alzheimer's disease and other serious health problems may have to be developed elsewhere.

Experience in doing and managing science has taught me that in order to succeed, any research project must have its own champions—people truly committed to focusing attention and energy on that specific project under all circumstances and for as long as necessary. I urge readers to champion a project as ambitious in its way as the Manhattan Project or today's search for an AIDS cure: the improvement of science education in our country.

P. ROY VAGELOS is chairman, president and chief executive officer of Merck & Co., Inc. He did research at the National Heart Institute and at the Washington University School of Medicine and then became director of the university's division of biology and biomedical sciences. Vagelos joined the health-products firm's Merck Sharp & Dohme Research Laboratories in 1975 as senior vice-president for research.


Dewar's
"White Label" Never Varies
"White Label"
DEWAR'S
BLENDED SCOTCH WHISKY
100% SCOTCH WHISKIES
PRODUCT OF SCOTLAND
John Dewar & Sons Ltd
750 ml
Cottage: Isle of Skye, Scotland
3 br. kitchen, 2 peat burning f.p.
Mt. and ocean vu. Not for sale.
In family forever.
To Let:
Taste Dewar's Scotland.
Where the good things in life stay that way.